



Universidad de Valladolid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Departamento de Teoría de la Señal y Comunicaciones e
Ingeniería Telemática

TESIS DOCTORAL

Analysis of Nocturnal Oximetry Recordings using Pattern Recognition Techniques to Assist in the Diagnosis of the Sleep Apnoea-Hypopnoea Syndrome

Presentada por José Víctor Marcos Martín para optar al grado de doctor por la
Universidad de Valladolid con Mención Europea

Dirigida por:
Roberto Hornero Sánchez

Año académico 2010/2011



Universidad de Valladolid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática

TESIS DOCTORAL

Analysis of Nocturnal Oximetry Recordings using Pattern Recognition Techniques to Assist in the Diagnosis of the Sleep Apnoea-Hypopnoea Syndrome

DOCTORANDO: JOSÉ VÍCTOR MARCOS MARTÍN

DIRECTOR: ROBERTO HORNERO SÁNCHEZ

DIRECTOR DE LA ESTANCIA EN ASTON UNIVERSITY: IAN T. NABNEY

TRIBUNAL

PRESIDENTE: DR. D.

VOCALES: DR. D.

DR. D.

DR. D.

SECRETARIO: DR. D.

SUPLENTES: DR. D.

DR. D.

FECHA DE LECTURA:

CALIFICACIÓN:



Universidad de Valladolid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática

DATOS DE LA ESTANCIA DE INVESTIGACIÓN

LUGAR: BIRMINGHAM (REINO UNIDO)

FECHAS: 27/05/2009 – 26/07/2009
14/09/2010 – 30/10/2010

DURACIÓN: 108 DÍAS (3.6 MESES)

INSTITUCIÓN: ASTON UNIVERSITY

DIRECTOR DE LA ESTANCIA: PROFESSOR IAN T. NABNEY

Abstract

Sleep apnoea-hypopnoea syndrome (SAHS) is the most common form of sleep disordered-breathing. SAHS is characterised by repetitive occlusion of the upper airway during sleep, causing intermittent cessations of breathing (apnoeas) or reduction in airflow (hypopnoeas). It is associated to hypoxaemia, bradycardia and fragmented sleep. SAHS has been pointed out as a major cause of traffic and industrial accidents due to excessive daytime sleepiness. Long-term effects can lead to severe cardiovascular and cerebrovascular diseases. Its prevalence is estimated at 5% of the adult population in western countries. Moreover, it is suspected that a high percentage of patients suffering from SAHS remain undiagnosed. Therefore, SAHS can be considered as a risk factor for public health.

Nowadays, nocturnal polysomnography (PSG) is the gold-standard for SAHS diagnosis. PSG must be performed in a special sleep unit and under supervision of a trained technician. Different physiological recordings and data are monitored during a complete night. They must be manually analysed by an expert to obtain a definitive diagnosis. It is based on the value of the apnoea-hypopnoea index (AHI), which measures the ratio of apnoeas and hypopnoeas per hour of sleep. Despite its high diagnostic performance, PSG presents some drawbacks since it is complex, expensive and time-consuming. Additionally, the demand for PSG studies is progressively growing as people and clinicians are becoming aware of SAHS while the available infrastructure is insufficient to support it. Thus, simplified diagnostic techniques are desirable.

This Thesis proposes to analyse oxygen saturation (SaO_2) signals recorded through nocturnal pulse oximetry to assist in SAHS diagnosis. Pulse oximetry is a non-invasive technique to monitor arterial blood oxygenation. It can be performed at patient's home, resulting in reduced complexity and cost in comparison with PSG. SaO_2 recordings reflect hypoxaemia due to airflow reduction during apnoea/hypopnoea events. As a result, SaO_2 signals from SAHS patients tend to be more unstable than those from control subjects due to the recurrence of apnoeas during sleep. This different

behaviour can be exploited to detect SAHS. In addition to visual inspection, several conventional oximetry indices have been suggested for automated interpretation of SaO₂ signals. The oxygen desaturation index over 3% (ODI3) and 4% (ODI4) as well as the cumulative time spent below 90% of saturation (CT90) are the most popular oximetry indices. In addition, other measurements such as the minimum SaO₂ value (minSaO₂), the saturation impairment time at 90% (SIT90) or the Δ index have been proposed. However, none of them has been accepted as a reliable predictor of SAHS.

A novel method for automated SaO₂ analysis is proposed in the Thesis. Pattern recognition techniques were used to model SAHS diagnosis from oximetry data. The applied methodology comprised three different stages: 1) feature extraction, 2) normalisation and dimensionality reduction, and 3) pattern analysis. In the first one, a total of 14 time-domain and frequency-domain features were extracted from SaO₂ recordings in order to characterise their dynamical behaviour. In the second stage, the distribution of each feature was normalised to have zero mean and unit variance. The utility of dimensionality reduction by means of principal component analysis (PCA) after normalisation was evaluated. In the third stage, pattern recognition techniques were used to process the features (or components) from the previous phase. Two different approaches were proposed to model SAHS diagnosis: classification and regression. The output of classification methods represents a categorical variable indicating one of two possible groups for the input pattern: SAHS-negative and SAHS-positive. The regression approach aims to provide accurate estimations of the AHI from information contained in the input pattern.

A database composed of 240 SaO₂ signals was available for this study. It was randomly divided into a training set (40%) with 96 SaO₂ signals (32 SAHS-negative and 64 SAHS-positive) and a test set (60%) with 144 SaO₂ signals (48 SAHS-negative and 96 SAHS-positive). The former was used for model selection and optimisation while the latter was allocated for assessing the trained algorithms. Several pattern recognition methods were evaluated for both tasks. To analyse the effect of dimensionality reduction, each of them was assessed with and without PCA in the second stage. A total of 16 classification algorithms and 14 regression algorithms were built. The classifier with the highest performance was based on the analysis of the complete set of normalised features by means of a multilayer perceptron (MLP)

network. A classification accuracy of 92.36% on the test set was reached (94.79% sensitivity and 87.50% specificity). This algorithm outperformed the Δ index, which provided the best classification results (88.19% accuracy) among the conventional oximetry indices. The highest accuracy in the regression task was achieved by two algorithms: one based on multivariate adaptive regression splines (MARS) for processing the complete set of normalised features and another based on Bayesian MLP networks using the components retained from PCA. Both reached an intraclass correlation coefficient (*ICC*) higher than 0.9. Moreover, the diagnostic accuracy provided by the estimated AHI was 87.50% (88.54% sensitivity and 85.42% specificity) for both of them. Despite ODI3 also achieved high *ICC*, it showed to be significantly inaccurate for predicting small AHI values. Most of the subjects misdiagnosed by the selected classification and regression algorithms had mild SAHS ($5 \text{ h}^{-1} \leq \text{AHI} \leq 15 \text{ h}^{-1}$). Thus, these algorithms represent valuable screening tools that could contribute to reduce the number of required PSG tests.

Resumen

El síndrome de la apnea-hipopnea del sueño (SAHS) es un trastorno respiratorio caracterizado por la oclusión intermitente y repetitiva de la vía aérea superior durante el sueño, produciéndose el cese completo de la respiración (apnea) o una reducción considerable del flujo aéreo (hipopnea). El SAHS se asocia a fenómenos como la hipoxemia y la bradicardia, a la vez que impide un descanso reparador. Además, es una de las causas principales de accidentes de tráfico y laborales producidos como consecuencia de la somnolencia diurna excesiva. A largo plazo, el SAHS puede dar lugar a severos trastornos de tipo cardiovascular y cerebrovascular. Su prevalencia se estima en el 5% de la población adulta occidental, aunque se cree que existe un elevado porcentaje de enfermos sin diagnosticar. Por tanto, el SAHS puede considerarse como un factor de riesgo para la salud pública.

Actualmente, la polisomnografía nocturna (PSG) es el método de referencia en el diagnóstico del SAHS. Esta prueba se realiza en una unidad del sueño especialmente acondicionada y bajo la supervisión de personal cualificado. Durante la noche se monitorizan varios datos y registros fisiológicos. Posteriormente, éstos han de ser analizados manualmente por un médico especialista para obtener un diagnóstico definitivo. El índice de apnea-hipopnea (AHI), que mide el número de eventos de apnea e hipopnea por hora de sueño, es el criterio empleado para establecer el diagnóstico. A pesar de su elevada precisión, la PSG presenta ciertos inconvenientes, ya que se trata de un procedimiento excesivamente tedioso para el especialista médico, complejo y de elevado coste. Por otro lado, la demanda de estudios polisomnográficos se incrementa progresivamente debido al creciente interés de pacientes y profesionales de la salud en la enfermedad del SAHS. Sin embargo, la infraestructura disponible actualmente resulta insuficiente para hacerla frente. Por tanto, sería deseable disponer de métodos de diagnóstico más simples y accesibles.

En esta Tesis se propone analizar las señales de saturación de oxígeno en sangre (SaO_2) obtenidas mediante pulsioximetría nocturna para la ayuda en el diagnóstico del SAHS. La pulsioximetría es una técnica no invasiva que permite

monitorizar el nivel de oxigenación en la sangre arterial. Esta prueba puede realizarse en el hogar del paciente, reduciéndose así la complejidad y el coste asociados a la PSG. Los registros de SaO_2 reflejan la hipoxemia producida durante los eventos de apnea o hipopnea. Por tanto, debido a la repetición de apneas durante el sueño, las señales de SaO_2 pertenecientes a pacientes con SAHS tienden a ser más inestables que en el caso de sujetos sanos. Las diferencias en el comportamiento dinámico de estas señales pueden emplearse para la detección del SAHS. Además de la mera inspección visual, el análisis de los registros de SaO_2 puede automatizarse mediante la utilización de los índices oximétricos convencionales. Entre éstos, los más comúnmente empleados son los índices de desaturación de oxígeno por encima del 3% (ODI3) y el 4% (ODI4), así como el tiempo acumulado con un nivel de saturación por debajo del 90% (CT90). Otros índices oximétricos son el valor mínimo de saturación (minSaO_2), el índice de severidad de las desaturaciones referenciado al 90% (SIT90) o el índice Δ . Sin embargo, ninguno de ellos puede ser considerado un indicador fiable del SAHS.

En esta Tesis se presenta un método novedoso para el análisis automático de los registros de SaO_2 . Éste se basa en la utilización de técnicas de reconocimiento de patrones para modelar el problema de diagnóstico del SAHS a partir de la señal de oximetría. La metodología empleada se estructura en tres etapas diferentes: 1) extracción de características, 2) normalización y reducción de la dimensionalidad, y 3) reconocimiento de patrones. En la primera de ellas, se definieron 14 características temporales y frecuenciales de los registros de SaO_2 para reflejar propiedades dinámicas indicadoras de SAHS. En la segunda etapa, cada característica fue normalizada, obteniéndose, para cada una de ellas, una distribución de valores con media nula y desviación típica igual a la unidad. Asimismo, se evaluó la utilidad del análisis por componentes principales (PCA) como herramienta para la reducción de la dimensionalidad del patrón de características. En la tercera etapa, se emplearon técnicas de reconocimiento de patrones para procesar las características (o componentes) procedentes de la etapa anterior. Se evaluaron dos modelos diferentes para el problema de diagnóstico del SAHS mediante el análisis de patrones multivariante: clasificación y regresión. En el caso de los métodos de clasificación, la salida del algoritmo de ayuda al diagnóstico es una variable categórica que indica a qué grupo pertenece el patrón de entrada: SAHS negativos o SAHS positivos. En cambio, los métodos para el análisis de

regresión tratan de proporcionar una estimación precisa del AHI a partir de la información contenida en el patrón de entrada.

Para el estudio se dispuso de una base de datos compuesta por 240 señales de SaO₂. Ésta se dividió aleatoriamente en un conjunto de entrenamiento (40%) con 96 señales (32 SAHS negativos y 64 SAHS positivos) y un conjunto de test (60%) con 144 señales (48 SAHS negativos y 96 SAHS positivos). El primero de ellos fue utilizado en las operaciones de selección y optimización de modelos, mientras que el segundo fue reservado para validar los algoritmos una vez entrenados. Se evaluaron varios métodos de reconocimiento de patrones tanto para la clasificación como para la regresión. Además, por cada uno de ellos, se implementaron dos algoritmos diferentes (en función de la utilización o no de PCA) para analizar el efecto de la reducción de dimensionalidad. Finalmente, un total de 16 algoritmos de clasificación y 14 algoritmos de regresión fueron evaluados. El clasificador con los mejores resultados fue aquél basado en el análisis del conjunto completo de características normalizadas mediante una red perceptrón multicapa (MLP). Este algoritmo logró una precisión del 92.36% en el conjunto de test (sensibilidad de 94.79% y especificidad de 87.50%). Estos resultados mejoraron los obtenidos mediante el índice Δ , el índice oximétrico con la mayor capacidad diagnóstica (precisión de 88.19%) sobre la base de datos utilizada. En cuanto a los modelos de regresión, los mejores resultados fueron alcanzados por dos algoritmos: aquél basado en el análisis del conjunto completo de características normalizadas mediante “*splines*” multivariante de regresión adaptativa (MARS) y el algoritmo de redes MLP Bayesianas para el procesado de las componentes derivadas de PCA. El valor del coeficiente de correlación intraclase (*ICC*) proporcionado por ambos fue superior a 0.90. Además, la precisión diagnóstica del AHI estimado por estos algoritmos fue 87.50% (sensibilidad de 88.54% y especificidad de 85.42%). A pesar de que mediante el ODI3 también puede alcanzarse un *ICC* elevado, se comprobó que la estimación proporcionada por este índice posee un error significativo para valores pequeños del AHI. Finalmente, se verificó que gran parte de los errores de diagnóstico cometidos por los algoritmos de clasificación y regresión seleccionados se producen en enfermos con SAHS leve ($5 \text{ h}^{-1} < \text{AHI} < 15 \text{ h}^{-1}$). Por tanto, estos algoritmos representan herramientas válidas para el estudio de pacientes, con lo que podrían contribuir a reducir el número de pruebas polisomnográficas.

Table of contents

1.	Introduction.....	1
1.1.	Context: biomedical signal processing.....	2
1.2.	Problem formulation.....	4
1.3.	Hypothesis.....	5
1.4.	Aims of the Doctoral Thesis.....	7
1.5.	Work plan.....	10
1.6.	Principles of pattern recognition.....	13
1.6.1.	The pattern classification problem.....	14
1.6.2.	Regression analysis.....	16
1.6.3.	Learning and generalisation.....	20
1.6.4.	The bias-variance trade-off.....	21
1.6.5.	Regularisation.....	23
1.6.6.	Maximum likelihood and Bayesian inference.....	24
1.7.	Structure of the document.....	26
2.	The sleep apnoea-hypopnoea syndrome.....	29
2.1.	Introduction.....	30
2.2.	Physiopathology.....	31
2.3.	Clinical presentation.....	32
2.4.	Epidemiology and clinical consequences.....	33
2.5.	Diagnosis.....	34
2.6.	Diagnostic utility of nocturnal pulse oximetry.....	36
2.6.1.	Principles of pulse oximetry.....	36
2.6.2.	Clinical applications.....	38
2.6.3.	Potentials and limitations of nocturnal pulse oximetry.....	39
2.7.	Treatment.....	40
3.	State of the art.....	43
3.1.	Introduction.....	44
3.2.	Signal processing-based methods to assist in SAHS diagnosis.....	45
3.2.1.	Automated analysis of nocturnal oximetry to assist in SAHS detection.....	46
3.2.2.	Other biomedical signals and data related to SAHS.....	51

3.3.	Utility of pattern recognition techniques as an aid tool for SAHS diagnosis	54
4.	Subjects and signals	59
4.1.	Population under study	60
4.2.	Sleep study and diagnosis	61
4.3.	Signal preprocessing	62
5.	Feature extraction, normalisation and dimensionality reduction	65
5.1.	Motivation	66
5.2.	Feature extraction	67
5.2.1.	Time-domain analysis	67
5.2.2.	Frequency-domain analysis	75
5.3.	Normalisation	79
5.4.	Principal component analysis (PCA)	79
6.	Pattern recognition methods for classification	83
6.1.	Introduction	84
6.2.	<i>K</i> -nearest neighbours (KNN)	84
6.3.	Fisher's linear discriminant (FLD)	86
6.4.	Logistic regression (LR)	89
6.5.	Probabilistic neural networks (PNN)	91
6.6.	Multilayer perceptron (MLP) networks for classification	92
6.7.	Radial basis function (RBF) networks for classification	98
6.8.	Bayesian (BY) networks for classification	103
6.9.	Support vector machines (SVM) for classification	109
7.	Pattern recognition methods for regression	115
7.1.	Introduction	116
7.2.	Multiple linear regression (MLR)	116
7.3.	Generalised regression neural networks (GRNN)	117
7.4.	Multilayer perceptron (MLP) networks for regression	119
7.5.	Radial basis function (RBF) networks for regression	120
7.6.	Bayesian (BY) networks for regression	122
7.7.	Multivariate adaptive regression splines (MARS)	126
7.8.	Support vector machines (SVM) for regression	131

8.	Results.....	137
8.1.	Introduction.....	138
8.2.	Performance assessment	138
8.2.1.	Performance measures for classification	138
8.2.2.	Performance measures for regression	140
8.3.	Model selection and performance estimation	143
8.4.	Feature extraction	144
8.5.	Feature normalisation and dimensionality reduction.....	151
8.6.	Classification results.....	152
8.6.1.	Design and training of classification algorithms	152
8.6.1.1.	<i>Design of KNN classifiers.....</i>	<i>153</i>
8.6.1.2.	<i>Design of FLD classifiers.....</i>	<i>154</i>
8.6.1.3.	<i>Design of LR classifiers.....</i>	<i>155</i>
8.6.1.4.	<i>Design of PNN classifiers.....</i>	<i>156</i>
8.6.1.5.	<i>Design of MLP classifiers.....</i>	<i>158</i>
8.6.1.6.	<i>Design of RBF classifiers.....</i>	<i>162</i>
8.6.1.7.	<i>Design of BY network classifiers.....</i>	<i>164</i>
8.6.1.8.	<i>Design of SVM classifiers.....</i>	<i>167</i>
8.6.2.	Testing classification algorithms	168
8.7.	Regression results	171
8.7.1.	Design and training of regression algorithms	171
8.7.1.1.	<i>Design of MLR algorithms</i>	<i>172</i>
8.7.1.2.	<i>Design of GRNN algorithms.....</i>	<i>173</i>
8.7.1.3.	<i>Design of MLP algorithms for regression.....</i>	<i>174</i>
8.7.1.4.	<i>Design of RBF algorithms for regression.....</i>	<i>177</i>
8.7.1.5.	<i>Design of BY network algorithms for regression.....</i>	<i>178</i>
8.7.1.6.	<i>Design of MARS algorithms</i>	<i>180</i>
8.7.1.7.	<i>Design of SVM algorithms for regression</i>	<i>182</i>
8.7.2.	Testing regression algorithms.....	184
8.8.	Conventional analysis of oximetry signals.....	188
9.	Discussion and conclusions	193
9.1.	Introduction.....	194
9.2.	Discussion.....	194
9.3.	Conclusions.....	207
9.4.	Original contributions.....	213
9.5.	Future work.....	216
	Appendix A. Glossary of terms.....	219

Appendix B. Publications derived from the Doctoral Thesis	223
B.1. Publications in peer-reviewed journals included in the Journal Citation Reports.....	224
B.2. Publications in international conferences	225
B.3. Publications in national conferences	226
Appendix C. Síntesis de la Tesis Doctoral en castellano	231
C.1. Introducción.....	232
C.2. El síndrome de la apnea-hipopnea del sueño.....	234
C.3. Estado de la técnica	235
C.4. Sujetos y señales	236
C.5. Extracción de características, normalización y reducción de dimensionalidad.....	237
C.6. Técnicas de reconocimiento de patrones para clasificación	239
C.7. Técnicas de reconocimiento de patrones para regresión	239
C.8. Resultados.....	239
C.9. Discusión y conclusiones.....	241
C.9.1. Conclusiones.....	243
C.9.2. Contribuciones originales	248
C.9.3. Líneas futuras	252
Bibliography.....	255

Contenidos

1.	Introducción.....	1
1.1.	Contexto: procesado de señales biomédicas.....	2
1.2.	Definición del problema	4
1.3.	Hipótesis de partida	5
1.4.	Objetivos.....	7
1.5.	Plan de trabajo	10
1.6.	Fundamentos del reconocimiento de patrones.....	13
1.6.1.	Clasificación de patrones.....	14
1.6.2.	Análisis de regresión	16
1.6.3.	Aprendizaje y generalización	20
1.6.4.	El compromiso entre sesgo y varianza	21
1.6.5.	Regularización.....	23
1.6.6.	Principios de máxima verosimilitud e inferencia Bayesiana	24
1.7.	Estructura del documento	26
2.	El síndrome de la apne-hipopnea del sueño.....	29
2.1.	Introducción.....	30
2.2.	Fisiopatología	31
2.3.	Manifestaciones clínicas.....	32
2.4.	Epidemiología e impacto clínico	33
2.5.	Diagnóstico	34
2.6.	Utilidad diagnóstica de la pulsioximetría nocturna	36
2.6.1.	Principio de funcionamiento.....	36
2.6.2.	Aplicaciones	38
2.6.3.	Ventajas e inconvenientes de la pulsioximetría nocturna	39
2.7.	Tratamiento.....	40
3.	Estado de la técnica.....	43
3.1.	Introducción.....	44
3.2.	Métodos de ayuda al diagnóstico del SAHS basados en el procesado de señales.....	45
3.2.1.	Análisis automático de la oximetría nocturna para la ayuda al diagnóstico del SAHS.....	46
3.2.2.	Otras señales biomédicas relacionadas con el SAHS	51

3.3.	Utilidad de las técnicas de reconocimiento de patrones como herramienta de ayuda al diagnóstico del SAHS	54
4.	Sujetos y señales	59
4.1.	Población de estudio	60
4.2.	Registro de señales y diagnóstico	61
4.3.	Preprocesado de las señales	62
5.	Extracción de características, normalización y reducción de dimensionalidad	65
5.1.	Motivación	66
5.2.	Extracción de características	67
5.2.1.	Análisis en el dominio del tiempo	67
5.2.2.	Análisis en el dominio de la frecuencia	75
5.3.	Normalización	79
5.4.	Análisis de componentes principales (PCA)	79
6.	Técnicas de reconocimiento de patrones para la clasificación	83
6.1.	Introducción	84
6.2.	<i>K</i> -vecinos próximos (KNN)	84
6.3.	Discriminante lineal de Fisher (FLD)	86
6.4.	Regresión logística (LR)	89
6.5.	Redes neuronales probabilísticas (PNN)	91
6.6.	Clasificación mediante redes perceptrón multicapa (MLP)	92
6.7.	Clasificación mediante redes de funciones de base radial (RBF)	98
6.8.	Clasificación mediante redes Bayesianas (BY)	103
6.9.	Clasificación mediante máquinas de vectores soporte (SVM)	109
7.	Técnicas de reconocimiento de patrones para la regresión	115
7.1.	Introducción	116
7.2.	Regresión lineal múltiple (MLR)	116
7.3.	Redes neuronales de regresión generalizada (GRNN)	117
7.4.	Regresión mediante redes perceptrón multicapa (MLP)	119
7.5.	Regresión mediante redes de funciones de base radial (RBF)	120
7.6.	Regresión mediante redes Bayesianas (BY)	122
7.7.	Splines multivariante de regresión adaptativa (MARS)	126
7.8.	Regresión mediante máquinas de vectores soporte (SVM)	131

8.	Resultados.....	137
8.1.	Introducción.....	138
8.2.	Evaluación del rendimiento.....	138
8.2.1.	Medidas del rendimiento en clasificación.....	138
8.2.2.	Medidas del rendimiento en regresión.....	140
8.3.	Selección del modelo y estimación del rendimiento.....	143
8.4.	Extracción de características.....	144
8.5.	Normalización de características y reducción de dimensionalidad.....	151
8.6.	Resultados de clasificación.....	152
8.6.1.	Diseño y entrenamiento de los algoritmos de clasificación.....	152
8.6.1.1.	<i>Diseño de los clasificadores KNN</i>	153
8.6.1.2.	<i>Diseño de los clasificadores FLD</i>	154
8.6.1.3.	<i>Diseño de los clasificadores LR</i>	155
8.6.1.4.	<i>Diseño de los clasificadores PNN</i>	156
8.6.1.5.	<i>Diseño de los clasificadores MLP</i>	158
8.6.1.6.	<i>Diseño de los clasificadores RBF</i>	162
8.6.1.7.	<i>Diseño de los clasificadores de red BY</i>	164
8.6.1.8.	<i>Diseño de los clasificadores SVM</i>	167
8.6.2.	Validación de los algoritmos de clasificación.....	168
8.7.	Resultados de regresión.....	171
8.7.1.	Diseño y entrenamiento de los algoritmos de regresión.....	171
8.7.1.1.	<i>Diseño de los algoritmos MLR</i>	172
8.7.1.2.	<i>Diseño de los algoritmos GRNN</i>	173
8.7.1.3.	<i>Diseño de los algoritmos de regresión de red MLP</i>	174
8.7.1.4.	<i>Diseño de los algoritmos de regresión de red RBF</i>	177
8.7.1.5.	<i>Diseño de los algoritmos de regresión de red BY</i>	178
8.7.1.6.	<i>Diseño de los algoritmos MARS</i>	180
8.7.1.7.	<i>Diseño de los algoritmos de regresión SVM</i>	182
8.7.2.	Validación de los algoritmos de regresión.....	184
8.8.	Métodos tradicionales para el análisis de la oximetría.....	188
9.	Discusión y conclusiones.....	193
9.1.	Introducción.....	194
9.2.	Discusión.....	194
9.3.	Conclusiones.....	207
9.4.	Contribuciones originales.....	213
9.5.	Líneas futuras.....	216
	Apéndice A. Glosario de términos.....	219

Apéndice B. Publicaciones derivadas de la Tesis Doctoral	223
B.1. Publicaciones con revisión por pares en revistas incluidas en el Journal Citation Reports.....	224
B.2. Publicaciones en congresos internacionales	225
B.3. Publicaciones en congresos nacionales	226
Apéndice C. Síntesis de la Tesis Doctoral en castellano	231
C.1. Introducción.....	232
C.2. El síndrome de la apnea-hipopnea del sueño.....	234
C.3. Estado de la técnica	235
C.4. Sujetos y señales.....	236
C.5. Extracción de características, normalización y reducción de dimensionalidad.....	237
C.6. Técnicas de reconocimiento de patrones para clasificación	239
C.7. Técnicas de reconocimiento de patrones para regresión	239
C.8. Resultados.....	239
C.9. Discusión y conclusiones.....	241
C.9.1. Conclusiones.....	243
C.9.2. Contribuciones originales	248
C.9.3. Líneas futuras	252
Bibliografía.....	255

List of Figures

Figure 1. Examples of SaO_2 signals corresponding to (a) a SAHS-negative subject ($\text{AHI} = 1.4 \text{ h}^{-1}$), (b) a doubtful SAHS-negative subject ($\text{AHI} = 8.1 \text{ h}^{-1}$), (c) a doubtful SAHS-positive subject ($\text{AHI} = 13.1 \text{ h}^{-1}$) and (d) a SAHS-positive subject ($\text{AHI} = 66.2 \text{ h}^{-1}$).

Figure 2. Graphical representation of Eq. (17), which indicates that a model optimised by minimising the sum-of-squares error approximates the conditional average of the target data. A one-dimensional function with a single independent variable is depicted (Bishop, 1995).

Figure 3. Effect of Bayesian inference. Distribution of adaptive parameters before and after the training data in D is observed. As shown in the picture, a narrower form is obtained for the posterior distribution indicating that some values of \mathbf{w} are preferred once training data is available (Bishop, 1995).

Figure 4. Scheme of the physiological cycle associated to apnoea/hypopnoea events.

Figure 5. A clinician observes the recordings from a patient during a PSG test in a sleep laboratory (Science Photo Library ©).

Figure 6. Absorption spectra of normal adult haemoglobin in saturated (OxyHb) and desaturated (DeOxyHb) oxyhaemoglobin states, carboxyhaemoglobin (COHb) and methaemoglobin (MetHb) (Moyle, 1996).

Figure 7. Example of a SaO_2 recording (a) before preprocessing and (b) after artefact removal using the proposed algorithm. As can be observed, drops to zero and subsequent abnormal values of saturation are removed.

Figure 8. Comparison of dynamic properties of oximetry signals corresponding to clear control subjects and subjects suffering from SAHS: (a) an one-hour epoch of a SaO_2 signal from a subject with $\text{AHI} = 1.4 \text{ h}^{-1}$; (b) an one-hour epoch of a SaO_2 signal from a subject with $\text{AHI} = 66.2 \text{ h}^{-1}$.

Figure 9. Distribution of the variable SaO_2 for (a) a SAHS-negative subject ($\text{AHI} = 1.4 \text{ h}^{-1}$), (b) a doubtful SAHS-negative subject ($\text{AHI} = 8.1 \text{ h}^{-1}$), (c) a doubtful SAHS-positive subject ($\text{AHI} = 13.1 \text{ h}^{-1}$) and (d) a SAHS-positive subject ($\text{AHI} = 66.2 \text{ h}^{-1}$).

Figure 10. PSD function corresponding to (a) a SAHS-negative subject ($\text{AHI} = 1.4 \text{ h}^{-1}$), (b) a doubtful SAHS-negative subject ($\text{AHI} = 8.1 \text{ h}^{-1}$), (c) a doubtful SAHS-positive subject ($\text{AHI} = 13.1 \text{ h}^{-1}$) and (c) a SAHS-positive subject ($\text{AHI} = 66.2 \text{ h}^{-1}$).

Figure 11. Example of the distribution of samples in two different classes. A high overlap is obtained if samples are projected in the direction for which the maximum difference between means is obtained (Bishop, 1995).

Figure 12. Scheme of a perceptron. Its output is obtained from the weighted sum of the input features, including a bias term, and a non-linear activation function.

Figure 13. Scheme of a generic MLP network with a single hidden layer with N_H hidden neurons and an output layer with N_O output neurons.

- Figure 14.** Average probability density function of the variable representing the SaO₂ amplitude computed from SAHS-positive and SAHS-negative subjects in the training set.
- Figure 15.** Average PSD function computed from SAHS-positive and SAHS-negative subjects in the training set.
- Figure 16.** Results from PCA analysis of the normalised features: (a) variance explained for each of the components and (b) cumulative variance for the first P_c components.
- Figure 17.** Influence of the number of nearest neighbours (K) on classification accuracy for (a) the *All-KNN_c* classifier and (b) the *PCA-KNN_c* classifier.
- Figure 18.** Influence of the spread parameter (controlled by τ) on classification accuracy for (a) the *All-PNN_c* classifier and (b) the *PCA-PNN_c* classifier.
- Figure 19.** Influence of the number of hidden nodes (N_H) and the regularisation parameter (ν) on classification accuracy for (a) the *All-MLP_c* classifier and (b) the *PCA-MLP_c* classifier.
- Figure 20.** Effect of weight decay regularisation on the size of network weights: (a) *All-MLP_c* algorithms with $\nu = 0.3$ and different N_H ; (b) *All-MLP_c* algorithms with $N_H = 25$ and different ν .
- Figure 21.** Effect of weight decay regularisation on the size of network weights: (a) *PCA-MLP_c* algorithms with $\nu = 0.8$ and different N_H ; (b) *PCA-MLP_c* algorithms with $N_H = 29$ and different ν .
- Figure 22.** Influence of the number of basis functions (N_B) and the kernel width parameter (controlled by τ) on classification accuracy for (a) the *All-RBF_c* classifier and (b) the *PCA-RBF_c* classifier.
- Figure 23.** Influence of the number of hidden nodes (N_H) on classification accuracy for (a) the *All-BY_c* classifier and (b) the *PCA-BY_c* classifier.
- Figure 24.** Evolution of the ratio (r_w) between the number of training samples and the number of adaptive parameters in a BY network architecture using all the input features and a BY network architecture using the components selected from PCA.
- Figure 25.** Results from ARD for the *All-BY_c* classification algorithm. The logarithm (base 10) of α_k is represented.
- Figure 26.** Influence of the regularisation parameter (C) and the kernel width parameter (σ_s) on classification accuracy for (a) the *All-SVM_c* classifier and (b) the *PCA-SVM_c* classifier.
- Figure 27.** Number of classification algorithms that provided an incorrect decision for each of the subjects in the test set.
- Figure 28.** Influence of the spread parameter (controlled by τ) on the ICC for (a) the *All-GRNN_r* and (b) the *PCA-GRNN_r* regression algorithms.
- Figure 29.** Influence of the number of hidden nodes (N_H) and the regularisation parameter (ν) on the ICC for (a) the *All-MLP_r* and (b) the *PCA-MLP_r* regression algorithms.

Figure 30. Effect of weight decay regularisation on the size of the weights: (a) *All-MLP_r* regression algorithms with $\nu = 50$ and different N_H ; (b) *All-MLP_r* regression algorithms with $N_H = 40$ and different ν .

Figure 31. Effect of weight decay regularisation on the size of the weights: (a) *PCA-MLP_r* regression algorithms with $\nu = 15$ and different N_H ; (b) *PCA-MLP_r* regression algorithms with $N_H = 22$ and different ν .

Figure 32. Influence of the number of basis functions (N_B) and the kernel width parameter (controlled by τ) on the *ICC* for (a) the *All-RBF_r* and (b) the *PCA-RBF_r* regression algorithms.

Figure 33. Influence of the number of hidden nodes (N_H) on the *ICC* for (a) the *All-BY_r* and (b) the *PCA-BY_r* regression algorithms.

Figure 34. Results from ARD for the *All-BY_r* regression algorithm. The logarithm (base 10) of α_k was represented.

Figure 35. Influence of the maximum number of basis functions (M_{max}) and the allowable level of interaction between variables (mi) on the *ICC* for (a) the *All-MARS_r* and (b) the *PCA-MARS_r* regression algorithms.

Figure 36. Influence of the regularisation parameter (C) and the kernel width parameter (σ_s) on the *ICC* for (a) the *All-SVM_r* and (b) the *PCA-SVM_r* regression algorithms.

Figure 37. Bland-Altman and regression plots for the *All-MLP_r* ((a) and (b)), the *PCA-MLP_r* ((c) and (d)), the *PCA-BY_r* ((e) and (f)) and the *All-MARS_r* regression algorithms.

Figure 38. Bland-Altman (a) and regression (b) plots for the linear model built from ODI3.

List of Tables

Table 1. Clinical and demographic features for the available database of subjects as well as for training and test sets. BMI: body mass index; AHI: apnoea-hypopnoea index.

Table 2. Structure of a confusion matrix for a two-class classification problem.

Table 3. Mean values of the time-domain and frequency-domain features extracted from oximetry recordings in the training set.

Table 4. Optimum decision threshold determined from the training set for each of the extracted features.

Table 5. Classification results individually achieved by each of the extracted features. *Tn*: true negatives; *Fp*: false positives; *Tp*: true positives; *Fn*: false negatives; *Se*: sensitivity; *Sp*: specificity; *Acc*: accuracy; *AUC*: area under the ROC curve; κ : Cohen's kappa coefficient.

Table 6. Coefficients of the linear regression models developed for each of the input features.

Table 7. Regression results individually achieved by each of the extracted features. *Tn*: true negatives; *Fp*: false positives; *Tp*: true positives; *Fn*: false negatives; *Se*: sensitivity; *Sp*: specificity; *Acc*: accuracy; *ICC*: intra-class correlation coefficient; *E_{RMS}*: root mean square error; *E_{rel}* (-/+): relative error per subject (SAHS-negative group/SAHS-positive group).

Table 8. Coefficients of the linear equation for the *All-FLD_c* algorithm.

Table 9. Coefficients of the linear equation for the *PCA-FLD_c* algorithm.

Table 10. Coefficients of the equation derived for the *All-LR_c* algorithm.

Table 11. Coefficients of the equation derived for the *PCA-LR_c* algorithm.

Table 12. Results achieved on the test set by the 16 classification algorithms implemented in the study. *Tn*: true negatives; *Fp*: false positives; *Tp*: true positives; *Fn*: false negatives; *Se*: sensitivity; *Sp*: specificity; *Acc*: accuracy; *AUC*: area under the ROC curve; κ : Cohen's kappa coefficient.

Table 13. Subjects in the test set misclassified by the *All-MLP_c* algorithm. AHI: apnoea-hypopnoea index; *y*: output value provided by the *All-MLP_c* algorithm.

Table 14. Coefficients of the linear equation of the *All-MLR_r* algorithm.

Table 15. Coefficients of the linear equation for the *PCA-MLR_r* algorithm.

Table 16. ANOVA function decomposition for the *All-MARS_r* algorithm. The standard deviation of each function reflects its influence on the output of the algorithm.

Table 17. Results achieved on the test set by the 14 regression algorithms implemented in the study. *Tn*: true negatives; *Fp*: false positives; *Tp*: true positives; *Fn*: false negatives; *Se*:

sensitivity; Sp : specificity; Acc : accuracy; ICC : intra-class correlation coefficient; E_{RMS} : root mean square error; E_{rel} (-/+): relative error per subject (SAHS-negative group/SAHS-positive group).

Table 18. Data of the subjects in the test set misdiagnosed by $PCA-BY_r$ and $All-MARS_r$ algorithms. $PCA-BY_r$: regression algorithm based on BY networks and components retained from PCA; $All-MARS_r$: regression algorithm based on MARS and all the features from oximetry data; AHI: apnoea-hypopnoea index.

Table 19. Mean values of the conventional oximetry indices computed from oximetry recordings in the training set.

Table 20. Optimum decision threshold determined from the training set for each of the oximetry indices.

Table 21. Classification results individually achieved by each of the oximetry indices. Tn : true negatives; Fp : false positives; Tp : true positives; Fn : false negatives; Se : sensitivity; Sp : specificity; Acc : accuracy; AUC : area under the ROC curve; κ : Cohen's kappa coefficient.

Table 22. Coefficients of the linear regression models developed for each of the oximetry indices.

Table 23. Regression results for each of the oximetry indices. Tn : true negatives; Fp : false positives; Tp : true positives; Fn : false negatives; Se : sensitivity; Sp : specificity; Acc : accuracy; ICC : intra-class correlation coefficient; E_{RMS} : root mean square error; E_{rel} (-/+): relative error per subject (SAHS-negative group/SAHS-positive group).

Table 24. Summary of the results reported in some previous studies based on the use of conventional oximetry indices to assist in SAHS diagnosis.

Table 25. Summary of some of the recent studies focused on SAHS diagnosis using signal processing techniques.

Table 26. Classification and regression algorithms with the highest diagnostic accuracy.

1. Introduction

- 1.1. Context: biomedical signal processing
- 1.2. Problem formulation
- 1.3. Hypothesis
- 1.4. Aims of the Doctoral Thesis
- 1.5. Work plan
- 1.6. Principles of pattern recognition
 - 1.6.1. The pattern classification problem
 - 1.6.2. Regression analysis
 - 1.6.3. Learning and generalization
 - 1.6.4. The bias-variance trade-off
 - 1.6.5. Regularisation
 - 1.6.6. Maximum likelihood and Bayesian inference
- 1.7. Structure of the document

1.1. Context: biomedical signal processing

The study presented in this Doctoral Thesis falls into the biomedical engineering field. Biomedical engineering can be defined as the development and application of engineering principles and techniques to provide effective solutions to problems in biology and medicine (Nebeker, 2002). It is a multidisciplinary domain that requires the combination of knowledge from medicine, biology, physics, mathematics or engineering. Biomedical engineering covers a wide range of activities such as biomedical image and signal processing, biomechanics, prosthetic devices and artificial organs, biomaterials, biotechnology, tissue engineering, neural engineering, biomedical instrumentation, bionanotechnology, physiological modelling, rehabilitation engineering, bioinformatics and genomics, clinical engineering or biosensors (Enderle, 2005). Specifically, the Thesis is focused on the biomedical signal processing area. It is the term used to refer the manipulation of a signal for the purpose of either extracting information from it, assessing the relationships of two (or more) signals or producing an alternative representation of the signal (Bruce, 2001).

Biomedical signals are produced by biological systems and structures. Hence, there is a wide variety of recordings. Considering their origin, they can be classified into different groups such as bioelectric signals (i.e., electroencephalogram, electrocardiogram, electromyogram and electrooculogram); biomagnetic signals (i.e., magnetoencephalogram and magnetocardiogram); bioacoustic signals (i.e., phonocardiogram and ecography); biomechanical signals (i.e., mechanomyogram and aerial flux pressure); biochemical signals (i.e., partial pressures of oxygen and carbon dioxide); biooptical signals (i.e., blood oxygen saturation); etc. In addition, biomedical signals can also be divided according to their field of application (i.e., neurology and cardiology) and their characteristics (i.e., continuous and discrete, stationary and non-stationary signals) (Cohen, 2006). Signal processing enables to automate the measurements of signal properties, resulting in reduced subjectivity and increased reliability. It is also of key importance when uncovering components with low amplitude and/or subtle variations in frequency that are very difficult to observe by visual inspection. Additionally, it enables to remove undesired components of technical and physiological origin so that the signal-to-noise ratio is improved and analysis is facilitated (Laguna and Sornmo 2009). Pattern recognition represents a specific area

into the signal processing field. Pattern recognition systems aims to provide appropriate decisions about the category of an object or a specific variable associated to it. Depending on the application, these objects can be images, signal waveforms or any type of measurements that need to be analysed. The generic term pattern is used to refer these objects. Pattern recognition techniques are used for classification and regression tasks. They have been extensively applied to different problems. Computer-aided diagnosis is one of the most popular applications of these techniques in biomedical research (Bishop, 1995; Haykin, 1999; Theodoridis and Koutroumbas 2009).

Signal processing is commonly used to extract relevant information from biomedical recordings that cannot be evaluated by visual inspection. The objective is often to discriminate abnormal from normal signals and to diagnose the presence of a disease related to the system generating them. In this Thesis, blood oxygen saturation (SaO_2) signals recorded through pulse oximetry were analysed. These recordings measure the level of oxygenation in arterial blood (Mendelson, 1992). Due to its non-invasive nature and the convenient patient interface consisting of a small sensor clipped or tapped onto the skin, pulse oximetry is commonly used in respiratory medicine (Bloch, 2003). Indeed, oximetry is widely applied in anaesthesia, surgery, critical care, hypoxaemia screening, exercise, transport from the operating room to the recovery room and emergency (Mendelson, 1992). In sleep medicine, nocturnal oximetry signals have been suggested as a screening tool to identify patients with sleep-disordered breathing (Netzer *et al.* 2001). In the present research, it is proposed to apply signal processing techniques to analyse SaO_2 recordings corresponding to subjects suspected of suffering from sleep apnoea-hypopnoea syndrome (SAHS). It is a common respiratory disorder with a severe impact on health. However, its diagnosis involves some difficulties and simpler procedures are desirable.

In this Thesis, pattern recognition methods are applied to oximetry data in order to assist in SAHS diagnosis. Most of the clinical diagnosis problems can be modelled using pattern recognition techniques. Their use in medical decision making has substantially increased in the last years (Lisboa, 2002). These methods may be of benefit for clinical practice. Pattern recognition techniques enable to simultaneously process a high quantity of data. Moreover, some of them are capable of modelling non-linear relationships that may arise in practical applications (Baxt, 1995). Two kinds of

pattern recognition tasks can be considered: classification and regression. Both schemes are proposed to model SAHS diagnosis.

1.2. Problem formulation

Nowadays, nocturnal polysomnography (PSG) is required to obtain a reliable diagnosis about SAHS. PSG is performed in a special sleep unit and enables the detection of apnoeas, hypopnoeas and arousals. It includes the acquisition of several recordings, such as electroencephalogram (EEG), electrocardiogram (ECG) or SaO₂, in addition to a range of other variables required for the complete description of sleep (Qureshi and Ballard 2003). The PSG study is supervised by qualified personnel (technicians and physicians) able to perform, monitor and interpret the examination (Gagnadoux *et al.* 2002). Once the test has been completed, a medical expert must manually analyse all the physiological recordings to provide a definitive diagnosis. Every disordered-breathing event is identified from the recorded data. According to the American Academy of Sleep Medicine, an apnoea is considered as a cessation of airflow for at least 10 seconds (Kushida *et al.* 2005). On the other hand, a hypopnoea event is considered as a reduction in airflow of at least 50% that is associated with an oxygen desaturation greater than 4% (Kushida *et al.* 2005). The frequency of apnoeas and hypopnoeas per hour of sleep is expressed as the apnoea-hypopnoea index (AHI), which is used to characterise the severity of SAHS. Usually, an AHI between 5 h⁻¹ and 15 h⁻¹ is used to determine a positive diagnosis (Caples *et al.* 2005).

Despite its diagnostic reliability, PSG presents some drawbacks. The sleep laboratory is an artificial environment and some patients present a disturbed sleep pattern due to the foreign setting. PSG is complex and time-consuming due to the great quantity of information that must be processed by the sleep specialist. In addition, it is an expensive procedure since specific equipment and personnel are required (Bennet and Kinnear 1999). Finally, the reception capacities of currently existing sleep laboratories are overwhelmed. The demand for this type of studies is progressively growing due to a better understanding of the disease and its associated risks. In contrast, the infrastructure to support it seems to be insufficient (Flemons *et al.* 2004). It is suspected that a high percentage of patients suffering from SAHS remain undiagnosed. The study by Young *et al.* (1997) found that 82% of men and 93% of women with

moderate to severe SAHS were not diagnosed. Early treatment is required in order to prevent other health complications. Flemons *et al.* (2004) evaluated the capacity of the sleep studies of five developed countries: United Kingdom, Belgium, Australia, United States and Canada. The average waiting time from the referral to a specialist to the beginning of treatment was estimated. It ranged from 2 months in Belgium up to 7-60 months in the United Kingdom. The number of beds prepared to perform PSG studies per 100000 inhabitants was also computed. It was 0.3 in the United Kingdom while the rate was around 1.4 in the other countries. Recently, a similar study analysed the delays in the diagnosis of SAHS in Spain (Masa *et al.* 2007). It concluded that the average time comprised between the sleep specialist and the beginning of treatment was approximately 10 months while the number of available beds for PSG per 100000 inhabitants was 0.49.

The mentioned statistics reveal the need for improving the protocol for SAHS diagnosis. Increasing the capacity for PSG studies by developing the existing sleep laboratories or creating new ones is required. However, it seems to be slow and an important investment on resources is needed. Other alternatives based on technological advances can be found. For instance, home unattended and hospital telemonitored PSG have been proposed (Gagnadoux *et al.* 2002). However, both are still complex, time-consuming and expensive. Research work related to the development of simplified diagnostic techniques has increased in the last years. The aim is to provide an accurate tool for SAHS detection while avoiding the difficulties of PSG. In particular, the usefulness of nocturnal oximetry for SAHS screening has been widely assessed. Oximetry recordings reflect respiratory dynamics during sleep. Pulse oximetry has the advantages of being readily available, relatively inexpensive and can be performed at home (Bennet and Kinnear 1999). Therefore, SaO₂ analysis has been suggested as a suitable approach to perform SAHS diagnosis and to reduce the need for PSG studies (Chiner *et al.* 1999; Caples *et al.* 2005).

1.3. Hypothesis

Pulse oximetry is an essential tool in sleep medicine to determine a patient's cardiorespiratory stability. Respiratory disturbances such as apnoeas and hypopnoeas are reflected in the SaO₂ profile. The saturation value decreases due to reduction of

airflow. Subsequently, it returns to its baseline level when respiration is restored. Therefore, SaO_2 recordings from SAHS-positive patients are characterised by marked fluctuations due to desaturation events, which reflect unstable ventilation (Bloch, 2003). In contrast, healthy respiratory patterns tend to present a near-constant SaO_2 waveform around 96% (Netzer *et al.* 2001). To illustrate this, the SaO_2 signals corresponding to a control subject ($\text{AHI} = 1.4 \text{ h}^{-1}$), a doubtful control subject ($\text{AHI} = 8.1 \text{ h}^{-1}$), a doubtful SAHS-positive patient ($\text{AHI} = 13.1 \text{ h}^{-1}$) and a SAHS-positive patient ($\text{AHI} = 66.2 \text{ h}^{-1}$) are depicted in Figure 1. As can be observed, a more unstable behaviour is reflected in patients with a higher AHI value. However, it is difficult to discriminate signals from doubtful subjects.

Traditionally, analysis of nocturnal oximetry was focused on visual inspection of tracings for identification of desaturation events, baseline and minimal

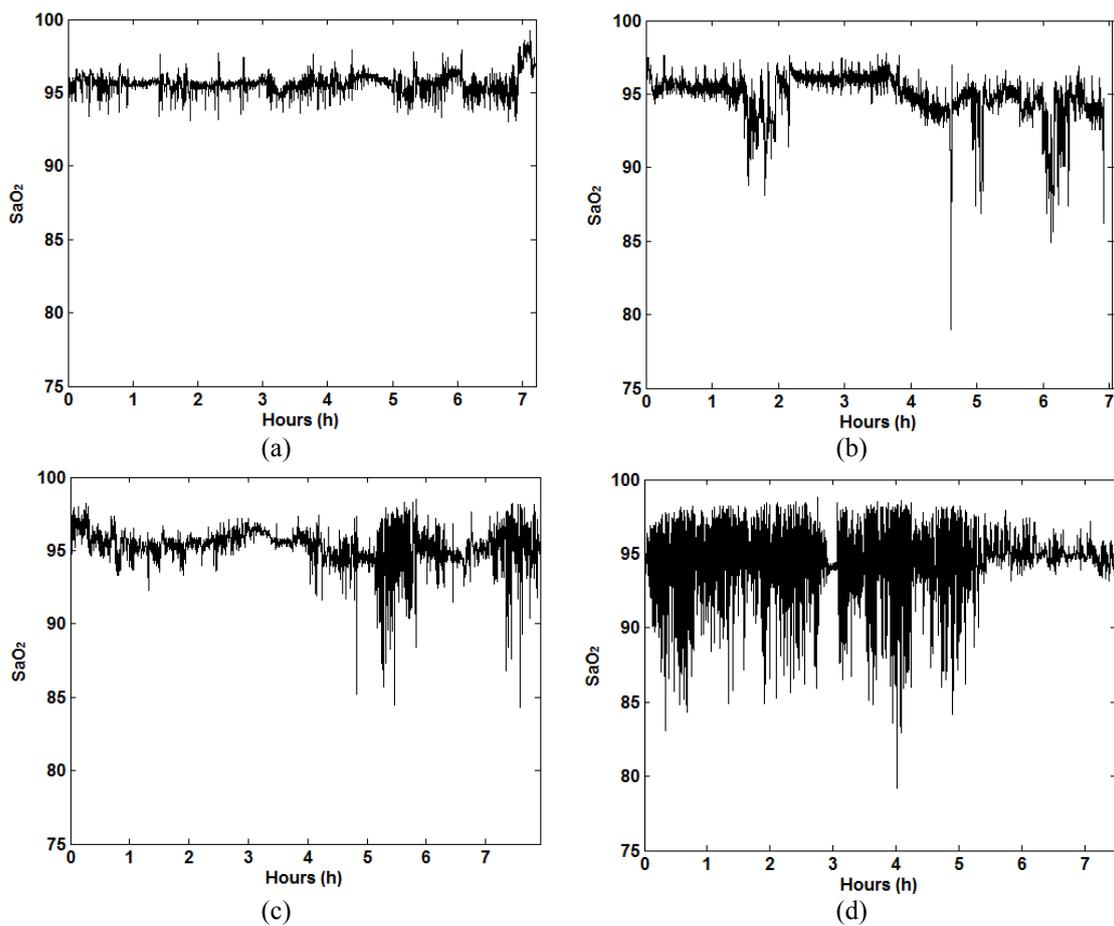


Figure 1. Examples of SaO_2 signals corresponding to (a) a SAHS-negative subject ($\text{AHI} = 1.4 \text{ h}^{-1}$), (b) a doubtful SAHS-negative subject ($\text{AHI} = 8.1 \text{ h}^{-1}$), (c) a doubtful SAHS-positive subject ($\text{AHI} = 13.1 \text{ h}^{-1}$) and (d) a SAHS-positive subject ($\text{AHI} = 66.2 \text{ h}^{-1}$).

values, and patterns of fluctuation in oxygen saturation (Bloch, 2003). However, it is tedious and subjective. A manual score must be provided for recordings extending over several hours and interpretation may differ among experts. Indeed, there is no universally accepted definition of an oxygen desaturation in sleep-disordered breathing (a decrease greater than 4% is usually applied) (Netzer *et al.* 2001). Several desaturation indices were proposed to perform automated and objective analysis of oxymetry signals. They reflect the variability of oxygen saturation. The most common indices are the oxygen desaturation index over 3% (ODI3) and 4% (ODI4) and the cumulative time spent below a given level of saturation, which is usually fixed at 90% (CT90) (Bloch, 2003). Their diagnostic utility has been evaluated in previous studies (Netzer *et al.* 2001). The reported results show a substantial difference between sensitivity and specificity. The former varied from 88 to 98% and the latter from 40 to 88%, when an AHI of 10 h^{-1} or 15 h^{-1} measured during PSG was defined as the upper limit of normal (Bloch, 2003). Any of these indices can be pointed out as a reliable measurement to perform SAHS diagnosis.

In this Doctoral Thesis, it is hypothesised that the diagnostic capability of oximetry recordings can be enhanced by using advanced signal processing techniques. According to the dynamic behaviour of SaO_2 signals, time and frequency analysis may be appropriate to identify features reflecting the presence of SAHS and its degree of severity. It is proposed to apply pattern recognition techniques to combine these features in order to achieve high diagnostic accuracy. In this context, two different approaches can be used to model the SAHS diagnosis problem. The first one corresponds to a classification scheme. In classification, the output of the model must provide a diagnostic decision about the signal under analysis. The second approach is based on regression techniques to estimate the value of the AHI from oximetry data. Both strategies are considered in this Thesis in order to provide an effective solution for SAHS diagnosis.

1.4. Aims of the Doctoral Thesis

The lack of general availability of PSG causes a relevant number of SAHS-positive subjects to remain undiagnosed. These patients are at risk for other severe health complications regarding cardiovascular and cerebrovascular systems. On the

other hand, a high economic cost is related to untreated SAHS. On average, SAHS-positive patients use more health care resources before their diagnosis compared with matched control subjects. These issues make SAHS an important social concern (Qureshi and Ballard 2003).

A growing number of research studies for the development of alternative diagnostic techniques has been performed during recent years. A wide variety of data have been used for this purpose. Several diagnostic models were developed from clinical and demographic variables such as the assessment of daytime sleepiness, age, gender, body mass index (BMI), neck circumference or morphological features of the upper airway (Caballero *et al.* 1998; El-Solh *et al.* 1999; Kirby *et al.* 1999). Additionally, personal questionnaires were used together with these features (Nettleton and Muñiz 2001). On the other hand, signal processing techniques have been applied to different biomedical recordings to assist in SAHS diagnosis. In particular, ECG signals have been extensively analysed (Penzel *et al.* 2002; de Chazal *et al.* 2003; Roche *et al.* 2003; de Chazal *et al.* 2004; Shouldice *et al.* 2004; Khandoker *et al.* 2009a; Khandoker *et al.* 2009b). Other signals used for this purpose include airflow (Várady *et al.* 2002, Al-Ani *et al.* 2004), thoracic effort (Macey *et al.* 1998), snoring (Fiz *et al.* 2010), videotape recordings (Sivan *et al.* 1996) or speech (Zigel *et al.* 2008). Moreover, the diagnostic utility of SaO₂ signals has been evaluated in numerous studies (Vázquez *et al.* 2000; Netzer *et al.* 2001; Roche *et al.* 2002; Magalang *et al.* 2003). Conventional oximetry indices represent the most common procedure for automated interpretation of SaO₂ recordings. Indeed, some of the commercial oximeters include the computation of these indices (Bloch, 2003). Signal processing techniques have been also used to characterise SaO₂ recordings. Spectral analysis of oximetry data revealed that signals corresponding to SAHS-positive subjects tend to present a higher percentage of the signal power in the band between 0.010 and 0.033 Hz (Zamarrón *et al.* 2003). Similarly, non-linear methods for signal analysis showed to be useful tools to study the dynamic properties of SaO₂ recordings in time domain. It has been proved that irregularity, variability and complexity of these signals measured by approximate entropy (*ApEn*), central tendency measure (*CTM*) and Lempel-Ziv complexity (*LZC*) are related to the presence of SAHS (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Hornero *et al.* 2007). In order to achieve high diagnostic ability, time-domain and frequency-domain features from SaO₂ data can be combined using pattern recognition techniques. This

methodology reported significant results in the SAHS diagnosis problem. For the pattern recognition stage, traditional statistical procedures such as discriminant analysis or logistic regression have been evaluated (Marcos *et al.* 2009; Marcos *et al.* 2010b). Other more advanced methods based on neural networks were also applied (Marcos *et al.* 2008a; Marcos *et al.* 2008b; Marcos *et al.* 2010a).

Pattern recognition techniques have been pointed out as a powerful tool to assist in clinical decision-making (Baxt, 1995). They are used to combine information from several features in classification and regression tasks. Pattern recognition techniques have been successfully applied to different biomedical engineering problems such as the detection of brain tumours (Preul *et al.* 1996), epilepsy events (Acir *et al.* 2005), Alzheimer's disease (Lehman *et al.* 2007), arrhythmias (Ozbay *et al.* 2006) or diabetic retinopathy (García *et al.* 2009). Specifically, SAHS diagnosis has been modelled by using pattern recognition techniques from clinical data and questionnaires (El-Solh *et al.* 1999; Kirby *et al.* 1999; Roche *et al.* 2002), ECG recordings (de Chazal *et al.* 2004; Khandoker *et al.* 2009a; Khandoker *et al.* 2009b) and snoring signals (Fiz *et al.* 2010). In comparison with pulse oximetry, the acquisition process used to collect this kind of data may require a great effort from clinicians (collecting data and questionnaires for every patient) or may result uncomfortable for patient's sleep.

The aim of this research is to develop automated methods based on signal processing techniques to assist in SAHS diagnosis from SaO₂ recordings. The diagnosis problem is modelled as a pattern recognition task. Multivariate data analysis techniques for classification and regression were used for this purpose. The classification approach consists in assigning SaO₂ signals to one of two possible categories: SAHS-positive or SAHS-negative. Regression analysis is applied to estimate the AHI from oximetry data. Time and spectral features from SaO₂ signals are proposed to define patterns used for classification and regression. In order to achieve these major objectives, the following partial objectives must be fulfilled:

1. To develop a database of SaO₂ recordings corresponding to subjects suspected of suffering from SAHS.

2. To research into the state of the art of pattern recognition techniques and their application to clinical diagnosis.
3. To design the methodology to model SAHS diagnosis from oximetry data using pattern recognition techniques.
4. To write the Matlab® code for the methods selected for feature extraction, normalisation and dimensionality reduction, and pattern recognition.
5. To apply the selected methods to SaO₂ signals from the database.
6. To quantify the diagnostic ability of the evaluated methods by analysing the statistical properties of their results.
7. To interpret the obtained results and identify the most effective methods for the proposed problem.
8. To publish the main findings of the research in international journals with a relevant impact factor.

1.5. Work plan

In order to achieve the proposed objectives, the following work plan has been designed:

1. Recruitment of control subjects and subjects affected by SAHS. The population available for the Thesis is composed of subjects suspected of suffering from SAHS. All the people gave their consent to participate in the study. Subjects involved in this research were referred to a PSG test at the sleep unit of the Hospital del Rio Hortega, Valladolid, Spain. Dr. Del Campo and his team, from Division of Pneumology, were the specialists in charge of analysing PSG data and providing a definitive diagnosis for each patient.

2. Recording of nocturnal SaO₂ signals. Oximetry signals were recorded during PSG tests together with other physiological data. The equipment used was a polygraph (Alice 5, Respironics, Philips Healthcare, The Netherlands). A finger probe interface was used for the acquisition of pulse oximetry recordings. Saturation values were recorded at a sampling frequency of 1 Hz.
3. Development of the database of SaO₂ recordings. The available signals were included in a database composed of control subjects and patients affected by SAHS. Each signal was preprocessed to remove artefacts due to a bad connection of the probe. In addition, the AHI value computed from PSG was included for each patient. Other clinical information in the database were age, BMI or gender. The available set of SaO₂ signals was randomly divided into a training set and a test set with 40% and 60% of the recordings, respectively. The former was used for optimising the proposed pattern recognition methods. Subsequently, the resulting trained algorithms were validated by using signals in the test set.
4. Feature extraction. Information in oximetry signals was summarised into a feature vector (pattern) in order to be analysed using pattern recognition techniques. The dynamic behaviour of SaO₂ signals can be reflected by evaluating appropriate properties. According to the proposed classification and regression approaches, the computed features are expected to provide useful information about the presence of SAHS and its degree of severity. Time-domain and frequency-domain analysis were applied to characterise SaO₂ signals.
5. Normalisation and dimensionality reduction. Linear re-scaling of the extracted features was applied to avoid possible differences between their magnitudes. Additionally, dimensionality reduction techniques were used as a previous step to pattern recognition. The improvement derived from the use of principal component analysis (PCA) was evaluated.

6. Bibliographic search about pattern recognition methods. An exhaustive analysis of the latest studies about the use of pattern recognition techniques on biomedical signals was performed. Classification and regression were considered to model the SAHS diagnosis problem. Some of the most popular and capable techniques were selected for both tasks, including conventional and complex procedures.
7. Software programming. The methods for feature extraction, normalisation and dimensionality reduction, and pattern recognition were implemented using Matlab® (version 7, R14, The Mathworks Inc.). It must be noted that some of the required algorithms are included in toolboxes developed by other researchers or could be a part of Matlab®.
8. Design of the pattern recognition methods. Most of the pattern recognition techniques evaluated in the Thesis require adjusting some design or smoothing parameters. A model selection process must be carried out to compare the performance of different configurations, i.e. models with a different value of these parameters. For this purpose, leave-one-out cross-validation was applied by only using data in the training set.
9. Processing SaO₂ recordings. The designed methods were applied to oximetry signals to obtain the final results. The processing stage was carried out in two phases:
 - Training. A pattern recognition algorithm was implemented from signals (patterns) in the training set using the design parameters determined from model selection.
 - Testing. The trained algorithm was applied to signals in the test set. An output value (response) was obtained for each of these signals.
10. Statistical analysis of the results. The output responses computed from signals in the test set were analysed to assess the performance of each pattern recognition method. The following measures were obtained:

- Sensitivity, specificity and accuracy in SAHS diagnosis were computed for classification and regression methods evaluated in the Thesis.
 - Receiving operating characteristic (ROC) analysis was performed for classification methods. The area under the ROC curve (*AUC*) was computed as a measure of classification performance.
 - The correlation between the predicted and actual AHI was computed to evaluate the performance of regression techniques.
11. Interpreting the results and drawing conclusions. The methods evaluated in the Thesis were compared. The goal was to identify the best classification and regression algorithms for the proposed problem. The main conclusions from these results were drawn. Finally, the results of the Thesis were compared with those reported in other previous studies.
 12. Publication of the results. Some of the preceding results of the Thesis were published in international journals related to the biomedical engineering field. In addition, national and international conferences were considered. Additional relevant findings will be considered for publication.

1.6. Principles of pattern recognition

Problems such as face and handwritten character recognition, fault diagnosis or clinical diagnosis are often carried out by humans. Pattern recognition aims to clarify the mechanisms and automate the decision-making process for this kind of problems (Bishop, 1995). It studies how machines can observe the environment, learn to distinguish patterns of interest from their background and make sound and reasonable decisions about the categories of the patterns (Jain, 2000). This Thesis analyses the application of pattern recognition techniques in a clinical diagnosis context. As indicated before, two different approaches are proposed to model the SAHS diagnosis problem: classification and regression.

The pattern recognition problem is typically defined from a finite set of samples D composed of N input-output pairs $\{(\mathbf{x}^n, \mathbf{t}^n)\}_{n=1, \dots, N}$, where $\mathbf{x}^n = (x_1^n, x_2^n, \dots, x_d^n)$ denotes an input pattern with d real-valued measurements or features and $\mathbf{t}^n = (t_1^n, t_2^n, \dots, t_c^n)$ is its corresponding target value. In a general form, the target can be viewed as a c -dimensional variable. From a statistical point of view, the most general and complete description of the data in D is given by the joint probability density $p(\mathbf{x}, \mathbf{t})$:

$$p(\mathbf{x}, \mathbf{t}) = p(\mathbf{t}|\mathbf{x})p(\mathbf{x}), \quad (1)$$

where $p(\mathbf{t}|\mathbf{x})$ denotes the probability density of \mathbf{t} given \mathbf{x} and $p(\mathbf{x})$ represents the unconditional density of \mathbf{x} . In order to make appropriate predictions of \mathbf{t} given the value of \mathbf{x} , it is the conditional density $p(\mathbf{t}|\mathbf{x})$ which need to be modelled by the output of the algorithm. Therefore, a common approach for implementing a pattern recognition algorithm is based on the estimation of $p(\mathbf{t}|\mathbf{x})$ from data in D . The functional $p(\mathbf{t}|\mathbf{x})$ can be used to obtain optimum solutions for classification and regression problems. In addition, other modelling strategies can be adopted. In the following points of this section, some topics related to the pattern recognition field are addressed.

1.6.1. The pattern classification problem

A classification problem occurs when an object needs to be assigned to one of c categories ($\omega_1, \omega_2, \dots, \omega_c$). The task of the classifier is to map the input pattern \mathbf{x} into a categorical variable \mathbf{t} that indicates the class label (Kulkarni *et al.* 1998; Jain, 2000):

$$f: \mathcal{R}^d \rightarrow \mathbf{t} \in \{\omega_1, \dots, \omega_c\}, \quad (2)$$

where f denotes the mapping function. The output of the classifier can be expressed as $\mathbf{y} = f(\mathbf{x}, \mathbf{w})$, where the dependence of the mapping on the input pattern \mathbf{x} and a set of adaptive parameters \mathbf{w} has been made explicit.

Statistical modelling assumes that patterns \mathbf{x} belonging to class ω_i have been randomly generated from the class-conditional probability function $p(\mathbf{x}|\omega_i)$ (Jain, 2000). In this context, the Bayes decision rule can be applied to minimise the risk of the prediction (Bishop, 1995; Haykin, 1999; Jain, 2000). It states that an input pattern \mathbf{x} must be assigned to class ω_i for which the conditional risk $R(\omega_i|\mathbf{x})$ is minimum:

$$R(\omega_i|\mathbf{x}) = \sum_{j=1}^c L(\omega_i, \omega_j) p(\omega_j|\mathbf{x}), \quad (3)$$

where $L(\omega_i, \omega_j)$ is the loss incurring in deciding ω_i when the true class is ω_j and $p(\omega_j|\mathbf{x})$ is the posterior probability. Usually, a 0/1 loss function is used (Jain, 2000):

$$L(\omega_i, \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}. \quad (4)$$

This function leads to a specific form of the Bayes' decision rule, known as the maximum a posteriori (MAP) rule. An input pattern \mathbf{x} is assigned to class ω_i if the following condition is satisfied:

$$p(\omega_i|\mathbf{x}) > p(\omega_j|\mathbf{x}) \quad \text{for all } j \neq i. \quad (5)$$

There exist several methods to perform pattern classification (Holmström *et al.* 1997; Jain, 2000). Those evaluated in the Thesis were the Fisher's linear discriminant (FLD), the K -nearest neighbour (KNN) rule, logistic regression (LR), probabilistic neural networks (PNN), multilayer perceptron (MLP) neural networks, Bayesian (BY) networks, radial basis function (RBF) neural networks and support vector machines (SVM). Most of them are based on the described statistical approach and use the Bayes' rule to provide a decision. In this context, the aim is to obtain accurate approximations for posterior probabilities $p(\omega_i|\mathbf{x})$ in order to minimise the risk of the prediction. It can be directly obtained (e.g. LR, MLP neural networks, BY networks and RBF neural networks) or through the estimation of the class-conditional probability $p(\mathbf{x}|\omega_i)$ (e.g. KNN or PNN) (Holmström *et al.* 1997). The latter approach makes use of the Bayes' theorem (Bishop, 1995), which shows that the posterior probability is given by:

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}, \quad (6)$$

where $P(\omega_i)$ is the prior probability associated to class ω_i . The term $p(\mathbf{x})$ vanishes when this expression is used to evaluate the Bayes' decision rule.

On the other hand, a different strategy for implementing pattern classifiers is based on directly determining optimum decision boundaries from samples in D rather than modelling the statistical distribution of the data (e.g. FLD and SVM). Usually, input patterns are projected into a new transformed space to facilitate the separation among different classes.

1.6.2. Regression analysis

The regression problem consists in adequately approximating a function of several variables given only the value of the function, often perturbed by noise, at various points in the dependent variable space (Friedman, 1991). The goal is to model a response continuous variable \mathbf{t} as a function of a set of independent variables or features grouped in the input pattern $\mathbf{x} = (x_1, x_2, \dots, x_d)$. For simplicity, one-dimensional functions will be considered.

In regression problems, the system that generated targets is assumed to be given by:

$$t = h(\mathbf{x}) + \varepsilon, \quad (7)$$

where \mathbf{x} is known, $h(\cdot)$ is the true underlying function (deterministic function) and ε is an additive stochastic component (noise) that reflects the dependence of t on quantities that are neither controlled nor observed (Friedman, 1991). A zero-mean Gaussian distribution is assumed for this variable:

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right). \quad (8)$$

Therefore, the distribution of target data t given the input pattern \mathbf{x} correspond to a Gaussian function of the form:

$$p(t|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[y(\mathbf{x}, \mathbf{w}) - t]^2}{2\sigma^2}\right\}, \quad (9)$$

where expressions in Eq. (7) and (8) have been combined and h has been replaced by the output $y = f(\mathbf{x}, \mathbf{w})$ of the regression model. The vector \mathbf{w} is used to denote the set of adaptive parameters in the model. The dependence of the output y on \mathbf{x} and \mathbf{w} has been indicated by using the form $y(\mathbf{x}, \mathbf{w})$. To assess the goodness of the approximation, the distance between h and $y(\mathbf{x}, \mathbf{w})$ is estimated from D by using some error measure. As it will be shown later, adjusting model parameters \mathbf{w} in order to maximise the probability of observing the original data set D (maximum likelihood criterion) leads to the sum-of-squares error function (Bishop, 1995). Thus, the regression model $y(\mathbf{x}, \mathbf{w})$ is chosen in order to minimise the following expression:

$$E_D = \frac{1}{2} \sum_{n=1}^N [y(\mathbf{x}^n, \mathbf{w}) - t^n]^2, \quad (10)$$

i.e. $y(\mathbf{x}, \mathbf{w})$ represents the least-squares solution to the problem of fitting data in D .

To interpret the output of the regression model, consider the limit in which the size N of the data set goes to infinity. The finite sum over patterns in the sum-of-squares error can be replaced by the following integral:

$$E_D = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=1}^N [y(\mathbf{x}^n, \mathbf{w}) - t^n]^2 = \frac{1}{2} \iint [y(\mathbf{x}, \mathbf{w}) - t]^2 p(t, \mathbf{x}) dt d\mathbf{x}, \quad (11)$$

where the extra factor of $1/N$ has been introduced in order to make the limiting process meaningful. The joint distribution $p(t, \mathbf{x})$ can be expressed as the product of the unconditional density function for the input data $p(\mathbf{x})$ and the target data density conditioned to the input vector $p(t|\mathbf{x})$ to give:

$$E_D = \frac{1}{2} \iint [y(\mathbf{x}, \mathbf{w}) - t]^2 p(t|\mathbf{x}) p(\mathbf{x}) dt d\mathbf{x}. \quad (12)$$

The following conditional averages of the target data are defined:

$$\langle t|\mathbf{x} \rangle \equiv \int t p(t|\mathbf{x}) dt \quad (13)$$

and

$$\langle t^2|\mathbf{x} \rangle \equiv \int t^2 p(t|\mathbf{x}) dt. \quad (14)$$

The first factor of the integrand in Eq. (12) can be written as:

$$(y-t)^2 = (y - \langle t|\mathbf{x} \rangle + \langle t|\mathbf{x} \rangle - t)^2 = (y - \langle t|\mathbf{x} \rangle)^2 + 2(y - \langle t|\mathbf{x} \rangle)(\langle t|\mathbf{x} \rangle - t) + (\langle t|\mathbf{x} \rangle - t)^2. \quad (15)$$

The integrand in Eq. (12) is replaced by this expression, leading to the following form for the sum-of-squares error:

$$E_D = \frac{1}{2} \int [y(\mathbf{x}, \mathbf{w}) - \langle t|\mathbf{x} \rangle]^2 p(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \int [\langle t^2|\mathbf{x} \rangle - \langle t|\mathbf{x} \rangle^2] p(\mathbf{x}) d\mathbf{x}. \quad (16)$$

The second term in Eq. (16) is independent of the model $y(\mathbf{x}, \mathbf{w})$. For the purpose of determining the parameters of the model by error minimisation, this term can be neglected. Since the integrand in the first term is non-negative, the absolute minimum of the error function occurs when this first term vanishes. It corresponds to the following result for the output of the model (Bishop, 1995):

$$y(\mathbf{x}, \mathbf{w}^*) = \langle t|\mathbf{x} \rangle, \quad (17)$$

where \mathbf{w}^* denotes the set of parameters that minimises the error. This result indicates that using a sum-of-squares error for the optimisation of the model, its output approximates the conditional average of the target data. It is known as the regression of t conditioned on \mathbf{x} . It is worth noting that any prior assumption about the architecture of the model was made. The expression of the regression function shows that if we could make several measurements of t for a given \mathbf{x} , we would obtain a set of random values

whose average (in the limit of an infinite sample) defines the value of the function h at point \mathbf{x} (Bishop, 1995). This is illustrated in Figure 2.

Different methods can be used for regression analysis. Some of them have a flexible architecture that allows for classification and regression modelling. For instance, most of the neural network models can adapt to both classification and regression tasks by modifying certain characteristics of their architecture. The difference is motivated by the nature of the variable to be approximated. It represents a probability value in a classification context while the output of regression models is an estimation of a continuous variable. As a result, optimisation principles differ from one approach to another. In this study, the methods evaluated for regression analysis were multiple linear regression (MLR), generalised regression neural networks (GRNN), MLP networks, BY networks, RBF networks, multivariate adaptive regression splines (MARS) and SVM.

Typically, the methods used for regression modelling assume a previous form of the functional dependence between the target variable and the input features. It depends on a set of unknown adaptive parameters that are fitted to the data in D . As explained, optimisation is performed by minimising an error measure that usually corresponds with the sum-of-squares error. This strategy is followed by MLR, MLP networks, BY networks, RBF networks, MARS and SVM. A different modelling scheme consists in using a density estimation approach to solve the regression problem,

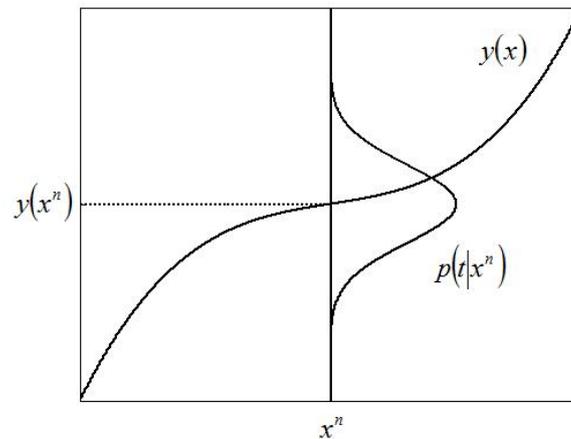


Figure 2. Graphical representation of Eq. (17), which indicates that a model optimised by minimising the sum-of-squares error approximates the conditional average of the target data. A one-dimensional function with a single independent variable is depicted (Bishop, 1995).

as some of the classification techniques previously cited. For example, GRNN are based on this principle (Bishop, 1995). An approximation to the probability density function $p(t|\mathbf{x})$ is computed. Then, the regression function is obtained by applying the expectation operator as indicated in its definition.

As can be observed, classification and regression represent different tasks. In this Thesis, both approaches are used to model SAHS diagnosis. However, a different problem definition is required for each of them. Pattern classifiers were applied to label SaO₂ signals as SAHS-positive or SAHS-negative, which corresponds to a two-class classification problem. On the other hand, regression techniques were used to approximate the value of the AHI as a function of the features extracted from SaO₂ signals. A wide variety of pattern recognition techniques can be used for classification and regression tasks. A subset of them was selected to be evaluated in this study.

1.6.3. Learning and generalisation

Pattern recognition models can be viewed as a mapping function that relates an input feature pattern with an output variable. For simplicity, a one-dimensional output will be considered. The form of the function is initially unknown and some examples are required to find a suitable relationship. Therefore, the mapping function $y = f(\mathbf{x}, \mathbf{w})$ is modelled in terms of some mathematical expression which contains a number of adjustable parameters (\mathbf{w}). Their value is determined on the basis of the input-output pairs in D . This process is called learning or training, and for this reason the data set of examples is referred to as a training set (Bishop, 1995). Training is usually performed by minimising an error function computed from the distance between the actual model output (y^n) and the desired output (t^n), given a training sample \mathbf{x}^n . In this case, it is termed as supervised learning, since for each input pattern the value of the desired output is specified. In contrast, unsupervised learning techniques do not require target data for training (Bishop, 1995; Haykin, 1999).

The goal of training is to build a statistical model of the process that generated the data rather than to learn an exact representation of the training samples. This is required to achieve a good generalisation capability. Generalisation refers to the ability to make good predictions for new inputs (Bishop, 1995). It is influenced by three

factors: the size of the training set and how representative is of the environment of interest, the number of adjustable parameters in the model and the physical complexity of the problem (Haykin, 1999; Jain, 2000). The two first factors are closely related whereas problem complexity cannot be controlled. The complexity of the model is linked to the number of adjustable parameters to be optimised during training. Theoretically, it should agree with problem complexity, i.e. more complex models are required to solve harder problems. The problem of finding the optimum model complexity must be addressed in the design of pattern recognition algorithms. A model with a small number of adjustable parameters, i.e. low flexibility, may be insufficient to achieve good generalisation. Conversely, a model with an excessive number of adjustable parameters may capture the noise present in the data, resulting in poor generalisation. These conditions are known as underfitting and overfitting, respectively (Bishop, 1995; Pardo and Sverbeglieri 2004). Underfitting could be avoided by increasing the flexibility of the model. As a result, a better fit of the training data would be achieved. Preventing overfitting requires the training set to grow accordingly to model complexity (Bebis and Georgiopoulos 1994; Haykin, 1999). The model is constrained by a more accurate specification of the problem. However, this cannot be usually done since the training set is fixed in most of the real applications. Therefore, the most suitable degree of complexity for the underlying problem must be determined.

1.6.4. The bias-variance trade-off

The problem of optimising model complexity leads to the bias-variance trade-off principle. A too simple or inflexible model will have a large bias and may lead to underfitting. In contrast, models with a high variance present a high flexibility regarding the size of the training set and could be affected by overfitting. Both are complementary quantities and the best generalisation is obtained when the best compromise between the conflicting requirements of small bias and small variance is achieved (Bishop, 1995).

The mathematical description of these concepts can be obtained by considering the regression model expressed in Eq. (7). Model parameters are adjusted in order to minimise the sum-of-squares error function (E_D) given in Eq. (10). As pointed

out previously, the minimum error (optimum mapping) is achieved when the model output represents the conditional average $\langle t|\mathbf{x} \rangle$ (Bishop, 1995; Haykin, 1999).

The error term E_D depends on the particular data set (D) on which it is optimised. The dependence can be eliminated by considering an average over the complete ensemble of data sets, which is written as:

$$E_T \left\{ \left[t - y(\mathbf{x}, \mathbf{w}) \right]^2 \right\}, \quad (18)$$

where $E_T\{\cdot\}$ denotes the expectation, or ensemble average, over all the possible training sets. The argument of this operator can be expressed as:

$$t - y(\mathbf{x}, \mathbf{w}) = (t - \langle t|\mathbf{x} \rangle) + [\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w})] = \varepsilon + [\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w})], \quad (19)$$

where ε is the noise term of the regression model. By substituting this expression into Eq. (18) and then expanding terms, the average error function is given in the following form (Haykin, 1999):

$$\begin{aligned} & E_T \left\{ \left[\varepsilon + (\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w})) \right]^2 \right\} = \\ & = E_T \{ \varepsilon^2 \} + E_T \left\{ \left[\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w}) \right]^2 \right\} + 2E_T \left\{ \varepsilon \left[\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w}) \right] \right\}, \end{aligned} \quad (20)$$

where the last expectation term in the right-hand side of the equation is zero since the error ε is uncorrelated with functions $\langle t|\mathbf{x} \rangle$ and $y(\mathbf{x}, \mathbf{w})$ (Haykin, 1999). As a result, the average error function is given by:

$$E_T \left\{ \left[\varepsilon + (\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w})) \right]^2 \right\} = E_T \{ \varepsilon^2 \} + E_T \left\{ \left[\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w}) \right]^2 \right\}. \quad (21)$$

The first term is the variance of the expected error ε , evaluated over the training sample D . This term represents the intrinsic error because it is independent of the parameter vector \mathbf{w} . Thus, it may be ignored to analyse the minimisation of E_D with respect to \mathbf{w} . The second term represents the natural measure of the distance between

the regression function $\langle t|\mathbf{x} \rangle$ and the approximating function $y(\mathbf{x}, \mathbf{w})$. It can be viewed as the average value of the estimation error between both functions, evaluated over the entire training sample D . This term can be expressed as (Bishop, 1995):

$$\begin{aligned} & E_T \left\{ \left[\langle t|\mathbf{x} \rangle - y(\mathbf{x}, \mathbf{w}) \right]^2 \right\} = \\ & = E_T \left\{ \left[E_T \{ y(\mathbf{x}, \mathbf{w}) \} - \langle t|\mathbf{x} \rangle \right]^2 \right\} + E_T \left\{ \left[E_T \{ y(\mathbf{x}, \mathbf{w}) \} - y(\mathbf{x}, \mathbf{w}) \right]^2 \right\} = \\ & = B^2(\mathbf{w}) + V(\mathbf{w}), \end{aligned} \quad (22)$$

where the first term in the right-hand side of the equation represents the square of the bias and the second term is referred to as the variance.

The meaning of both quantities is clearly identified from their mathematical representation. The bias measures the extent to which the average (over all data sets) of the mapping function implemented by the model differs from the desired function. Conversely, the variance measures the extent to which the mapping function $y(\mathbf{x}, \mathbf{w})$ is sensitive to the particular choice of data set (Bishop, 1995).

1.6.5. Regularisation

High generalisation capability involves finding the optimum balance between bias and variance. As indicated, underfitting and overfitting are related to models with high bias and variance, respectively. One way to simultaneously reduce both quantities is to increase the number of data points. As a result, more complex models could be applied, leading to reduced bias. At the same time, the implemented models would be more heavily constrained by the data, thereby also reducing variance. In order to do this, it is required that the size of the training set grow rapidly in comparison with model complexity. In this case, the model would be a consistent estimator of the regression function since it could be approximated with arbitrary accuracy as the number of data points goes to infinity (Bishop, 1995).

However, this situation cannot be achieved in practice. As explained before, the training set is usually fixed and the problem becomes to find the optimum model complexity. This optimisation problem provides an example of Occam's razor (Bishop,

1995; Haykin, 1999). This principle states that simpler models should be preferred to more complex ones. This preference should be traded off against the extent to which the models fit the data. A certain degree of flexibility is required to capture the properties of the problem to be solved but a smoothing mapping should be obtained to prevent high variance and overfitting. This effect is related to irregular mapping functions that reflect the noise on the training data, resulting in poor generalisation. Overfitting is a common problem when designing a pattern recognition model (Bishop, 1995). Regularisation techniques encourage smooth network mappings to avoid this problem and improve generalisation capability. The aim is to control the effective complexity of the model, i.e. to reduce the number of adaptive parameters to be adjusted during training.

There exist several approaches to implement regularisation. Usually, the error function is modified by adding a new term that penalises complex mappings. Weight decay regularisation corresponds to this approach. Other alternatives to control model complexity are early stopping, weight sharing and pruning algorithms (Bishop, 1995). In addition, it has been proposed to assess the generalisation performance of a model by only using data in the training set. A generalised prediction error (GPE) measure was developed by Moody (1992). It is applicable to both linear and non-linear models and is composed of the sum of two terms: the training error and a complexity term. Simple models will give a large value of GPE since the training error is high. A large value of GPE is also obtained by complex models since the complexity term is high. Thus, the minimum value of the GPE criterion represents a trade-off between these two competing effects (Bishop, 1995).

1.6.6. Maximum likelihood and Bayesian inference

The aim of training is to model the statistical properties of the generator of data in D , which is completely described by the joint probability density function $p(\mathbf{x}, \mathbf{t})$. However, for the purpose of making predictions of \mathbf{t} for new patterns \mathbf{x} , it is the conditional density $p(\mathbf{t}|\mathbf{x})$ which must be modelled. Usually, a predefined architecture is used to model this distribution. It includes a set of adjustable parameters (\mathbf{w}) that must be defined to obtain the definitive model. This dependence can be made explicit by expressing $p(\mathbf{t}|\mathbf{x}, \mathbf{w})$. Adjustable parameters in the model are optimised during learning in order to provide a good estimation of the density. A training set D composed of N

examples is available for this purpose. Two learning approaches can be considered: maximum likelihood and Bayesian inference.

Maximum likelihood seeks to find the optimum value of \mathbf{w} by maximising a likelihood function derived from the training data (Bishop, 1995). Considering that samples in D were drawn independently from the distribution $p(\mathbf{x}, \mathbf{t})$, the joint probability density of observing the whole data set D is given by:

$$p(D|\mathbf{w}) = \prod_{n=1}^N p(\mathbf{x}^n, \mathbf{t}^n | \mathbf{w}) = \prod_{n=1}^N p(\mathbf{t}^n | \mathbf{x}^n, \mathbf{w}) p(\mathbf{x}^n) \equiv L_D(\mathbf{w}), \quad (23)$$

where $L_D(\mathbf{w})$ is a function of \mathbf{w} given a set D and is referred to as the likelihood. The technique of maximum likelihood sets the value of \mathbf{w} by maximising $L_D(\mathbf{w})$. This corresponds to the idea of choosing the \mathbf{w} which is most likely to give rise to the observed data (Bishop, 1995). In practice, the negative logarithm of the likelihood is considered:

$$E_D = -\ln L_D(\mathbf{w}) = -\sum_{n=1}^N \ln p(\mathbf{t}^n | \mathbf{x}^n, \mathbf{w}) - \sum_{n=1}^N \ln p(\mathbf{x}^n), \quad (24)$$

where E_D , the error function, is minimised through the optimisation process. As indicated, the pattern recognition method will be used to model $p(\mathbf{t}|\mathbf{x})$. Thus, the second term in the equation does not depend on \mathbf{w} and can be dropped from the error function, which is then given by:

$$E_D = -\sum_{n=1}^N \ln p(\mathbf{t}^n | \mathbf{x}^n, \mathbf{w}). \quad (25)$$

However, the model which minimises the error in D is not guaranteed to achieve the best generalisation ability. In the Bayesian approach, the parameters \mathbf{w} are described by a probability distribution, which represents the uncertainty about their value. Before training data are observed, the parameters of the model are described by a prior probability density $p(\mathbf{w})$. Typically, a broad function is used to represent the lack of knowledge about their appropriate initial value. Once training data D is observed, the

Bayes' theorem can be used to find the corresponding posterior probability density (Bishop, 1995):

$$p(\mathbf{w}|D) = \frac{p(D|\mathbf{w})p(\mathbf{w})}{p(D)}, \quad (26)$$

where $p(D|\mathbf{w})$ is precisely the likelihood function introduced in Eq. (23) and $p(\mathbf{w}|D)$ is referred to as the posterior distribution for \mathbf{w} . Since some values of \mathbf{w} are more consistent with D , the posterior distribution is narrower than the prior distribution. This phenomenon, known as Bayesian learning, is illustrated in Figure 3.

Instead of choosing a specific value for \mathbf{w} , the Bayesian approach uses the posterior distribution $p(\mathbf{w}|D)$ to compute the distribution of the output for a new input pattern \mathbf{x} . It is given by the weighted average over all values of \mathbf{w} , as indicated in the following expression (Bishop, 1995):

$$p(\mathbf{t}|\mathbf{x}, D) = \int p(\mathbf{t}|\mathbf{x}, \mathbf{w})p(\mathbf{w}|D)d\mathbf{w}. \quad (27)$$

1.7. Structure of the document

In addition to some key concepts, hypothesis, objective and methodology of the Thesis have been introduced. In this point, the organisation of the document is described. It is divided into nine chapters, including the present chapter that serves as

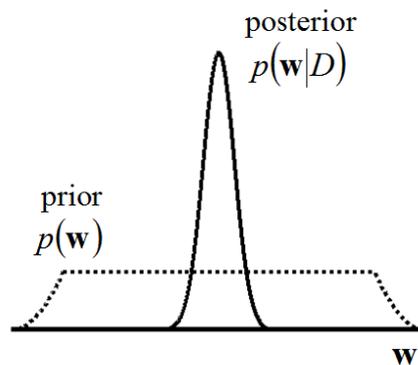


Figure 3. Effect of Bayesian inference. Distribution of adaptive parameters before and after the training data in D is observed. As shown in the picture, a narrower form is obtained for the posterior distribution indicating that some values of \mathbf{w} are preferred once training data is available (Bishop, 1995).

introduction. Their content is detailed in the following paragraphs:

- Chapter 2. The sleep apnoea-hypopnoea syndrome. A detailed description of SAHS is given in this chapter. Firstly, the definition of the disease and a complete explanation of the physiological phenomena associated to SAHS are provided. Afterwards, symptoms, epidemiology and consequences are addressed. Subsequently, the standard procedure for SAHS diagnosis is described, explaining its associated problems. Nocturnal pulse oximetry is presented as a useful technique for simplifying SAHS diagnosis. Finally, the treatment options for SAHS are commented.
- Chapter 3. State of the art. An exhaustive analysis of other proposed approaches for SAHS diagnosis from SaO_2 data is detailed. In addition, those methods based on the analysis of other different recordings are included.
- Chapter 4. Subjects and signals. This chapter describes the population under study. The acquisition process and properties of SaO_2 signals are explained. Finally, the details about training and test sets are explained, providing the statistics of subjects in both sets.
- Chapter 5. Feature extraction, normalisation and dimensionality reduction. The fifth chapter deals with the feature extraction process. Time-domain and frequency-domain analyses are proposed to characterise SaO_2 signals in order to reflect the presence of SAHS and its degree of severity. The reasons motivating the use of the selected methods are explained. Subsequently, the transformation (normalisation and dimensionality reduction) of the extracted features before the pattern analysis stage is described. An explanation about PCA is provided.
- Chapter 6. Pattern recognition for classification. Pattern recognition techniques used for classification of SaO_2 signals are described in this chapter. A wide set of methods have been applied to model SAHS diagnosis

as a classification problem. The following ones are included: KNN, FLD, LR, PNN, MLP networks, RBF networks, BY networks and SVM.

- Chapter 7. Pattern recognition for regression. It has been proposed to model SAHS diagnosis as a regression problem to estimate the AHI from SaO₂ data. This chapter is focused on the description of the methods used for this task. These include MLR, GRNN, MLP networks, RBF networks, BY networks, MARS and SVM. Some of these methods were also cited among the techniques used for classification. The differences between both modelling approaches are explained.
- Chapter 8. Results. The main results achieved in the study are presented in this chapter. It follows a similar scheme to that of the methodology proposed in the Thesis. Initially, the performance measures used to evaluate classification and regression techniques are presented. A brief analysis about the features extracted from oximetry data is provided. Their individual capability for solving the proposed classification and regression problems is evaluated. Subsequently, the design (model selection) process for each method is described. It is shown how the smoothing parameters were defined. Finally, the results obtained on the test set by the selected models are given.
- Chapter 9. Discussion and conclusions. The results presented in the previous chapter are exhaustively analysed. The main findings of the study are highlighted. In addition, a comparison with other previous methods is carried out. The main contributions of the Thesis are identified as well as future work for this study.

2. The sleep apnoea-hypopnoea syndrome

- 2.1. Introduction
- 2.2. Physiopathology
- 2.3. Clinical presentation
- 2.4. Epidemiology and clinical consequences
- 2.5. Diagnosis
- 2.6. Diagnostic utility of nocturnal pulse oximetry
 - 2.6.1. Principles of pulse oximetry
 - 2.6.2. Clinical applications
 - 2.6.3. Potentials and limitations of nocturnal pulse oximetry
- 2.7. Treatment

2.1. Introduction

The sleep apnoea-hypopnoea syndrome was independently described by two European research groups in 1965. Gastaut, Tassinari and Duron (1965) in France as well as Jung and Kuhlo (1965) in Germany reported their findings of the Pickwickian syndrome of sleep apnoea. The term Pickwickian syndrome was adopted by Burwell *et al.* (1956) from the somnolent fat boy described in the Pickwick papers by Charles Dickens (Kryger, 1985; Van Houwelingen *et al.* 1999). However, they evaluated their somnolent obese patients only while awake. In 1970, Tassinari and Lugaresi (1970) performed clinical sleep investigations in a large series of patients. They described sleep apnoea syndrome in obese and non-obese patients, showed a correlation with cardiovascular diseases and identified snoring and hypersomnolence as diagnostic indicators. Guilleminault *et al.* (1973) introduced the first definition for apnoea and AHI in 1973.

SAHS is characterised by periodic complete or partial upper airway obstruction during sleep, causing intermittent cessations of breathing (apnoeas) or reductions in airflow (hypopnoeas) despite ongoing respiratory effort (Qureshi and Ballard 2003). An apnoea entails a cessation of airflow for 10 or more seconds. It usually indicates complete obstruction of the upper airway (Kushida *et al.* 2005). Hypopnoea is related to a reduction in airflow greater than 30% for at least 10 seconds associated with a 4% decrease in oxygen saturation (Kushida *et al.* 2005). It connotes a transient reduction in inspiratory airflow caused by increased upper airway resistance. Three different types of apnoeas can be identified (Qureshi and Ballard 2003):

- An apnoea is considered obstructive if there is continued respiratory effort despite cessation of airflow.
- In the case of central apnoeas, there is no concurrent respiratory effort. They are caused by a failure of the central nervous system respiratory centers.
- The term mixed apnoea is used to define those events composed of obstructive and central apnoeas within the same episode.

The AHI characterises the severity of SAHS and is used to determine whether a patient is affected by this condition. There is no consensus about the decision threshold to be applied on AHI, varying between 5 h^{-1} and 15 h^{-1} (Van Houwelingen *et al.* 1999; Caples *et al.* 2005). Commonly, the upper limit for SAHS-negative subjects is established at $\text{AHI} = 10 \text{ h}^{-1}$ (Bloch, 2003).

2.2. Physiopathology

Upper airway obstruction characterising apnoeas and hypopnoeas typically lies in the pharynx. The structure of the upper airway imparts an inherent tendency to collapsibility. It is demonstrated by the observation that sleep onset in healthy people is associated with increased airway resistance and radiographic evidence of the upper airway (Caples *et al.* 2005). In healthy awake human subjects, pharyngeal patency is maintained by continual neuromuscular activation of the pharyngeal muscles by the central nervous system. During normal sleep, various protective mechanisms maintain partial patency of the upper airway. More than 20 muscles play a role in airway dilation and wall stiffening (Qureshi and Ballard 2003; Caples *et al.* 2005). However, muscle activation is typically reduced during sleep, compromising the patency of the upper airway. Additionally, this activation can be attenuated by alcohol, sleep deprivation, anaesthesia and sedative-hypnotics (Qureshi and Ballard 2003). The combination of reduced neural activation with anatomic abnormalities of the pharynx, such as excess posterior pharyngeal tissue, an enlarged tongue or a low-hanging palate, sets the stage for obstructive apnoeas and hypopnoeas in patients with SAHS. Imaging and endoscopic studies have shown that during wakefulness and sleep, SAHS-positive patients have a smaller upper airway lumen compared with controls (Caples *et al.* 2005).

Patients affected by SAHS suffer numerous episodes of apnoea/hypopnoea during sleep. Each of them begins with total occlusion or critical narrowing of the upper airway, which either fully eliminates or substantially reduces ventilation. This in turn allows the development of hypercarpnia and hypoxia. Chemoreflexes mediate the ventilatory response to these phenomena. Peripheral chemoreceptors, the most important of which are located in the internal carotid arteries, primarily respond to blood oxygen tension. Brainstem central chemoreceptors are most sensitive to carbon

dioxide and acid-base balance (Caples *et al.* 2005). In some people, hypoxemic stimulation of the peripheral chemoreceptors simultaneously increases sympathetic output to muscles while activation of cardiac vagal activity results in bradycardia, which is known as the diving reflex (Caples *et al.* 2005). It is characterised by peripheral vasoconstriction, which preserves blood flow to the brain and heart vessels, and profound bradycardia as a means of limiting cardiac oxygen demand. Respiratory effort progressively increases in response to these chemostimuli and eventually triggers an arousal. This leads to a surge of pharyngeal dilator muscle activity and resolution of the upper airway obstruction. This cycle can repeat itself continuously during the night, allowing intermittent hypoxia and hypercarpnia, fragmenting sleep and triggering adrenergic surges (Qureshi and Ballard 2003). Figure 4 depicts a scheme of the physiological phenomena associated to apnoeas.

2.3. Clinical presentation

There is a broad spectrum of clinical features related to SAHS. The cardinal symptom of SAHS is excessive daytime sleepiness resulted from sleep fragmentation. It is difficult to be quantified by physicians because patients may use varied adjectives to describe sleepiness and they may confuse sleepiness with fatigue. Subjective tools, such as the Epworth Sleepiness Scale (Johns, 1991), are quick and simple to complete but do not correlate well with the severity of the AHI (Caples *et al.* 2005). Another prominent

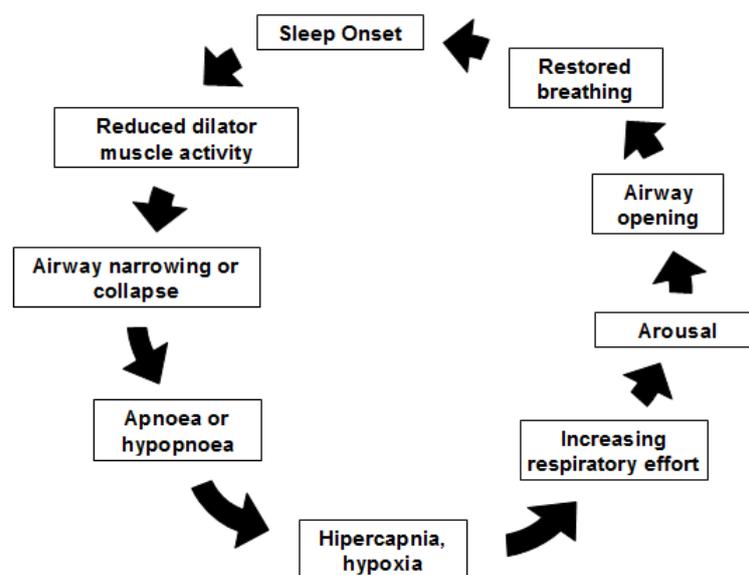


Figure 4. Scheme of the physiological cycle associated to apnoea/hypopnoea events.

symptom is snoring. Some snorers are not aware of their problem or its severity. An interview with the partner or other relatives is essential to obtain an accurate description. A pattern of intermittent loud snoring, with periods of silence lasting more than 10 seconds, is suggestive of the occurrence of sleep apnoea (Van Houwelingen *et al.* 1999). Data from the Sleep Heart Health Study suggest that snoring is associated with daytime sleepiness independently of the severity of the AHI (Gottlieb *et al.* 2000).

Other possible symptoms include nocturnal arousals with or without choking spells, nocturnal diaphoresis, abnormal motor activity during sleep, gastroesophageal reflux, headaches, automatic behaviour, chest pain, diminished libido, impotence, loss of memory and concentration, personality changes, fatigue and depression (Van Houwelingen *et al.* 1999; McNicholas, 2008). Some features evaluated by physical examination are suggestive of SAHS. Patients affected by this condition are frequently overweight and hypertensive. Neck circumference, which is believed to be indicative of central obesity, is often increased in patients with SAHS (Qureshi and Ballard 2003). Examination of the oral cavity may reveal abnormalities such as enlarged tonsils, thickened uvula or a long and redundant soft palate. In patients with severe SAHS, cardiovascular symptoms are also common (Van Houwelingen *et al.* 1999).

2.4. Epidemiology and clinical consequences

Different studies have evaluated the prevalence of SAHS in order to anticipate health care needs and allocate appropriate resources. It is estimated that 1 of 5 adults with an average BMI of 25 to 28 kg/m² has an AHI ≥ 5 h⁻¹ and 1 of 15 has an AHI ≥ 15 h⁻¹ (Durán *et al.* 2001; Young *et al.* 1993; Young *et al.* 2002). Specific epidemiologic studies about SAHS have reported an incidence up to 5% of adults in western countries (Young *et al.* 2002). Some of the factors associated with progression of disease are baseline obesity, older age and the presence of snoring (Caples *et al.* 2005). In addition, the association between AHI and sex, waist-hip ratio and cholesterol concentration has been pointed out (Tishler *et al.* 2003). Potentially modifiable risk factors that influence the development of SAHS are smoking and alcohol. Menopause is an additional risk factor for women (Young *et al.* 2002).

Progression studies show the time evolution of the prevalence. An 8-year follow-up study of 282 participants revealed that the overall mean AHI increased from 2.5 to 5.1 h⁻¹ during this period (Young *et al.* 2002). Significant increases of AHI were found in all strata of sex, BMI, age and snoring. Whereas the change in mean AHI was not considerably greater in men compared with women, progression was significantly greater in obese compared with non-obese, older compared with younger and habitually snoring compared with non-habitually snoring subjects. Consistent with the elevation in AHI, there were more people whose AHI increased than decreased. In addition, many people who had an AHI of 1 h⁻¹ or less increased to an AHI greater than 1 h⁻¹ during the 8-year period, whereas few opposite cases were found (Young *et al.* 2002).

Diverse consequences may be derived from SAHS. Epidemiological data support the concept that SAHS can participate in the initiation or progression of several diseases mainly related to cardiovascular and cerebrovascular systems (Durán *et al.* 2001; Bradley and Floras 2009). Indeed, there is an association of sleep apnoea with morbidity and mortality due to these diseases (Van Houwelingen *et al.* 1999). Even mild SAHS (i.e. $5 \text{ h}^{-1} \leq \text{AHI} \leq 15 \text{ h}^{-1}$) is associated with significant morbidity. Undiagnosed patients, with or without symptoms, are related to increased likelihood of cardiovascular conditions such as hypertension, ischemic heart disease, arrhythmias congestive cardiac failure or stroke, pulmonary disorders such as pulmonary hypertension, asthma or chronic obstructive pulmonary disease, cerebrovascular disease, damage of the neurocognitive function and upper airway inflammation (Young *et al.* 2002; Qureshi and Ballard 2003). Moreover, SAHS is associated to motor vehicle and industrial accidents as well as diminished quality of life (Young *et al.* 2002). Habitually sleepy drivers have a higher incidence of respiratory disturbance during sleep and significantly more automobile accidents if their AHI is of greater than 15 h⁻¹ (Qureshi and Ballard 2003).

2.5. Diagnosis

SAHS diagnosis is based on the combined assessment of clinical features together with objective sleep study findings. Nocturnal PSG is required for the latter (McNicholas, 2008). Nowadays, PSG is considered the gold-standard for the diagnosis of SAHS. Several physiological signals and data are monitored during PSG studies

including EEG, ECG, electrooculogram (EOG), electromyogram (EMG), SaO₂, oral and nasal airflow, respiratory effort, body position, snore microphone and leg movement (Qureshi and Ballard 2003; McNicholas, 2008). PSG provides detailed information on sleep state and respiratory and gas exchange abnormalities. Manual scoring of the recorded data performed by a specialist is required to interpret the PSG study, which enables the detection of apnoeas and hypopnoeas. Finally, the value of the AHI is derived from the test. It is used to determine the severity of SAHS. Figure 5 shows a PSG study being carried out in a special sleep unit.

PSG enables exhaustive assessment of sleep architecture, identifying stages and arousals. However, there can be considerable interobserver variability in the scoring of arousals. The respiratory variables in PSG studies are reproducible on a night-to-night basis in moderate to severe cases of SAHS. Mild cases can demonstrate a significant variability in the AHI, leading to false-negative cases (McNicholas, 2008). This variability can relate to a number of different factors, particularly time spent in the supine position where AHI is typically higher. Thus, the inclusion of body position is required to perform a complete analysis of sleep. Other factors that may contribute to night-to-night variability of the AHI include alcohol and drug consumption in addition to variability in sleep stage distribution, particularly REM sleep (McNicholas, 2008).

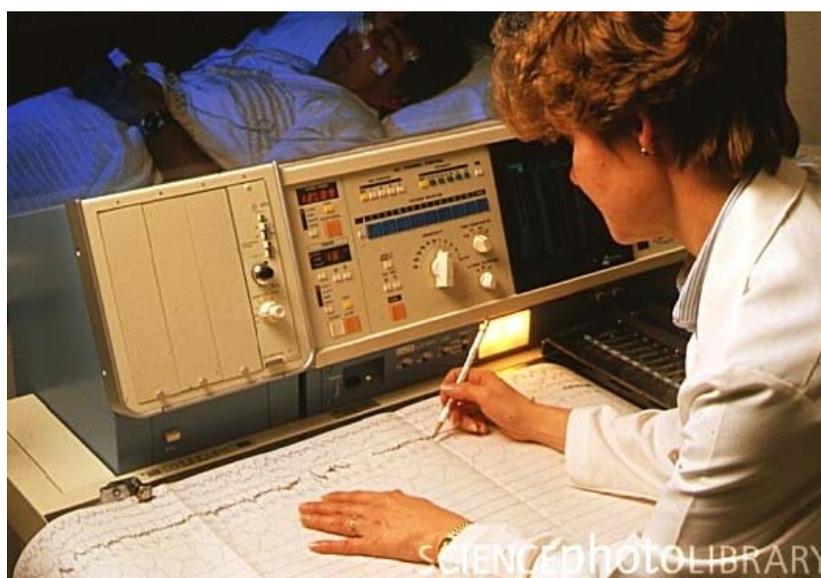


Figure 5. A doctor observes the recordings from a patient during a PSG test in a sleep laboratory (Science Photo Library ©).

PSG studies are resource intensive since they require to be performed in a full sleep laboratory and a trained technician must be in charge of the test (McNicholas, 2008). Thus, long waiting lists for diagnostic sleep studies are common (Flemons *et al.* 2004). Moreover, the high prevalence of SAHS may require clinicians outside sleep centres to assess patients instead of clinicians with specific training in sleep medicine. Therefore, it is necessary to consider other simplified approaches for SAHS diagnosis (McNicholas, 2008). In this context, nocturnal pulse oximetry represents a potential tool for developing alternative diagnostic methods.

2.6. Diagnostic utility of nocturnal pulse oximetry

2.6.1. Principles of pulse oximetry

Accurate assessment of blood gases is essential in different areas of medicine such as critical care or pulmonary medicine. These measurements assist the clinician in determining a patient's oxygenation status and to protect the patient against dangerous hypoxic conditions. Previously, blood gases were measured by invasive sampling, either through an indwelling arterial catheter or by arterial puncture, and analysed in a clinical laboratory by a blood gas analyser (Mendelson, 1992). This practice presents significant drawbacks as the prolonged delay between sample acquisition and the availability of the laboratory results. In neonatal applications, for example, frequent blood sampling can cause significant blood loss, especially for very small infants. Furthermore, blood gas values are available only intermittently and, therefore, indicate the status of the patient only at the time the blood sample was drawn. The inevitable delay and lack of continuous information can lead to potential diagnostic errors, particularly in critically ill patients. These can be severely affected by rapid and often life-threatening cardiopulmonary changes within a short time. Continuous monitoring of blood gases, on the other hand, offers a major advantage: instantaneous changes in blood gas concentrations can be recognised and adequately corrected before irreversible tissue damage occurs. Furthermore, continuous monitoring also provides information on trends in the patient's condition to assist the clinician in assessing the response to therapeutic procedures (Mendelson, 1992).

Pulse oximetry is a non-invasive technique for monitoring arterial blood saturation. It was developed by Millikan during World War II as a method to assess the

oxygenation of pilots in high altitude flying (Tremper, 1989). Pulse oximetry has become a standard in clinical practice. The most important advantage of pulse oximeters is the capability to provide continuous, safe and effective monitoring of blood oxygenation non-invasively at the patient's bedside. Pulse oximeters are easy to use and versatile, require no user calibration, heating of the skin and changing of electrolytes or membranes and provide virtually maintenance-free operation (Mendelson, 1992; Jubran, 2004). Arterial oxygen saturation provides relevant information about respiratory patterns. In the early years of pulmonary medicine, pulse oximetry was the key means to identify patients with Pickwickian syndrome or severe sleep apnoea by detecting the saw-tooth pattern on SaO_2 waveforms (Netzer *et al.* 2001). Nowadays, the technique is an integral component of PSG and, by itself, is commonly used as a simple tool in the evaluation of sleep apnoea (Bloch, 2003).

Pulse oximeters determine the SaO_2 level by measuring light absorption of arterial blood at two specific wavelengths: 660 nm (red) and 940 nm (infrared) (Jubran, 2004). The ratio of absorbencies at these wavelengths is calibrated empirically against direct measurements of arterial blood oxygen saturation, and the resulting calibration curve is used to generate the pulse oximeter's estimation of arterial saturation (Tremper, 1989). Figure 6 shows the absorption spectra of normal adult haemoglobin in its saturated and desaturated states. The oximeter probe comprises two light emitting

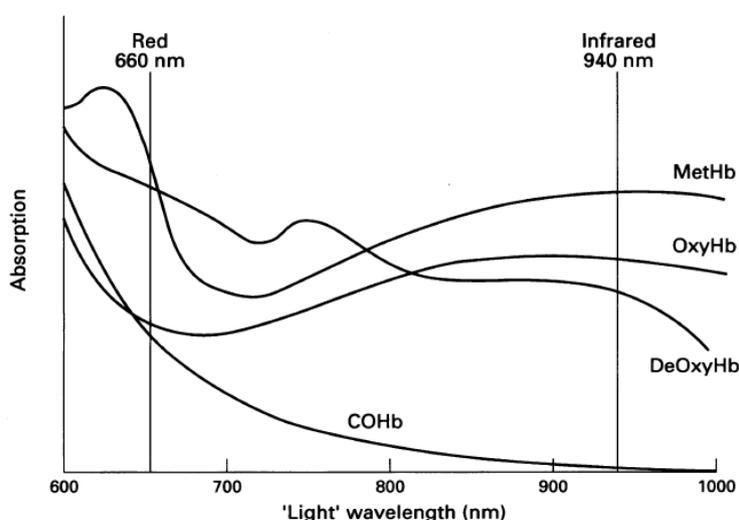


Figure 6. Absorption spectra of normal adult haemoglobin in saturated (OxyHb) and desaturated (DeOxyHb) oxyhaemoglobin states, carboxyhaemoglobin (COHb) and methaemoglobin (MetHb) (Moyle, 1996).

diodes (red and infrared) and a detector. The emitters and detector are placed so as to face each other through tissue about 5-10 mm in thickness. Usually, a finger-probe interface is used to connect the pulse oximetry equipment to the patient (Hanning and Alexander-Williams 1995; Moyle, 1996). The diodes are switched on and off in rapid sequence so that each measurement set include an estimation of transmission of red light, infrared light and ambient light. After correction for ambient light the ratio of red to infrared light is determined. Hundreds of individual measurements are made each second and fed into an algorithm in a microprocessor, which accepts or rejects them. Then, the obtained measurements are weighted using formulas unique to each manufacturer. Finally, the displayed value is an average based on the previous 3-6 seconds of recording and updated every 0.5-1 seconds (Hanning and Alexander-Williams 1995).

2.6.2. Clinical applications

Pulse oximetry enables to detect hypoxaemia, defined as a SaO₂ value less than 90% (Jubran, 2004). Hypoxaemia is commonly found in all aspects of clinical practice and is a major cause of organ dysfunction and death. Continuous oximetry is regarded as essential during anaesthesia and critical care of adults and neonates. Frequent applications of pulse oximetry include preoperative evaluations, the operating room, postanaesthesia recovery suites and stroke units (Netzer *et al.* 2001). It is also found in other acute care areas such as accident units and endoscopy suites. In addition, SaO₂ measurements are usually included in the routine vital signs of temperature, pulse, respiratory rate and blood pressure (Hanning and Alexander-Williams 1995).

In sleep medicine, oximetry recordings contain essential information about SAHS and play a crucial role for interpreting PSG studies. Pulse oximetry enables to track the fluctuations in SaO₂ reflecting abnormal ventilation in patients with SAHS (Bloch, 2003). These tend to present unstable SaO₂ signals due to frequent drops in the saturation value. Apnoeas and hypopnoeas are usually accompanied by marked desaturation events due to the lack of airflow. Subsequently, respiration is restored and saturation recovers its baseline level. A different behaviour tends to be observed in healthy patients. Recordings from these subjects reflect normal ventilation, for which

the saturation value remains near 96% (Netzer *et al.* 2001). As a result, oximetry signals can be used to develop novel diagnostic procedures for SAHS detection (Bloch, 2003).

2.6.3. Potentials and limitations of nocturnal pulse oximetry

The debate about the utility of pulse oximetry for screening patients with sleep-disordered breathing is open. Previous studies suggested that pulse oximetry could be enough for diagnosing moderate-to-severe SAHS because of the characteristic pattern of repetitive desaturation (Cooper *et al.* 1991). Pulse oximetry enables to overcome the difficulties of PSG. It is well suited to ambulatory assessment and could be performed at home, resulting in significant lower cost and resources, as well as reduced waiting lists. Additionally, the inherent complexity of PSG could be avoided by including a single recording for sleep monitoring. The oximetry equipments provide digital signal acquisition systems. Thus, mathematical and statistical techniques can be applied to perform computer-assisted analysis of SaO₂ data (Bloch, 2003). It enables to avoid manual scoring, resulting in a notable reduction in the time required to interpret the sleep study. Therefore, accurate assessment of the diagnostic utility of oximetry signals has become a common subject of study for the scientific community (Netzer *et al.* 2001).

The use of pulse oximetry alone for SAHS detection is limited by several factors. As previously indicated, only PSG can provide information about sleep states and arousals. In addition, oxygen desaturation can be absent with hypopnoeas or in events associated with increased upper airway resistance (McNicholas, 2008). Moreover, it is not possible to differentiate desaturations occurring secondary to obstructive apnoeas, central apnoeas, primary pulmonary disease and cardiac disease using oximetry (Bennet and Kinnear 1999). Finally, SaO₂ could be severely affected by artefacts. Pulse oximetry is vulnerable to the effects of poor peripheral arterial blood flow. Therefore, body movements, vasoconstriction and hypotension can affect SaO₂ readings. In sleep medicine, movement artefacts are common since patients often have fragmented sleep with a lot of body movements. Additionally, changes in the haemoglobin structure and quantity can cause artificially high or low readings that are not due to respiratory disturbances. Tissue optics in very obese patients can result in the same effect. The impact of these limitations can be reduced in in-laboratory sleep

studies since other parameters are also monitored during sleep (Netzer *et al.* 2001). However, they should be considered in the design of diagnostic techniques based on nocturnal oximetry alone.

2.7. Treatment

Once a positive diagnosis of SAHS is confirmed, effective treatment is required to prevent additional complications. The first approach to treat a patient with SAHS is lifestyle modification in order to avoid risk factors. Patients should be encouraged to lose weight, avoid sleep deprivation and refrain from tobacco, alcohol and sedatives (Qureshi and Ballard 2003). A critical weight loss must often occur before a significant AHI reduction is seen. Additionally, positional therapy can be used in a selected group of patients. It consists in having a lateral position during sleep or propping the head at 30 to 60 degrees. This therapy enables to stabilise the pharynx, reducing the overall apnoea frequency (Qureshi and Ballard 2003).

Continuous positive airway pressure (CPAP) remains the most effective therapy for SAHS. It eliminates upper-airway flow limitation in almost any patient and is associated with improvement in daytime symptoms and objective measures of sleepiness in patients with mild or severe abnormalities in the AHI (Caples *et al.* 2005). CPAP treats apnoeas and hypopnoeas by providing air under positive pressure through a nasal or facial mask. It prevents collapse of the pharyngeal airway. The pressure of the CPAP mask depends on the patient under treatment and must be adjusted. The use of CPAP mitigates or reverses many of the acute pathophysiologic responses related to SAHS. Despite its effectiveness, compliance rates for CPAP between 65 to 90% have been reported due to diverse causes such as nasal congestion, mask discomfort, claustrophobia, difficult exhaling or bed-partner intolerance (Qureshi and Ballard 2003).

Other possibilities of treatment include surgery (recommended in patients with craniofacial abnormalities) and oral appliances, which aim to modify the position of the mandible so as to enlarge the upper airway or reduce its collapsibility (Van Houwelingen *et al.* 1999). The role of surgery is controversial. The most appropriate indication is clearly reversible causes of upper-airway obstruction. Uvulopalatopharyngoplasty, which involves the excision of soft palatal tissue, results in

an AHI less than 20 in a minority of patients (Sher *et al.* 1996). This suboptimal performance may be explained in part by the role of anatomic factors other than soft palatal tissue.

3. State of the art

- 3.1. Introduction
- 3.2. Signal processing-based methods to assist in SAHS diagnosis
 - 3.2.1. Automated analysis of nocturnal oximetry to assist in SAHS detection
 - 3.2.2. Other biomedical signals and data related to SAHS
- 3.3. Utility of pattern recognition techniques as an aid tool for SAHS diagnosis

3.1. Introduction

Signal processing and artificial intelligence techniques have been proposed as a reasoning tool to support clinical-decision making since the earliest days of computing (Ledley and Lusted 1960). The theoretical and practical problems encountered have led to important developments in statistics and computer science. However, it was during the last decade of the last century that decision support systems have been routinely used in clinical practice on a significant scale (Lisboa, 2002).

Sleep medicine is an active area in the development of systems for automated data analysis and decision support (Penzel and Conradt 2000). Sleep analysis requires to record physiological signals during several hours, providing high quantity of variables and information. Thus, specific software for data analysis represents a helpful tool for sleep physicians. For instance, automatic systems play an important role in sleep scoring from PSG recordings (Penzel and Conradt 2000). The use of computer-based tools for diagnostic purposes has extended in the last years (Pang and Terris 2006). As previously indicated, conventional PSG is the gold-standard in SAHS diagnosis but it suffers from limited resources, high cost, long waiting lists and intense labour requirements. SAHS is the most common sleep disorder and may severely affect patient's health and well-being. Thus, more effective diagnostic procedures are needed. Recently, numerous novel diagnostic methods have been proposed for SAHS. Most of them were based on automated signal analysis.

SAHS is related to different physiological functions involving respiratory, cardiac or nervous systems. It is also associated to limb movements and snoring during sleep. Moreover, diverse clinical and demographic variables such as increased neck circumference and BMI, daytime sleepiness, smoking habit or age usually accompany SAHS. Thus, several useful sources of information can be used to identify risk patients without the requirement of in-laboratory sleep studies. The diagnostic utility of different physiological recordings and clinical variables has been evaluated using signal processing techniques. These include conventional statistical measurements, spectral analysis, non-linear methods and pattern recognition techniques.

3.2. Signal processing-based methods to assist in SAHS diagnosis

Diverse methodologies have been suggested to overcome the inherent problems of conventional in-laboratory PSG. However, it is the only available procedure that ensures a reliable diagnosis. Other researchers evaluated the utility of home unattended PSG in order to preserve high reliability while reducing the demand for specific sleep units. Gagnadoux *et al.* (2002) assessed this strategy in a group of 111 patients. The percentage of ineffective recordings was computed as the performance measure. It was estimated at 23.4% for home unattended recordings. This approach was contrasted with in-laboratory telemonitored PSG, which provided 11.2% of ineffective recordings. Similarly, Portier *et al.* (2000) compared home unattended and conventional in-laboratory PSG in a population of 103 subjects. Home tests were not feasible for 33% of the patients. These results question the utility of PSG out of the laboratory environment. The proposed approaches could contribute to decrease the number of in-laboratory sleep studies but they are still complex and time-consuming due to the quantity of data that should be manually scored.

Automated interpretation of PSG recordings was suggested to assist physicians in scoring tasks. Some examples of this approach can be found in the literature (Guimaraes *et al.* 2001; Cabrero-Canosa *et al.* 2003; Cabrero-Canosa *et al.* 2004; Al-Ani *et al.* 2004; Álvarez-Estévez and Moret-Bonilo 2009). The methods proposed in these studies are based on the use of artificial intelligence techniques for automated detection of apnoea and hypopnoea events from signals recorded during PSG. The SAMOA system proposed by Cabrero-Canosa *et al.* (2004) enables to identify apnoea-hypopnoea events with a 93.9% agreement using if-then rules. The detected events are used to estimate the AHI, which provides a final diagnosis for subjects under study. A sensitivity of 87.5% and a specificity of 100% were obtained using a small set with 13 subjects. Nevertheless, these studies assumed that data acquisition is performed in a laboratory environment.

The mentioned methods can partially solve the problems derived from conventional PSG. In order to provide an effective solution for SAHS diagnosis, the new diagnostic techniques should be cheap, straightforward and automatic.

Furthermore, they are expected to enable home unattended monitoring. The cost of home studies has been estimated in 143.86 euros per study less expensive than PSG (Golpe *et al.* 2002). Therefore, researchers have focused their efforts on developing diagnostic methods based on a reduced set of signals or data recorded through simple devices. Subsequently, some kind of signal processing is applied to automatically analyse the recorded data. A wide variety of diagnostic methods adopting this approach have been proposed in the last years. Indeed, some of them were implemented in commercial devices that have been evaluated in previous studies. For instance, Watch_PAT100 (Ayas *et al.* 2003; Zou *et al.* 2006), MESAM 4 (Stoohs and Guilleminault 1992; Cirignotta *et al.* 2001), Autoset (Bradley *et al.* 1995; Gugger, 1997; Mayer *et al.* 1998; Kiely *et al.* 1996), ARES (Westbrook *et al.* 2005), Edentec Monitoring System (Parra *et al.* 1997), NovaSom QSG (Reichert *et al.* 2003), SNAP (Su *et al.* 2004), Embletta (Dingli *et al.* 2003), OxiFlow (Baltzan *et al.* 2000; Herer *et al.* 2002) or Nonin WristOx 3100 (Nigro *et al.* 2009) can be cited.

Several vital signs such as respiration or cardiac rhythm may reveal the presence of apnoeas during sleep. Neuronal activity is also affected and reflects arousals due to apnoea events (Sugi *et al.* 2009). Additionally, SAHS is related to specific clinical features. Thus, diverse physiological and clinical data can be used to develop new alternative diagnostic techniques. Indeed, the main difference between them relies on the signals and data selected for automated analysis. As in the present study, nocturnal pulse oximetry has been commonly used for this purpose since it reflects respiratory dynamics (Netzer *et al.* 2001). Clinical score, morphological features and questionnaires as well as other recordings such as ECG, EEG or respiratory signals have been also assessed. Moreover, the combination of information from several sources is a common approach to develop novel diagnostic procedures.

3.2.1. Automated analysis of nocturnal oximetry to assist in SAHS detection

Nocturnal oximetry recordings represent a useful tool for SAHS diagnosis (Netzer *et al.* 2001). Merely visual inspection of SaO₂ signals can provide relevant information about respiratory dynamics during sleep. The diagnostic utility of this approach was assessed by Rodríguez *et al.* (1996). A sensitivity of 91% and a

specificity of 69% were estimated using a population with 96 subjects. However, manual scoring is required. Several oxygen desaturation indices were proposed for computer-assisted processing of oximetry signals. They are calculated as the number of falls in the saturation value over a given threshold divided by the total hours of sleep (Bennet and Kinnear 1999). Usually, ODI3 and ODI4 are considered (Magalang *et al.* 2003). In addition, CT90 is also included among conventional oximetry measures (Netzer *et al.* 2001). These indices are commonly used in clinical practice since some commercial oximeters include the software to compute them (McNicholas, 2008). Their utility in SAHS diagnosis was evaluated in previous studies. A complete assessment of all of them can be found in the study of Gyulay *et al.* (1993). A population of 98 subjects was used to estimate the diagnostic capability of the three indices: ODI3, ODI4 and CT90. A positive diagnosis of SAHS was considered for $AHI \geq 15 \text{ h}^{-1}$. High differences between sensitivity and specificity statistics were found for all the indices. The highest sensitivity was achieved by CT90 with a value of 93% whereas its specificity was 51%. ODI4 provided 98% specificity, which was the highest value for this statistic. In contrast, its sensitivity was 40%. Despite their utility as a straightforward method for automated interpretation of SaO_2 data, the diagnostic performance of oximetry indices is unclear. Even, a strong dependence between their discrimination ability and BMI was observed (Nakano *et al.* 2004a). The published results significantly vary from one study to another. For instance, ODI4 provided substantially higher specificity than sensitivity in the study of Ryan *et al.* (1995) (31% sensitivity and 100% specificity), Hussain and Fleethan (2003) (33% sensitivity and 88% specificity) and Golpe *et al.* (1999) (32% sensitivity and 97% specificity). An opposite result can be found in the study of Magalang *et al.* (2003). Sensitivity and specificity values estimated for ODI4 in a population of 224 subjects were 94% and 44%, respectively. More balanced diagnostic results of this parameter were reported by other researchers. A sensitivity of 71% and a specificity of 93% were observed by Chiner *et al.* (1999) using data from 275 subjects. Oximetry signals from a larger population (359 subjects) were available for the study of Gurubhagavatula *et al.* (2001). ODI4 reached 85.2% sensitivity and 90% specificity when the criterion to define SAHS was $AHI \geq 5 \text{ h}^{-1}$.

Pepin *et al.* (1991) proposed an additional measure (Δ index) for automated interpretation of oximetry signals. It evaluates the variability of these recordings. This

parameter achieved 75% sensitivity and 86% specificity. Subsequently, Δ index was assessed by other researchers. It provided 88.5% sensitivity and 39.6% specificity in the study Olson *et al.* (1999), which analysed a total of 793 subjects. Magalang *et al.* (2003) achieved similar results for this index, with a sensitivity of 91% and a specificity of 59%. These results were improved by using the Δ index together with other conventional indices such as ODI3, ODI4 and the cumulative time spent below several levels of saturation. A sensitivity of 90% and a specificity of 70% were reached (Magalang *et al.* 2003). Other parameters from oximetry data were suggested for diagnostic purposes. For instance, Epstein and Dorlac (1998) focused on the detection of two patterns: a change greater than 4% in the saturation leading to a SaO₂ value lower than 90% and the occurrence of repetitive short-duration fluctuations in saturation. The former performed better, achieving 73.6% sensitivity and 89.4% specificity (Epstein and Dorlac 1998). This feature was previously suggested by Series *et al.* (1993). In this study, sensitivity and specificity were 98.2% and 47.7%, respectively. Kirby *et al.* (1992) proposed a measure called saturation impairment time (SIT) to account for the number and degree of desaturation events. It computes the area between the SaO₂ profile and a given threshold on the saturation value. Usually, it is fixed at 90% of saturation. Chesson *et al.* (1993) found that the value of this parameter in subjects with severe SAHS differed from subjects with mild and moderate SAHS. However, no differences were observed between mild and moderate subjects. Finally, a notable result was reported in the study by Vázquez *et al.* (2000). A sensitivity of 98% and a specificity of 88% were reached in a population of 245 subjects using a threshold of 15 h⁻¹ to define SAHS. An algorithm to estimate the number of respiratory disturbances during sleep was proposed. The algorithm detects drops in SaO₂. If an increased is detected, it checks whether three previous points were marked as drops. If one of them was equal or greater than 4%, a respiratory disturbance event was designated. The respiratory disturbance index (RDI) was computed as the number of events divided by the total hours of sleep. The correlation coefficient between AHI derived from PSG and RDI was 0.97. Sensitivity and specificity respectively decreased to 97% and 80% when the cutoff point for the definition of SAHS was placed at AHI = 10 h⁻¹.

Advanced signal processing techniques have been also applied to oximetry data. Conventional statistics include easy-to-calculate measures that can provide very

useful information about SAHS. The mean SaO_2 value has been used in previous studies as a SAHS indicator (Lacassagne *et al.* 1997; Herer *et al.* 1999; Heneghan *et al.* 2008a; de Chazal *et al.* 2009). Other statistics such as variance, skewness and kurtosis have been recently proposed to evaluate dispersion, asymmetry and sharpness, respectively, for the distribution of SaO_2 samples (Marcos *et al.* 2010). On the other hand, spectral analysis of SaO_2 signals revealed significant differences between SAHS-positive and SAHS-negative subjects. Zamarrón *et al.* (1999) used the peak amplitude and the relative signal power, computed from the power spectral density (PSD), in the band between 0.010 and 0.033 Hz to identify subjects from both groups. They found that these parameters tend to be higher in subjects affected by SAHS due to repeated desaturation events during sleep reflecting apnoeas. The peak amplitude in the selected frequency band provided 94% sensitivity and 65% specificity. Sensitivity and specificity reached by measuring the relative power were 91% and 67%, respectively. A similar spectral behaviour was observed for oximetry recordings from SAHS-positive subjects in the study by El-Solh *et al.* (2003). It was focused on the classification of signals into one of three groups: normal, SAHS-positive patients and subjects affected by Cheyne-Stokes respiration. Additionally, non-linear methods have been pointed out as a valuable instrument to evaluate SAHS-related properties of oximetry signals. *ApEn* has been adopted for measuring the irregularity of these recordings (Del Campo *et al.* 2006; Hornero *et al.* 2007). A more irregular behaviour is usually presented by signals from subjects affected by SAHS. The diagnostic utility of *ApEn* was evaluated by Hornero *et al.* (2007). A sensitivity of 82.09% and a specificity of 86.96% were achieved. In addition, *CTM* and *LZC* have been suggested for diagnostic purposes by measuring variability and complexity, respectively, of SaO_2 signals (Álvarez *et al.* 2006; Álvarez *et al.* 2007). Previous results show that the presence of SAHS is associated with more variable and complex signals. Álvarez *et al.* (2006) applied these methods to nocturnal oximetry signals from a population of 187 subjects. *CTM* provided the highest diagnostic performance, reaching 90.1% sensitivity and 82.9% specificity. Similar methods for evaluating variability were applied in subsequent studies (Morillo *et al.* 2009; Schmittendorf *et al.* 2009). Furthermore, several non-linear measures from oximetry data were combined for implementing diagnostic algorithms based on pattern recognition techniques. Marcos *et al.* (2008b) designed a MLP network to simultaneously process *ApEn*, *CTM* and *LZC* from oximetry signals. A sensitivity of 89.8% and a specificity of 79.4% were achieved. Similar results were obtained by

applying RBF networks, reaching 89.4% sensitivity and 81.4% specificity (Marcos *et al.* 2008a). In addition, the combination of spectral and non-linear features from SaO₂ recordings achieved promising results in SAHS diagnosis. Traditional statistical classifiers were used in the study by Marcos *et al.* (2009) to compare the utility of both groups of features as well as the combination approach. The highest diagnostic accuracy was reached by using LR and linear discriminant analysis (LDA) with spectral features. A sensitivity of 91.05% and a specificity of 82.61% were achieved by the latter approach. In order to select complementary input features, BY neural networks were proposed by Marcos *et al.* (2010) for pattern classification. These networks achieved 87.76% sensitivity and 82.39% specificity. *ApEn* and *LZC* together with mean and kurtosis measures of the SaO₂ variable were selected by the algorithm for effective pattern classification.

Clinical data or other physiological recordings have been analysed together with nocturnal oximetry signals to perform SAHS diagnosis. Indeed, it should be noted that SaO₂ recordings are usually considered for developing portable monitoring devices. For instance, SaO₂ signals have been used together with airflow signals (Baltzan *et al.* 2000; Herer *et al.* 2002; Reichert *et al.* 2003), snoring sound (Issa *et al.* 1993), peripheral arterial tonometry (Ayas *et al.* 2003) or heart rate (Stoohs *et al.* 1992; Cirignotta *et al.* 2001; Westbrook *et al.* 2005) for this purpose. Combining clinical data and oximetry features has been a common practice in previous researchers. This approach achieved a diagnostic accuracy of 75% in the study of Lacassagne *et al.* (1997). Features such as the mean SaO₂ value, CT90 or BMI were used. Roche *et al.* (2002) developed two algorithms for SAHS detection using clinical score and oximetry. LR was used for signal classification. It provided an accuracy of 53%. Additionally, a MLR model was built to estimate the AHI. An accuracy of 62.1% and a correlation coefficient of 0.38 were reached. Similarly, a pattern classification approach based on LR was proposed by Herer *et al.* (1999). In addition to clinical and oximetry features, it used data from pulmonary function testing and blood gas tension measurements. This method yielded a diagnostic accuracy of 72.5% in a population of 102 patients. On the other hand, combining SaO₂ and ECG recordings is a common approach to perform automated SAHS diagnosis. Raymond *et al.* (2003) proposed the cardiac oximetry disturbance index (CODI) for evaluating SAHS. It is computed from heart rate variability (HRV) and SaO₂ series as the number of autonomic arousals (assessed on the

basis of changes in the heart interbeat interval) that were coincident with a rise in oximetry. This index achieved 52% sensitivity and 89% specificity. It did not improve conventional oximetry indices. Zamarrón *et al.* (2003) analysed the spectral properties of HRV and SaO₂ signals. A diagnostic criterion based on the presence of a peak between 0.010 and 0.033 Hz in either of both PSD was used. A sensitivity of 94% and a specificity of 82% were reached. De Chazal *et al.* (2009) developed a system for identification of 1-minute apnoea events from oximetry signals and ECG. The RR interval series and the ECG-derived respiratory (EDR) signal were obtained from ECG. Time-domain and frequency-domain features from these signals together with time-domain oximetry features (mean and minimum SaO₂ values, number of samples lower than 92%, mean of the absolute differences between successive SaO₂ samples and the spread of SaO₂ values) were applied as input to a linear discriminant classifier. The event detection scheme enabled to estimate the AHI, which is used for classification of patients. A sensitivity of 94.8% and a specificity of 83.3% were reported. A similar methodology was presented in the study by Heneghan *et al.* (2008a). Finally, Yadollahi *et al.* (2010) recently proposed to combine information from oximetry data and tracheal sound signals to estimate the AHI. The acquisition of the sound signal required to place a neckband around patient's neck with a microphone attached to it. The method was validated on 66 subjects and it achieved a correlation of 0.96 with the reference AHI. Sensitivity and specificity were 83% and 91%, respectively (Yadollahi *et al.* 2010).

3.2.2. Other biomedical signals and data related to SAHS

The diagnostic ability of other methodologies based on data distinct from nocturnal oximetry has been assessed in previous studies. Clinical information, subjective impression and questionnaires can be easily acquired. Thus, they have been broadly used as a simplified means for SAHS detection. Hypertension, habitual snoring, neck circumference, BMI, age, sex and alcohol or tobacco consumption have been identified as clinical predictors of SAHS (Flemons *et al.* 1994; Dixon *et al.* 2003). In addition, Berlin questionnaire, Epworth Sleepiness Scale and Visual Analog Scale for snoring have been suggested for a rapid evaluation of the patient (Netzer *et al.* 1999; Friedman *et al.* 2010). Patients or their bed partners are required to complete these questionnaires. LR (Crocker *et al.* 1990; Deegan and McNicholas 1996) and MLR (Davies *et al.* 1992; Hoffstein and Szalai 1993) have been usually applied to develop

prediction rules using clinical features as those mentioned. They provided high sensitivity values ranging from 78% to 95%. However, specificity was significantly lower, with values from 41% to 63% (Flemons and McNicholas 1997). El-Solh *et al.* (1999) and Kirby *et al.* (1999) proposed more complex models based on neural networks to process clinical data. The former achieved 94.9% sensitivity and 64.7% specificity whereas the latter provided a sensitivity of 98.9% and a specificity of 80%.

Numerous methods based on ECG analysis can be found among the studies related to automated SAHS diagnosis. Penzel *et al.* (2002) compiled the methods presented to Computers in Cardiology Challenge 2000. It aimed to find the best algorithm to identify apnoea events from ECG signals. The event detection stage is then used to provide a definitive diagnosis. Among the presented methods, the algorithms of de Chazal *et al.* (2000) and Jarvis and Mitra (2000) achieved the highest performance in SAHS diagnosis. The former was based on the analysis of RR and PR interval series in time and frequency domains. LDA was used for pattern classification. The latter proposed time-frequency analysis of ECG signals using spectrograms. The KNN approach was used for classification. Both methods correctly classified all the thirty subjects in the test set. Subsequent studies based on the scheme proposed for this competition in de Chazal *et al.* (2000) were presented. For instance, de Chazal *et al.* (2003) used the same database to evaluate a similar methodology using the EDR signal and quadratic discriminant analysis (QDA). Heneghan *et al.* (2008b) proposed a method to estimate the AHI from apnoea events identified in ECG. The correlation with true AHI from PSG was 0.88. A sensitivity of 92% and a specificity of 69% were achieved in a population of 92 subjects. The same objective is considered in the study by Méndez *et al.* (2009). The following series were analysed: the RR interval and the area of the QRS complex. Temporal and spectral features from these series were used as input to KNN and MLP classifiers. The proposed method provided a correct diagnosis for the twenty-five subjects in the test set. These authors presented an analogous approach in a subsequent article (Méndez *et al.* 2010). In this case, empirical mode decomposition and wavelet analysis were applied to ECG for feature extraction. The same database was used for validation. Also, the new method achieved an accuracy of 100% for subjects in the test set. Other notable study related to ECG analysis was developed by Roche *et al.* (2003). A total of 147 subjects took part in the study. Feature extraction was performed by applying wavelet analysis to HRV series. A decision tree algorithm was used for

classification, providing 75.8% sensitivity and 64.2% specificity. Khandoker *et al.* (2009b) proposed a pattern recognition-based approach to diagnose SAHS from ECG. Features computed from wavelet analysis of HRV and EDR signals were used as input to a SVM classifier. Using a data set with 125 recordings, the method achieved a sensitivity of 92.31% and a specificity of 93.75%. Finally, Zamarrón *et al.* (2006) measured irregularity of HRV series using *ApEn*. ROC analysis was used to determine the optimum threshold for this parameter in order to discriminate SAHS-positive subjects. A sensitivity of 71.2% and a specificity of 78.9% were achieved.

Other physiological signals have been considered for developing novel procedures to assist in SAHS diagnosis. Respiratory signals such as airflow were used for implementing the AutoSet, a portable CPAP device that can be used in diagnosis mode. The system estimates the respiratory airflow by measuring pressure variations. It was evaluated in several previous studies (Bradley *et al.* 1995; Kiely *et al.* 1996; Gugger, 1997; Mayer *et al.* 1998). Mayer *et al.* (1998) used the largest population (95 subjects) for testing AutoSet. It reached 92% sensitivity and 79% specificity. Recently, an interesting approach based on short-time daytime recording of the nasal airway pressure was proposed by Salisbury and Sun (2007). Only a 5-minute recording is required to obtain a final diagnosis about SAHS. Time-frequency analysis of the recorded signals is performed using the Hilbert-Huang transform. The method was validated on two small groups of signals. An accuracy of 100% was achieved in the first group with 18 recordings. The second group was composed of 16 recordings. Sensitivity and specificity were 85.7% and 100%, respectively. On the other hand, physical examination has shown to provide relevant features about SAHS. Galvin *et al.* (1989) concluded that SAHS-positive patients have smaller oropharyngeal and nasopharyngeal airways. Similarly, Andersson *et al.* (1991) found that patients affected by SAHS showed a reduced posterior airway and a posterior rotation of the mandible. Thus, physical examination has been used as a predictor of SAHS in several studies. Caballero *et al.* (1998) studied if the calibre of the upper airway using computer tomography could be useful to distinguish patients with SAHS. Several features were measured in a population of 29 subjects. Stepwise feature selection and classification using KNN were used for implementing the pattern recognition system. A sensitivity of 64.29% and a specificity of 86.67% were reached. The study presented by Tsai *et al.* (2003) also used upper airway physical examination. Several features were measured in

75 subjects. Three of them were selected to develop the final prediction rule: cricomenal space of 1.5 cm or less, pharyngeal grade of more than II and the presence of overbite. The achieved sensitivity and specificity were 40% and 96%, respectively. Also, sound signals such as snoring have been proposed for SAHS detection. Usually, spectral analysis of these signals is performed to evaluate their properties. Fiz *et al.* (2010) proposed to study sound recordings from a tracheal microphone. Intensity and spectral features of these signals were used to classify subjects. A sensitivity of 80% and a specificity of 90% were reached. Nakano *et al.* (2004b) used spectral features from these recordings to estimate the AHI. A correlation of 0.93 was achieved, resulting in 79% sensitivity and 95% specificity. Ng *et al.* (2008) assessed the diagnostic utility of snoring signals. The first formant was used to discriminate SAHS-positive subjects. This feature provided a sensitivity of 88% and a specificity of 82%. Finally, the utility of EEG recordings in SAHS diagnosis was analysed. Spectral analysis of EEG is a common technique in order to recognise arousals accompanying events of apnoea. The EEG signal tends to present higher variability when these events occur, resulting in a larger amount of signal power associated to higher frequency components (Dingli *et al.* 2002). Some examples of this approach are the studies of Lin *et al.* (2006) and Sugi *et al.* (2009). Another relevant study was developed by Liu *et al.* (2008). A method to identify subjects affected by SAHS or narcolepsy using EEG and pupil size was proposed. Data from 5 SAHS-positive subjects, 4 narcoleptic patients and 6 controls were collected. Spectral analysis of EEG and the pupil size were used for feature extraction. A neural network based on adaptive resonance theory was used for classification. Nine out of eleven subjects affected by sleep disorders were correctly identified. Recently, Abdullah *et al.* (2010) evaluated the correlation between spectral features from EEG and HRV in different sleep stages. Results showed that EEG delta, sigma and beta bands exhibited a strong correlation with HRV features (low-frequency power, high-frequency power and the ratio of both quantities).

3.3. Utility of pattern recognition techniques as an aid tool for SAHS diagnosis

The analysis of biomedical signals with diagnostic purposes usually involves the use of pattern recognition methods. In medical diagnosis problems, a set of measurements or features of different nature is collected in order to take a decision or

estimate the value of a variable. Pattern recognition techniques are used to perform these tasks. They have been applied to several diagnostic problems including SAHS detection. Most of the cited studies were developed by using these methods.

There exist a wide variety of pattern recognition techniques. Models based on discriminant analysis or LR are probably the most commonly used. For instance, LDA and QDA classifiers were applied to the SAHS diagnosis problem from clinical data (Friedman *et al.* 2010), ECG recordings (de Chazal *et al.* 2000; McNames and Fraser 2000; de Chazal *et al.* 2003; de Chazal *et al.* 2004; Shouldice *et al.* 2004; Zywiets *et al.* 2004; Redmond and Heneghan 2006; Heneghan *et al.* 2008b; Méndez *et al.* 2010) and the combination of oximetry and ECG features (Heneghan *et al.* 2008a; de Chazal *et al.* 2009). On the other hand, several examples of diagnostic methods using LR can be found. Clinical score (Crocker *et al.* 1990; Deegan *et al.* 1996), nocturnal oximetry (Herer *et al.* 1999; Roche *et al.* 2002) and snoring signals (Fiz *et al.* 2010) were used to develop them. Furthermore, stepwise analysis has been used together with LR to perform feature selection (Viner *et al.* 1991; Flemons *et al.* 1994; Roche *et al.* 1999; Álvarez *et al.* 2010). KNN is another conventional technique for pattern classification that has been applied to SAHS detection. Two diagnostic algorithms based on this method were proposed to process ECG patterns (Jarvis and Mitra 2000; Méndez *et al.* 2009) and features derived from physical examination of the upper airway (Caballero *et al.* 1998). The diagnostic utility of LDA, QDA, LR and KNN classifiers was compared in a previous study (Marcos *et al.* 2009). Spectral and non-linear features from SaO₂ signals were used to assign subjects into two possible groups: SAHS-positive or SAHS-negative.

MLR has been traditionally employed for multivariate pattern analysis in regression problems. It enables to estimate the value of a variable from a set of predictors. This technique has been suggested in previous studies for automated SAHS detection using clinical and oximetry data (El-Solh *et al.* 1999; Roche *et al.* 2002). Moreover, MLR has been used to identify those features that best predict SAHS from a given set (Davies *et al.* 1992; Hoffstein and Szalai 1993; Wang *et al.* 1998).

Neural networks represent a more complex approach for classification and regression. They have been commonly applied to diverse problems including medical

diagnosis. MLP are the most known and used neural network. Several algorithms based on these networks have been proposed for SAHS diagnosis and detection of apnoea events. Different recordings have been considered:

- Clinical and anthropomorphic data (El-Solh *et al.* 1999).
- Oximetry (Taktak *et al.* 2000; Marcos *et al.* 2008b).
- ECG (Bystricky and Safer 2004; Al-Abed *et al.* 2007; Méndez *et al.* 2009).
- Respiratory recordings such as airflow or thoracic effort (Várady *et al.* 2002; Fontela-Romero *et al.* 2005; Weinreich *et al.* 2008; Sezgin and Tagluk 2009; Tagluk *et al.* 2010).
- EEG (Lin *et al.* 2006).

Other neural network models and advanced multivariate pattern analysis techniques have been also used for developing diagnostic methods for SAHS:

- RBF networks were used for classifying SaO₂ signals as SAHS-negative or SAHS-positive (Marcos *et al.* 2008a).
- A method based on PNN classifiers was proposed to identify Cheyne-Stokes and SAHS patients from SaO₂ data (El-Solh *et al.* 2003).
- A GRNN algorithm was used to approximate the AHI from clinical features (Kirby *et al.* 1999).
- An intelligent system for automated identification of respiratory disturbances was proposed by using self-organising maps (Guimaraes *et al.* 2001). Information from airflow, SaO₂, snoring and abdominal movements was processed.

- The utility of BY network classifiers to identify SaO₂ signals from SAHS patients was evaluated (Marcos *et al.* 2010).
- Adaptive resonance theory networks were also applied to SAHS detection using EEG data (Liu *et al.* 2008).
- SVM classifiers were used for SAHS detection from ECG features (Patangay *et al.* 2007; Khandoker *et al.* 2009a; Khandoker *et al.* 2009b)
- Several studies used Hidden Markov models to develop novel diagnostic techniques (McNames and Fraser 2000; Al-Ani *et al.* 2004; Chua *et al.* 2007).
- Gaussian mixture models were used for classifying ECG recordings from potential SAHS patients (Raymond *et al.* 2000).
- A MARS model was built to compute the AHI from several oximetry indices (Magalang *et al.* 2003).
- Decision trees were applied to process wavelet features from HRV series in order to provide a diagnosis about SAHS (Roche *et al.* 2003).
- Fuzzy logic has been also used for implementing diagnostic methods for SAHS (Nazeran *et al.* 2001; Álvarez-Estévez *et al.* 2009).

4. Subjects and signals

- 4.1. Population under study
- 4.2. Sleep study and diagnosis
- 4.3. Signal preprocessing

4.1. Population under study

The population available for this study was composed of 240 subjects suspected of suffering from SAHS. There was no distinction about the type of apnoeas (obstructive, central or mixed) for recruitment of subjects. All of them presented typical symptoms such as daytime sleepiness, loud snoring, nocturnal choking and awakenings or apnoea events reported by the subject or a bedmate. The data set consists of 80 SAHS-negative and 160 SAHS-positive subjects. The SAHS-negative group is composed of 52 men and 28 women. Their mean \pm standard deviation age, BMI and AHI are 47.23 ± 12.15 years, 27.83 ± 3.84 kg/m² and 3.94 ± 2.40 h⁻¹, respectively. On the other hand, a total of 134 men and 26 women form the SAHS-positive group. Their mean age and BMI are 54.78 ± 13.79 years and 30.83 ± 4.34 kg/m², respectively. The average AHI value in this group is 36.63 ± 25.73 h⁻¹. The percentage of males was significantly higher in the population of patients. As can be observed, the proportion of SAHS-positive patients is substantially greater than that of controls since subjects were referred to sleep study because of their symptoms and suspicions about SAHS.

The population was divided into training and test sets in order to provide an unbiased evaluation of the algorithms implemented in the Thesis. Approximately, 40% of the subjects under study were assigned to the training set whereas the remaining group was used for testing. Several random divisions of the initial population were carried out. Ideally, both sets should present similar mean values for age, percentage of males, BMI and recording time. Thus, the partition with the best balance for these variables was selected. The training set is allocated for algorithm optimisation. Usually, pattern recognition techniques require some adaptive parameters to be optimised according to the statistical properties of the problem. The training set, composed of a set of representative examples of the problem, is used for this purpose. In this study, the training set consisted of 96 subjects (32 SAHS-negative and 64 SAHS-positive). Once the optimisation process is completed, the trained algorithm is ready to be used. A test set composed of previously unseen samples is required to objectively assess the performance of the algorithm in the proposed problem. In this study, the oximetry signals from 144 subjects (48 SAHS-negative and 96 SAHS-positive) composed the test set. These signals were not previously included in the data set used for training. Table 1

summarises clinical and demographic data for the whole data set as well as for training and test sets separately.

4.2. Sleep study and diagnosis

Overnight conventional PSG was performed from midnight to 08:00 AM in the Sleep Unit of Hospital Universitario Pío del Río Hortega of Valladolid, Spain. The Review Board on Human Studies approved the protocol and each subject gave his informed consent to participate in the study. Patients were continuously monitored using a polysomnograph (Alice 5, Respiromics, Philips Healthcare, The Netherlands). A

ALL SUBJECTS			
	All	SAHS-positive	SAHS-negative
Subjects	240	160	80
Age (years)	52.26 ± 13.71	54.78 ± 13.79	47.23 ± 12.15
Males (%)	77.50	83.75	65.00
BMI (kg/m ²)	29.83 ± 4.38	30.83 ± 4.34	27.83 ± 3.84
Recording Time (h)	7.25 ± 0.55	7.23 ± 0.64	7.28 ± 0.31
AHI (h ⁻¹)	25.73 ± 26.09	36.63 ± 25.73	3.94 ± 2.40
TRAINING SET			
	All	SAHS-positive	SAHS-negative
Subjects	96	64	32
Age (years)	52.35 ± 13.76	54.88 ± 14.53	47.31 ± 10.59
Males (%)	77.08	84.38	62.50
BMI (kg/m ²)	29.83 ± 4.17	30.61 ± 3.86	28.27 ± 4.38
Recording Time (h)	7.25 ± 0.33	7.25 ± 0.35	7.25 ± 0.29
AHI (h ⁻¹)	24.75 ± 25.19	35.01 ± 25.16	4.23 ± 2.22
TEST SET			
	All	SAHS-positive	SAHS-negative
Subjects	144	96	48
Age (years)	52.19 ± 13.73	54.71 ± 13.35	47.17 ± 13.20
Males (%)	77.78	83.33	66.67
BMI (kg/m ²)	29.83 ± 4.53	30.98 ± 4.65	27.54 ± 3.26
Recording Time (h)	7.24 ± 0.66	7.22 ± 0.78	7.30 ± 0.33
AHI (h ⁻¹)	26.39 ± 26.74	37.71 ± 26.17	3.75 ± 2.51

Table 1. Clinical and demographic features for the available database of subjects as well as for training and test sets. BMI: body mass index; AHI: apnoea-hypopnoea index.

full description of the equipment and its technical specifications can be found in the Philips Respironics website (Philips Respironics). PSG studies included EEG, EOG, chin EMG, airflow (thermal and pressure flow), ECG, measurement of body position and oximetry. Signals are sent from patient interface to headbox through sensor cables. Data is sampled and sent to the base station where it is stored on a disk or sent through an ethernet connection to a computer. A sleep specialist analysed these recordings according to the system proposed by Rechtschaffen and Kales (Rechtschaffen and Kales 1968). The derived AHI is used to provide a reliable diagnosis for each subject. It served as the reference for validating the results obtained in the Thesis.

The methods proposed for automated SAHS diagnosis are based on SaO₂ analysis. Oximetry signals obtained through PSG composed the database available for this study. The polysomnograph equipment included a Nonin PureSAT pulse oximeter (Nonin Medical Inc., USA) with 3 seconds or faster averaging interval at a minimum heart rate of 60 beats per minute or greater. A finger probe was used to record these signals at a sampling frequency of 1 Hz. The recordings were saved to separate files to be off-line processed.

4.3. Signal preprocessing

One of the limitations of overnight pulse oximetry is given by motion artefacts caused by any movement of the probe on the extremity. Pulse oximetry is especially sensitive to this problem as the energy absorption due to arterial pulsation is only 1-2% of the total absorption (Moyle, 1996). Drops to zero observed in SaO₂ signals reflect artefacts occurring during sleep. These samples should be removed before performing analysis of oximetry data. However, removing zero samples may not be enough since some signals epochs could be affected by frequent drops. In this case, non-zero samples between consecutive drops do not provide a reliable measure of SaO₂ and the complete signal epoch should be removed.

A preprocessing stage was applied to SaO₂ signals to remove these artefacts. Magalang *et al.* (2003) proposed to eliminate all changes of SaO₂ signals between consecutive sampling intervals greater than 4%/s and any SaO₂ sample lower than 20%. These criteria were taken into account for implementing the preprocessing algorithm

developed in the Thesis. It evaluates each sample in the recording. The aim was to detect zero-samples and markedly affected signal epochs. The preprocessing algorithm involves the following steps:

1. Samples corresponding to the first 5 minutes of the signal are discarded.
2. A moving window with a length of 10 minutes is used to define a reference level. It is computed as the median value of the samples in this window. Initially, it is empty and new valid samples are progressively included in the window.
3. A pointer is used to analyse each sample of the signal. Three different states are defined:
 - 3.1. If the sample is distinct to zero, it is accepted as a valid sample. The moving window is updated and the pointer moves towards the next sample.
 - 3.2. If the sample is equal to zero, it is discarded and the pointer moves to the next sample.
 - 3.3. If a non-zero sample is found after a zero-sample, it will be accepted if the two following conditions are met:
 - The absolute difference between the value of the sample and the reference level obtained from the moving window should be less than 5% of the latter.
 - The next one-minute signal epoch (counted from the sample that met the previous condition) should present non-zero values.
4. If these two conditions are satisfied, the one-minute signal epoch after the non-zero sample is marked as valid, the moving window is updated and the pointer is placed in the next sample after this epoch. If there are zero values,

the pointer is placed in the next sample after the last zero value in the one-minute epoch. The sample contained by the pointer is evaluated. The algorithm will take one of the three possible states described in step 3.

Each time the algorithm accepts a signal epoch in state 3.3, the new segment, which was preceded by zero samples, must be assembled with the previous part of the signal. The following criterion was used:

5. A maximum step parameter of 4% (expressed in units of SaO_2) is considered for joining two different signals epochs (Magalang *et al.* 2003). The difference between extreme samples, i.e., the last sample identified as valid and the first sample of the new signal epoch is computed. It is divided by the step parameter and the obtained result is rounded towards minus infinity. The result provides the number of new samples to be included between both extreme samples. Interpolation techniques are used to compute the value of the signal in these new points.

Figure 7 shows an example of a SaO_2 recording after being processed for artefact removal. As can be observed, drops to zero and signal epochs containing repetitive frequent artefacts are automatically removed. The dynamic properties of the signal are preserved since no digital filtering is applied.

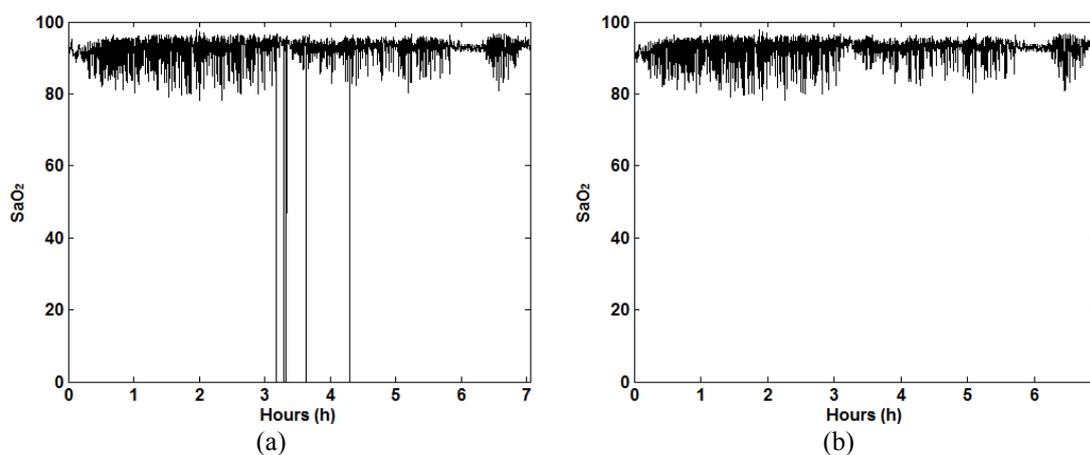


Figure 7. Example of a SaO_2 recording (a) before preprocessing and (b) after artefact removal using the proposed algorithm. As can be observed, drops to zero and subsequent abnormal values of saturation are removed.

5. Feature extraction, normalisation and dimensionality reduction

- 5.1. Motivation
- 5.2. Feature extraction
 - 5.2.1. Time-domain analysis
 - 5.2.2. Frequency-domain analysis
- 5.3. Normalisation
- 5.4. Principal component analysis (PCA)

5.1. Motivation

Pattern recognition algorithms were used to analyse SaO₂ recordings in order to help in SAHS diagnosis. Oximetry signals were recorded during a complete night of sleep. The average recording time from all signals used in the Thesis was 7.25 hours, which corresponds to more than 26000 samples. Using each sample in the recording as an input variable to the pattern recognition algorithm would be an impractical approach. The length of the recording varies from one subject to another. In addition, the input dimension would be extremely high. This would lead to an intractable system that would probably provide poor generalisation because of the curse of dimensionality (Bishop, 1995). This term is used to explain the relationship between the size of the training set and the input dimension. The curse of dimensionality shows that the quantity of training data needed to specify the underlying problem exponentially grows with respect to the dimension of the input patterns. Thus, the number of training samples should be increased according to the input dimension in order to avoid overfitting and achieve good generalisation.

Therefore, it is necessary to transform the signal into a new representation before the pattern recognition stage. Feature extraction is required to map the SaO₂ signal into a reduced set of variables or features that summarise the information in the recording. The term feature extraction is used to denote the process of forming linear or non-linear combinations of original data to generate new variables (Bishop, 1995). Each of them represents a specific attribute of the signal. Feature extraction contributes to alleviate the problem of overfitting due to the curse of dimensionality. A model with fewer inputs has fewer adaptive parameters to be determined and these are more likely to be properly constrained by a training set of limited size. In addition, another two important benefits are derived from feature extraction. First, it enables to incorporate prior knowledge in the pattern recognition algorithm (Bishop, 1995). The designer selects those features to be extracted from the original data in order to enhance the most relevant signal properties for the problem under study. Second, the computational resources needed for training are substantially reduced.

In this study, SAHS diagnosis was modelled by applying classification and regression techniques to oximetry signals. Prior knowledge about the behaviour of these

signals regarding SAHS should be used to evaluate appropriate properties for both approaches. For classification, the extracted features are expected to provide useful measures from oximetry data in order to discriminate signals from SAHS-positive subjects. In regression modelling, they should correlate with the degree of severity of SAHS to obtain a suitable AHI estimation. A single feature set was selected for both models. According to the domain used for analysing SaO₂ signals, the extracted features were divided into two groups: time-domain and frequency-domain features. The former is related to the analysis of the temporal evolution of signals. The latter refers to the study of their frequency components. Both groups of features were used to characterise the dynamic properties of SaO₂ signals regarding SAHS.

5.2. Feature extraction

5.2.1. Time-domain analysis

Time representation of oximetry signals reflects respiratory dynamics during sleep. Apnoea events are characterised by a decrease in the SaO₂ value due to airway obstruction and reduced airflow. Therefore, signals from SAHS-positive subjects are usually associated to instability. They reflect continuous drops and subsequent restorations of the saturation value due to the repetition of apnoeas. In contrast, signals from SAHS-negative subjects tend to present a near constant saturation value around 96% (Netzer *et al.* 2001). Figure 8 shows a one-hour epoch from two recordings

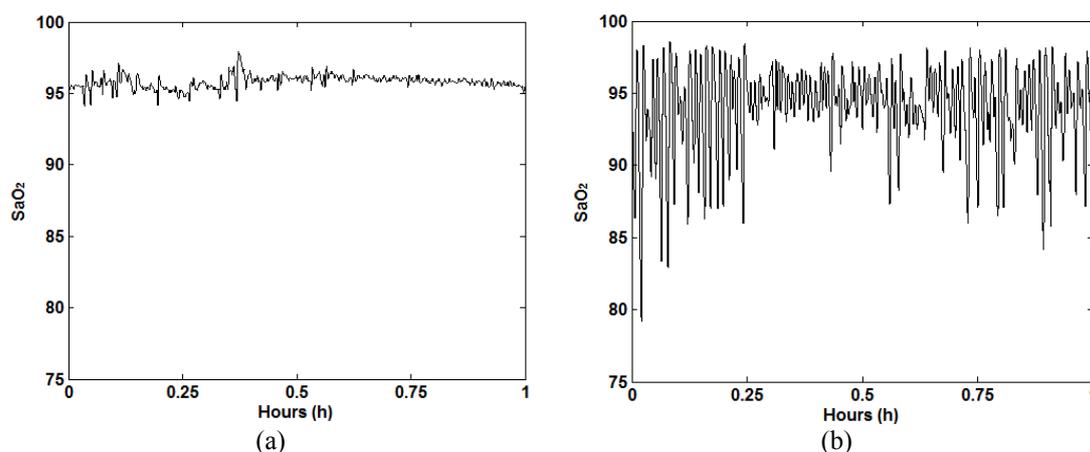


Figure 8. Comparison of dynamic properties of oximetry signals corresponding to clear control subjects and subjects suffering from SAHS: (a) an one-hour epoch of a SaO₂ signal from a subject with AHI = 1.4 h⁻¹; (b) an one-hour epoch of a SaO₂ signal from a subject with AHI = 66.2 h⁻¹.

corresponding to a SAHS-negative subject ($AHI = 1.4 \text{ h}^{-1}$) and a patient with severe SAHS ($AHI = 66.2 \text{ h}^{-1}$), respectively. Notable differences between both tracings can be observed. Marked events of desaturation distinguish the SaO_2 signal from the SAHS-positive subject.

In order to quantify these dynamic differences, oximetry recordings were analysed in the time domain using conventional statistics and non-linear methods. The former represent a straightforward tool to evaluate the behaviour of SaO_2 signals. The first four standard moments in the time domain were computed for the variable representing the saturation value. To illustrate the statistical behaviour of this variable, the distributions of SaO_2 samples corresponding to a SAHS-negative subject, a doubtful SAHS-negative subject, a doubtful SAHS-positive subject and a SAHS-positive subject are shown in Figure 9. Non-linear methods represent a more advanced approach for processing oximetry data. *ApEn*, *CTM* and *LZC* were applied. The use of these features

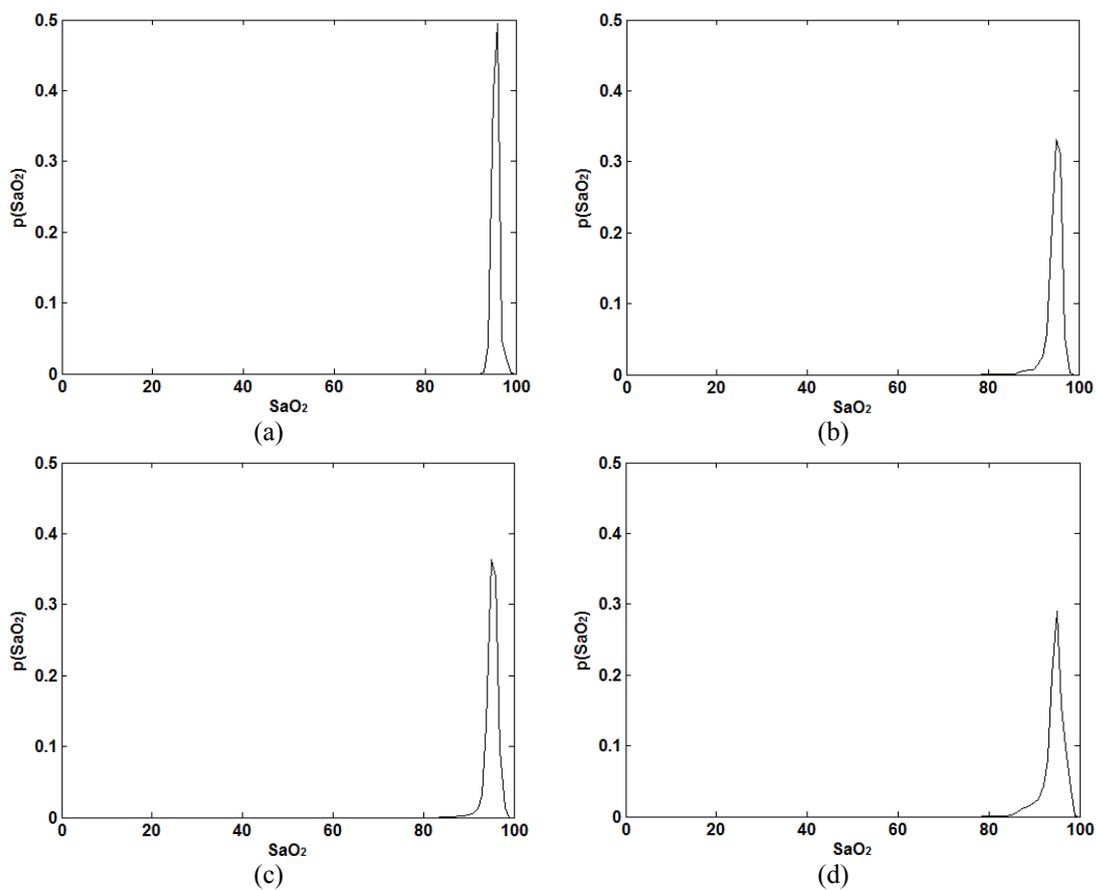


Figure 9. Distribution of the variable SaO_2 for (a) a SAHS-negative subject ($AHI = 1.4 \text{ h}^{-1}$), (b) a doubtful SAHS-negative subject ($AHI = 8.1 \text{ h}^{-1}$), (c) a doubtful SAHS-positive subject ($AHI = 13.1 \text{ h}^{-1}$) and (d) a SAHS-positive subject ($AHI = 66.2 \text{ h}^{-1}$).

as indicators of SAHS was suggested in preceding studies (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Álvarez *et al.* 2007; Hornero *et al.* 2007; Marcos *et al.* 2010a).

The statistical properties of the SaO₂ variable were analysed. The first four standard moments of this variable were computed. To evaluate these features, the probability density function was modelled with a discrete distribution, assigning to each saturation value in the series $\mathbf{s} = [s(1), s(2), \dots, s(T)]$ the probability given by its relative frequency, i.e. the number of times that the SaO₂ value was observed out of the total number of samples given by T . The signal properties evaluated by these features are explained in the following points (Jobson, 1991a):

- Feature 1. First statistical moment in the time domain or mean (μ_t). It quantifies the central tendency for the distribution of SaO₂ samples. The mean is supposed to be lower for SAHS-positive patients due to frequent drops in the saturation value associated to apnoea events. The expression used to compute μ_t is given by:

$$\mu_t = \frac{1}{T} \sum_{i=1}^T s(i). \quad (28)$$

- Feature 2. Second statistical moment in the time domain or variance (σ_t^2). It indicates the degree of dispersion for the distribution of SaO₂ samples. SAHS-positive patients are expected to provide higher values of σ_t^2 due to instability of their recordings. The variance was computed as:

$$\sigma_t^2 = \frac{1}{(T-1)} \sum_{i=1}^T [s(i) - \mu_t]^2. \quad (29)$$

- Feature 3. Third statistical moment in the time domain or skewness (γ_t). It measures the asymmetry for the distribution of SaO₂ samples. A negative skew indicates that the tail on the left side of the distribution is longer than on the right side and most of the values lie on the right of the mean. Conversely, a positive skew is obtained for distributions with a longer tail on the right side. Typically, this feature is negative in positive subjects due

to a higher concentration of samples with low values of saturation. The definition of skewness is given by:

$$\gamma_t = \frac{\frac{1}{(T-1)} \sum_{i=1}^T [s(i) - \mu_t]^3}{\sigma_t^3}. \quad (30)$$

- Feature 4. Fourth statistical moment in the time domain or kurtosis (δ_t). It evaluates the sharpness for the distribution of SaO₂ samples. Higher kurtosis reflects the existence of infrequent values that contribute to increase the variance. Kurtosis is expected to be higher in control subjects since their saturation value tends to be constant. It was computed as:

$$\delta_t = \frac{\frac{1}{(T-1)} \sum_{i=1}^T [s(i) - \mu_t]^4}{\sigma_t^4}. \quad (31)$$

Non-linear analysis was also used to assess oximetry dynamics in the time domain. Irregularity, variability and complexity of SaO₂ recordings were evaluated by using *ApEn*, *CTM* and *LZC*, respectively. It has been shown that the measures provided by these methods correlate with the severity of SAHS (Álvarez *et al.* 2006; Hornero *et al.* 2007). The following points provide an explanation of the non-linear techniques used for feature extraction:

- Feature 5. Approximate entropy (*ApEn*). *ApEn* is a family of parameters and statistics introduced to quantify irregularity in data without any a priori knowledge about the system generating them (Pincus, 2001). *ApEn* assigns a non-negative number to a time series, with larger values corresponding to more irregularity in the data. The distance between two runs composed of m consecutive samples of the original series $\mathbf{s} = [s(1), s(2), \dots, s(T)]$ is evaluated to compute *ApEn*. They are considered close if the obtained value is less than a threshold r (Pincus, 2001). Parameters m and r must be specified by the user. *ApEn(m,r)*, or *ApEn(m,r,T)*, with T the number of points in the time series, must be considered as a family of parameters:

comparisons between time series segments can only be made with the same values of m and r (Pincus, 2001). Formally, given T data points from a time series \mathbf{s} , the *ApEn* is computed as follows.

1. Form vectors with length m , from \mathbf{s}_1 to \mathbf{s}_{T-m+1} , defined as $\mathbf{s}_i = [s(i), s(i+1), \dots, s(i+m-1)]$, $i = 1, \dots, T-m+1$. These vectors represent m consecutive s values, commencing with the i th point.
2. Calculate the distance between \mathbf{s}_i and \mathbf{s}_j , $d[\mathbf{s}_i, \mathbf{s}_j]$ as the maximum absolute difference between their respective scalar components:

$$d[\mathbf{s}_i, \mathbf{s}_j] = \max_{k=1,2,\dots,m} \{|s(i+k-1) - s(j+k-1)|\}. \quad (32)$$

3. For a given \mathbf{s}_i count the number of j ($j = 1, 2, \dots, T-m+1$) for $j \neq i$ such that $d[\mathbf{s}_i, \mathbf{s}_j] \leq r$ denoted as $N^m(i)$. Then, for $i = 1, 2, \dots, T-m+1$:

$$C_r^m(i) = \frac{N^m(i)}{T-m+1}. \quad (33)$$

The $C_r^m(i)$ value measures, within a tolerance r , the regularity (frequency) of patterns similar to a given one of window length m .

4. Compute the natural logarithm of each $C_r^m(i)$ and obtain its average over i :

$$\phi^m(r) = \frac{1}{T-m+1} \sum_{i=1}^{T-m+1} \ln C_r^m(i), \quad (34)$$

where $\phi^m(r)$ represents the average frequency of all the m -point patterns in the sequence remain close to each other.

5. Increase the dimension to $m + 1$. Repeat steps from 1 to 4 and find $C_r^{m+1}(i)$ and $\phi^{m+1}(r)$.
6. Theoretically, the *ApEn* is defined as:

$$ApEn(m, r) = \lim_{T \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)]. \quad (35)$$

Usually, high values of *ApEn* are associated with SaO₂ signals from SAHS-positive subjects. These tend to present a more irregular behaviour due to frequent changes in the saturation value (Hornero *et al.* 2007).

- Feature 6. Central tendency measure (*CTM*). *CTM* quantifies signal variability by using second-order difference plots. These are graphs centred in the origin (scatter diagram) and are used to assess the degree of chaos in a data set (Cohen *et al.* 1996). A plot of $[s(i+2) - s(i+1)]$ vs. $[s(i+1) - s(i)]$ is made to represent the graph. The *CTM* is computed by selecting a circular region of radius ρ_{CTM} around the origin, counting the number of points that fall within the radius and dividing by the total number of points. Given a series with T data points, the total number of points in the scatter plot is $T - 2$. Then, the *CTM* can be computed as (Cohen *et al.* 1996):

$$CTM = \frac{\sum_{i=1}^{T-2} \delta(d_i)}{T-2}, \quad (36)$$

where

$$\delta(d_i) = \begin{cases} 1, & \text{if } \left\{ [s(i+2) - s(i+1)]^2 + [s(i+1) - s(i)]^2 \right\}^{1/2} < \rho_{CTM} \\ 0, & \text{otherwise} \end{cases} \quad (37)$$

Low *CTM* values are obtained from series with high variability. Typically, this is the case for oximetry recordings from SAHS-positive patients due to the recurrence of apnoeas (Álvarez *et al.* 2006).

- Feature 7. Lempel-Ziv complexity (*LZC*). *LZC* is a non-parametric, simple-to-calculate measure of complexity in a one-dimensional signal that does not require long data segments to compute (Lempel and Ziv 1976). It is related to the number of distinct substrings and the rate of their recurrence along the given sequence. *LZC* analysis is based on a coarse-graining of the measurements, so before calculating the complexity measure $c(T)$ the signal must be transformed into a finite symbol sequence. The simplest way is to convert the time series values $s(i)$ ($i = 1, \dots, T$) into a two-symbol sequence. The median value of the samples is estimated as a threshold T_d . By comparison with the threshold, the signal data is converted into a 0-1 sequence $P_{seq} = [u(1), u(2), \dots, u(T)]$ (Lempel and Ziv 1976):

$$u(i) = \begin{cases} 0, & \text{if } s(i) < T_d \\ 1, & \text{otherwise} \end{cases} \quad (38)$$

The sequence P_{seq} is scanned from left to right and the complexity counter $c(T)$ is increased by one unit every time a new subsequence of consecutive characters is encountered. The complexity measure can be estimated using the following algorithm (Lempel and Ziv 1976):

1. Let S and Q denote two subsequences of P_{seq} and SQ be the concatenation of S and Q , while sequence $SQ\pi$ is derived from SQ after its last character is deleted (π means the operation to delete the last character in the sequence). Let $v(SQ\pi)$ denote the vocabulary of all different subsequences of $SQ\pi$. At the beginning, $c(T) = 1$, $S = u(1)$, $Q = u(2)$, therefore, $SQ\pi = s(1)$.
2. In general, $S = u(1), u(2), \dots, u(r)$, $q = u(r + 1)$, then $SQ\pi = u(1), u(2), \dots, u(r)$; if Q belongs to $v(SQ\pi)$, then Q is a subsequence of $SQ\pi$, not a new sequence.
3. Renew Q to be $u(r + 1), u(r + 2)$ and judge if Q belongs to $v(SQ\pi)$ or not.

4. Repeat the previous steps until Q does not belong to $v(SQ\pi)$. Now $Q = u(r+1), u(r+2), \dots, u(r+i)$ is not a subsequence of $SQ\pi = u(1), u(2), \dots, u(r+i-1)$, so increase $c(T)$ by one.
5. Thereafter, S is renewed to be $S = u(1), u(2), \dots, u(r+i)$ and $Q = u(r+i+1)$.

These procedures have to be repeated until Q is the last character. At this time, the number of different subsequences in P (i.e. the measure of complexity) is $c(T)$.

In order to obtain a complexity measure which is independent of the sequence length, $c(T)$ should be normalised. If the length of the sequence is T and the number of different symbols in the symbol set is α_{sym} , it has been proved that the upper bound of $c(T)$ is given by (Lempel and Ziv 1976):

$$c(T) < \frac{T}{(1 - \varepsilon_T) \log_{\alpha_{sym}}(T)}, \quad (39)$$

where ε_T is a small quantity and $\varepsilon_T \rightarrow 0$ ($T \rightarrow \infty$). In general, $T / \log_{\alpha_{sym}}(T)$ is the upper bound of $c(T)$, where the base of the logarithm is α_{sym} . The following equality can be written:

$$\lim_{T \rightarrow \infty} c(T) = b(T) \equiv \frac{T}{\log_{\alpha_{sym}}(T)}. \quad (40)$$

For a 0-1 sequence $\alpha_{sym} = 2$, therefore:

$$b(T) \equiv \frac{T}{\log_2(T)}. \quad (41)$$

The counter $c(T)$ can be normalised via $b(T)$:

$$LZC \equiv C(T) \equiv \frac{c(T)}{b(T)}. \quad (42)$$

The variable $C(T)$ (representing the LZC) reflects the arising rate of new patterns along with the sequence (Lempel and Ziv 1976). A high value of LZC is obtained from time series with higher complexity, which is supposed for oximetry recordings from SAHS positive patients (Álvarez *et al.* 2006).

5.2.2. Frequency-domain analysis

Spectral analysis of SaO_2 signals reveals differences between SAHS-positive and SAHS-negative subjects. Preceding studies found that signal power associated with frequency components between 0.010 and 0.033 Hz is usually higher in subjects with SAHS than in controls (Zamarrón *et al.* 2003). Apnoeas originate phase-lagged changes in SaO_2 signals. Their duration usually ranges from 30 seconds to 2 minutes, including the awakening response after the event. Patients suffering from SAHS may have several consecutive breathings reflecting apnoeas or hypopnoeas. Thus, the repetition of these events during these periods of sleep will be produced at a rate between 30 seconds and 2 minutes, which correspond with a frequency of 0.033 Hz and 0.010 Hz, respectively. Relevant frequency components in this range denote fluctuations in oximetry recordings due to phases with repetitive apnoeas. This behaviour can be examined through spectral analysis of oximetry data. To illustrate this, Figure 10 shows the PSD computed from SaO_2 signals corresponding to the four subjects used in previous examples. The power associated to frequency components between 0.010 and 0.033 Hz is almost zero for the control subject. In contrast, the recording from the SAHS-positive subject presents significant frequency components located in that band.

The non-parametric Welch's method was used to compute the PSD of oximetry recordings (Welch, 1967). Briefly, the original recording $\mathbf{s} = [s(1), s(2), \dots, s(T)]$ is divided into M overlapping sequences of length L by applying a window function \mathbf{w}_f . The modified periodogram is computed for each of them by using the finite Fourier transform. It is given by:

$$\hat{S}_m(k) = \frac{1}{LU} \left| \sum_{i=0}^{L-1} s(i) w_f(i) e^{-\frac{j2\pi ki}{L}} \right|^2, \quad (43)$$

where

$$U = \frac{1}{L} \sum_{i=0}^{L-1} |w_f(i)|^2. \quad (44)$$

Finally, the PSD estimation is obtained as the average of the periodograms:

$$\hat{S}(k) = \frac{1}{M} \sum_{m=0}^{M-1} \hat{S}_m(k). \quad (45)$$

Two sets of features were derived from the PSD function (Marcos *et al.* 2009; Marcos *et al.* 2010). Initially, statistical analysis was used to characterise the PSD

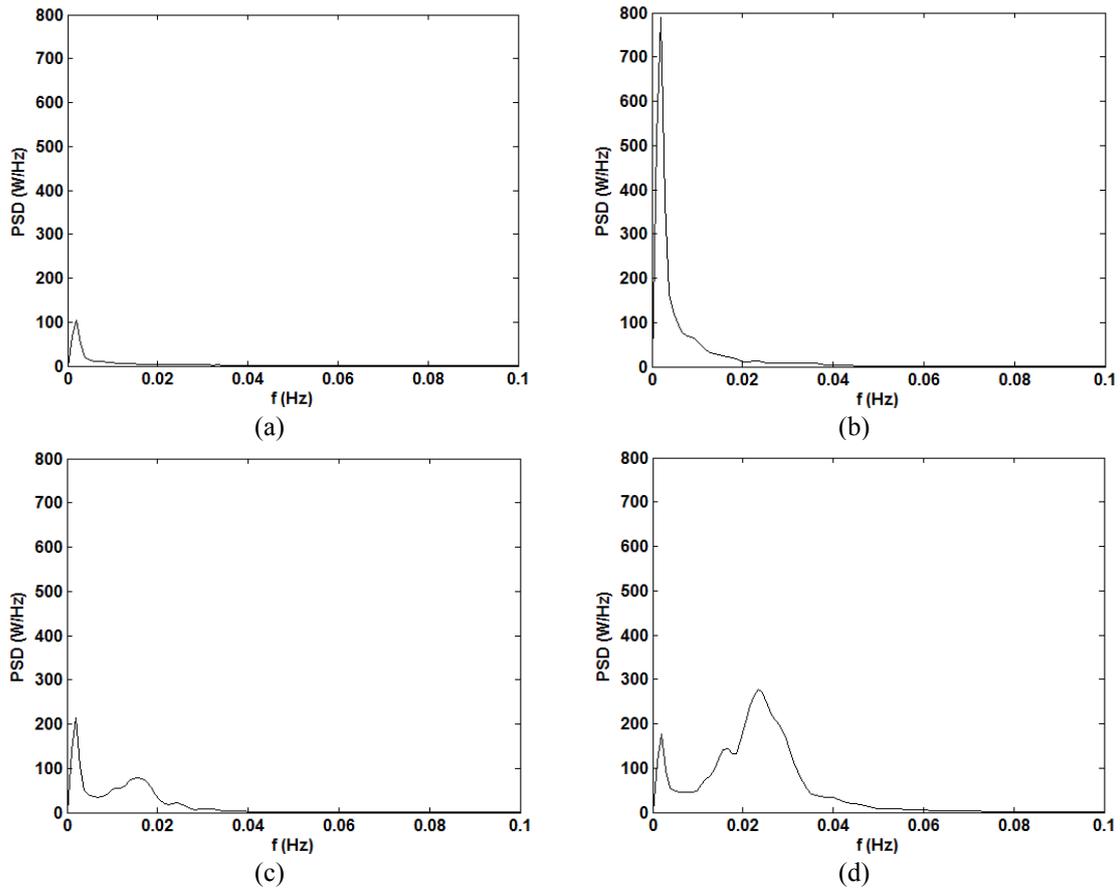


Figure 10. PSD function corresponding to (a) a SAHS-negative subject ($AHI = 1.4 \text{ h}^{-1}$), (b) a doubtful SAHS-negative subject ($AHI = 8.1 \text{ h}^{-1}$), (c) a doubtful SAHS-positive subject ($AHI = 13.1 \text{ h}^{-1}$) and (d) a SAHS-positive subject ($AHI = 66.2 \text{ h}^{-1}$).

of SaO₂ signals. The variable representing the frequency component was modelled using the normalised PSD (PSD_n) as its probability density function. Only the positive half of the original PSD was considered. As proposed in time-domain analysis, the first four standard moments were used to characterise this variable:

- Feature 8. First statistical moment in the frequency domain or mean (μ_f). It estimates the central tendency of the variable defined by the frequency component. It is expected to be higher for SAHS-positive patients since their SaO₂ signals tend to present a higher bandwidth. The value of μ_f is computed as:

$$\mu_f = \sum_{f=0}^{f_s/2} PSD_n(f) f. \quad (46)$$

- Feature 9. Second statistical moment in the frequency domain or variance (σ_f^2). It represents the degree of dispersion of the variable defined by the frequency component. Similarly, higher values of this feature are associated to patients affected by SAHS. The following expression was used to compute the variance of variable f :

$$\sigma_f^2 = \sum_{f=0}^{f_s/2} PSD_n(f) (f - \mu_f)^2. \quad (47)$$

- Feature 10. Third statistical moment in the frequency domain or skewness (γ_f). It measures the asymmetry of the variable defined by the frequency component. It is positive for both groups of subjects. Higher values are associated to signals from SAHS-negative subjects since their PSD tends to be concentrated on frequencies close to zero. Feature γ_f was obtained as:

$$\gamma_f = \frac{\sum_{f=0}^{f_s/2} PSD_n(f) (f - \mu_f)^3}{\sigma_f^3}. \quad (48)$$

- Feature 11. Fourth statistical moment in the frequency domain or kurtosis (δ_f). It evaluates the sharpness of the distribution defined by the frequency component. It is expected to be higher for SAHS-negative subjects since the power of the signal is concentrated in low frequency components, resulting in a peaked PSD function. Kurtosis was computed as:

$$\delta_f = \frac{\sum_{f=0}^{f_s/2} PSD_n(f)(f - \mu_f)^4}{\sigma_f^4}. \quad (49)$$

In addition, three features were directly computed from the PSD function. The properties evaluated by them are described in the following points:

- Feature 12. Total area under the PSD (S_T). S_T evaluates the power of the SaO₂ signal under study. Higher values are associated to signals from SAHS-positive subjects due to frequent changes and variability. This feature is obtained as:

$$S_T = \sum_{f=0}^{f_s/2} PSD(f). \quad (50)$$

- Feature 13. Area enclosed in the band of interest (S_B). S_B measures the signal power corresponding to frequencies in the band between 0.010 and 0.033 Hz. As indicated before, it is usually higher for SAHS-positive patients. The expression to compute S_B is given by:

$$S_B = \sum_{f=0.010}^{0.033} PSD(f). \quad (51)$$

- Feature 14. Peak amplitude of the PSD in the band of interest (PA). PA represents the most significant frequency component contained in the band between 0.010 and 0.033 Hz. It is expected to be higher in SAHS-positive patients due to repetitive changes in SaO₂ signals corresponding to these frequencies. The definition of PA is given by:

$$PA = \max_{PSD(f)} \{PSD(f)\}, f \in [0.010, 0.033] (Hz). \quad (52)$$

5.3. Normalisation

The extracted features represent different properties of the signal under analysis. Therefore, their values may differ significantly. Simple linear rescaling is useful to avoid differences between the magnitudes of the input features. Each feature (\tilde{x}_i) was independently treated. Its mean (μ_i) and standard deviation (σ_i) were computed with respect to the training set. A set of rescaled variables were obtaining by applying the following linear transformation (Bishop, 1995):

$$x_i^n = \frac{\tilde{x}_i^n - \mu_i}{\sigma_i}, \quad (53)$$

where n and i are the sample and feature indices, respectively, x_i^n is the normalised value of feature i for sample n and \tilde{x}_i^n is its corresponding raw value. The transformed variables have zero mean and unit standard deviation.

It should be noted that this linear transformation does not have any effect when the pattern recognition technique is also linear. However, normalisation is particularly important in non-linear methods. For instance, MLP networks require the input variables to have similar values for suitable weight initialisation and optimisation. Differences in the magnitude of the inputs should be balanced with the value of the network weights. Furthermore, the need for input normalisation is evident in kernel methods such as RBF or SVM. They compute the distance between the input pattern and some representative patterns to provide their output value. If one of the input variables has a much smaller range of values than the others, the distance will be very insensitive to this variable (Bishop, 1995).

5.4. Principal component analysis (PCA)

There is not an exact rule to determine the appropriate model complexity from the size of the training set. Overfitting may arise if the complexity of the pattern

recognition technique is too high regarding the number of available training samples. Methods for dimensionality reduction could be applied after feature extraction and normalisation processes in order to prevent overfitting due to the curse of dimensionality and excessive variance. In general, a reduction in the dimensionality of the input space is accompanied by a loss of information. Therefore, the goal is to find the smallest set of inputs that can result in satisfactory predictive performance.

One of the most popular methods for dimensionality reduction is PCA (Jolliffe, 2002). In this study, pattern recognition techniques with and without PCA were compared to evaluate the improvement derived from its use. PCA rotates vectors in the original feature space onto a new basis. The covariance matrix (Σ_x) of the variable representing original vectors $\mathbf{x} = (x_1, \dots, x_d)$, with d the number of extracted features, is computed. Its eigenvectors represent an orthonormal basis in the original space. PCA produces an uncorrelated set of variables or components $\mathbf{x}_{PC} = (PC_1, \dots, PC_d)$ by projecting the input data on the new basis:

$$\mathbf{X}_{PC} = \mathbf{X}^T \mathbf{A}_x. \quad (54)$$

where matrix \mathbf{X}_{PC} ($d \times N$) contains the N samples of dimension d with components produced by PCA, matrix \mathbf{X} ($d \times N$) contains the N samples of dimension d with the original features and matrix \mathbf{A} ($d \times d$) is the orthogonal matrix whose k th column is the k th eigenvector of matrix Σ_x . Each new variable PC_i can be viewed as a linear combination of features x_i in the original space.

The projected components are ranked by PCA in decreasing importance. It is the most relevant property of this technique. The amount of variance explained by each component is represented by its associated eigenvalue. The obtained order can be used to perform dimensionality reduction. A subset of the components is selected to define a destination space with dimension $d_{PCA} \leq d$. To reduce the loss of information, the d_{PCA} components with the highest eigenvalues are retained. They represent the inputs to the pattern recognition method. In practice, the algorithm proceeds as follows (Jolliffe, 2002):

1. The mean of the vectors in the original space is computed and subtracted.
2. Matrix Σ_x is calculated and its eigenvectors and eigenvalues are found.
3. The eigenvectors corresponding to the d_{PCA} largest eigenvalues are retained.
4. The original vectors are projected onto the eigenvectors to give the components of the transformed vectors in the d_{PCA} -dimensional space.

There exist different criteria for selecting the appropriate number of components. The Cattell's scree test (Cattell, 1966) is a graphical method in which the values of the eigenvalues are plotted in descending order. The envelope curve is obtained and a change in its slope is used to define the selected components. The criterion proposed by Kaiser (1960) suggests using only those components that explain at least the variance equal to the average variance of the original features (i.e. components with eigenvalues larger than one). Finally, another common rule consists in fixing a fraction of the variance to be retained and selecting the components needed to explain this percentage (Nabney, 2002). The latter approach was used in this study. A percentage of 90% was applied to the variance of original data to determine the number of components to be selected.

6. Pattern recognition methods for classification

- 6.1. Introduction
- 6.2. *K*-nearest neighbours (KNN)
- 6.3. Fisher's linear discriminant (FLD)
- 6.4. Logistic regression (LR)
- 6.5. Probabilistic neural networks (PNN)
- 6.6. Multilayer perceptron (MLP) networks for classification
- 6.7. Radial basis function (RBF) networks for classification
- 6.8. Bayesian (BY) networks for classification
- 6.9. Support vector machines (SVM) for classification

6.1. Introduction

The first approach to model the SAHS diagnosis problem was based on classification techniques. In a classification context, each feature pattern representing an oximetry recording must be labelled as SAHS-positive or SAHS-negative. It corresponds to a two-class classification task. Diverse methods can be used to address it. In this Thesis, several pattern recognition techniques were proposed for classification. In this chapter, a detailed description is provided for them.

6.2. *K*-nearest neighbours (KNN)

The KNN method is a simple approach for density estimation. Patterns \mathbf{x} are supposed to be drawn from an unknown density function $p(\mathbf{x})$. KNN is based on two previous assumptions to estimate $p(\mathbf{x})$ (Bishop, 1995). First, the probability associated to region R in the input space is approximated as:

$$P \approx \frac{K}{N}, \quad (55)$$

where N is the number of available patterns and K represents the number of patterns that fall in the region R . Second, the probability density function is assumed to be constant around point \mathbf{x} . Therefore, the probability P would be given by:

$$P = \int_R p(\mathbf{x}') d\mathbf{x}' \approx p(\mathbf{x})V, \quad (56)$$

where V is the volume of the region R . From both previous assumptions, the approximation for the probability density function $p(\mathbf{x})$ is obtained as:

$$p(\mathbf{x}) \approx \frac{P}{V} = \frac{K}{NV}. \quad (57)$$

The validity of both assumptions is governed by the choice of the region R . The first approximation will be more accurate if the volume of R is large. However, the opposite occurs for the second assumption since it states that $p(\mathbf{x})$ is constant inside this region. If R is too large, the estimated $p(\mathbf{x})$ could be over-smoothed. Therefore, there is

a choice to be made regarding both assumptions. For a given set of data, it is expected that there will be some optimum value for the size of R that will give the best estimation of $p(\mathbf{x})$ (Bishop, 1995).

This approximation is implemented by KNN. The value of K is fixed and the corresponding volume V is determined from the data. A hypersphere centred at a point \mathbf{x} is considered. Its radius grows until it contains K data points. Then, the value of $p(\mathbf{x})$ is obtained from Eq. (57). The KNN method can be used to construct classifiers through the use of Bayes' theorem. It involves modelling the class-conditional density $p(\mathbf{x}|\omega_i)$ of each class ω_i according to the described approach. The following expression would be obtained:

$$p(\mathbf{x}|\omega_i) \approx \frac{K_i}{N_i V}, \quad (58)$$

where K_i represents the total number of samples belonging to class ω_i that are found in a volume V centred on \mathbf{x} , given that N_i is the total number of training samples of class ω_i . The priors can be estimated using

$$P(\omega_i) \approx \frac{N_i}{N}. \quad (59)$$

Assuming that the unconditional distribution $p(\mathbf{x})$ is given by Eq. (57), the posterior probability $p(\omega_i|\mathbf{x})$ can be obtained from the Bayes' theorem (Bishop, 1995):

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})} = \frac{K_i}{K}. \quad (60)$$

Once posterior probabilities are known, pattern classification is carried out by applying the Bayes' decision rule. In practice, it is only required to identify the number of nearest neighbours to the input pattern (K). Then, posterior probabilities are computed by counting the number of neighbours in each class. KNN classifiers define non-linear decision boundaries in the input space.

6.3. Fisher's linear discriminant (FLD)

FLD is a technique for dimensionality reduction that can be used for pattern classification. FLD linearly projects input patterns \mathbf{x} onto a one-dimensional space. The projected value is obtained as:

$$y = \mathbf{w}^T \mathbf{x}, \quad (61)$$

where \mathbf{w} denotes the set of adaptive parameters (Fisher, 1936). It defines the direction of the projection. The elements of this vector are selected in order to maximise the separation between classes, which is defined as the separation between the projected mean vectors $\boldsymbol{\mu}_i$ of patterns belonging to class ω_i . Considering a two-class classification problem, the weight vector \mathbf{w} must maximise:

$$\mu_2 - \mu_1 = \mathbf{w}^T (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \quad (62)$$

where

$$\mu_i = \mathbf{w}^T \boldsymbol{\mu}_i \quad (63)$$

is the class mean of the projected data from class ω_i . The expression in Eq. (62) can be made arbitrarily large by increasing the magnitude of \mathbf{w} . Thus, this vector is constrained to have unit length. On the other hand, the optimisation problem must take into account the within-class spread of the data. Maximising the distance between projected means does not ensure to improve the separability between classes. Undesired solutions for distributions like those shown in Figure 11 could be produced.

As can be observed, projection on x_1 -axis allows to maximise the distance between the transformed means. However, the projected data is considerably overlapped. The Fisher's discriminant criterion suggests preventing these situations by maximising a function which represents the difference between the projected class means, normalised by a measure of the within-class scatter along direction of \mathbf{w} (Jobson, 1991b; Bishop, 1995). The within-class variance of the projected patterns from class ω_i is computed as:

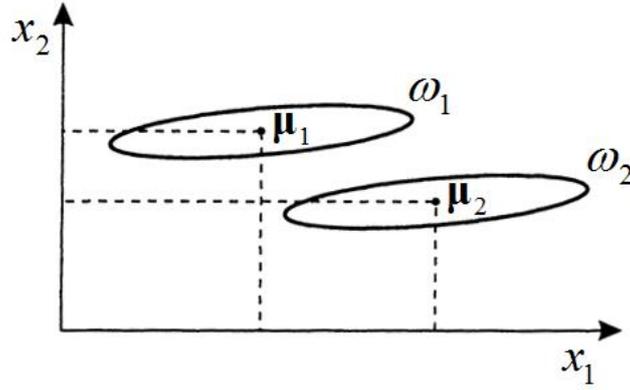


Figure 11. Example of the distribution of samples in two different classes. A high overlap is obtained if samples are projected in the direction for which the maximum difference between means is obtained (Bishop, 1995).

$$\sigma_i^2 = \sum_{n \in \omega_i} (y^n - m_i)^2 \quad (64)$$

and the total within-class covariance can be defined as $\sigma_1^2 + \sigma_2^2$, for the two-class problem. Therefore, the Fisher's criterion consists in maximising the expression given by (Bishop, 1995):

$$J(\mathbf{w}) = \frac{(\mu_2 - \mu_1)^2}{\sigma_1^2 + \sigma_2^2}, \quad (65)$$

which can be written as

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}, \quad (66)$$

where \mathbf{S}_B is the between-class covariance matrix and \mathbf{S}_W is the total within-class covariance matrix. They are computed as:

$$\mathbf{S}_B = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \quad (67)$$

and

$$\mathbf{S}_W = \sum_{n \in \omega_1} (\mathbf{x}^n - \boldsymbol{\mu}_1)(\mathbf{x}^n - \boldsymbol{\mu}_1)^T + \sum_{n \in \omega_2} (\mathbf{x}^n - \boldsymbol{\mu}_2)(\mathbf{x}^n - \boldsymbol{\mu}_2)^T, \quad (68)$$

respectively.

If function $J(\mathbf{w})$ is differentiated with respect to \mathbf{w} , the following condition is obtained:

$$(\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} = (\mathbf{w}^T \mathbf{S}_W \mathbf{w}) \mathbf{S}_B \mathbf{w}. \quad (69)$$

Multiplying both sides by \mathbf{S}_W^{-1} , the previous equation can be expressed as:

$$\mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w} = \lambda \mathbf{w}, \quad (70)$$

which corresponds to an eigenvalue problem. The optimum direction of projection \mathbf{w} is given by the eigenvector of matrix $\mathbf{S}_W^{-1} \mathbf{S}_B$ associated to the largest eigenvalue λ (Jobson, 1991b).

The Fisher's discriminant requires to evaluate the distance from the projection of the input patter to the projected class means to determine its membership. The following rule is applied (Jobson, 1991b):

$$\text{Decide } \omega_i \text{ for } \mathbf{x} \text{ if } |\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \boldsymbol{\mu}_i| = \min_i |\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \boldsymbol{\mu}_i| \quad (71)$$

The FLD defines a linear decision boundary in the input space. The hyperplane is represented by vector \mathbf{w} . It should be noted that any assumption about prior probabilities on group membership is required. Under the hypothesis of multivariate normality with common covariance matrix, the Fisher's discriminant is equivalent to decide the group for which the class-conditional density is the highest (Jobson, 1991b).

6.4. Logistic regression (LR)

LR models are a special case of generalised linear models, which represent a generalisation of linear regression. These models linearly relates a set of explanatory variables or features (grouped in pattern \mathbf{x}) with a response variable (t) via a link function. Each outcome of the response variable t is assumed to be generated from a probability distribution of the exponential family. In LR modelling, t is a dummy variable used to represent two possible categories corresponding to $t = 1$ (ω_1) or $t = 0$ (ω_0), respectively. Thus, its probability density function is modelled by a binomial (Bernoulli) distribution given by:

$$p(t|\pi) = \pi^t (1-\pi)^{(1-t)}, \quad (72)$$

where $\pi = E\{t\}$ is the probability associated to $t = 1$ (Jobson, 1991b).

The aim of LR is to express π as a function of \mathbf{x} . It is assumed that the value of π depends on the linear combination of the input features, i.e. $\pi = \pi(l)$, with $l = w_0 + \sum_{i=1}^d w_i x_i$ and $\mathbf{w} = (w_0, w_1, \dots, w_d)$ representing the set of model adaptive parameters. This functional dependence is modelled by LR using the logistic function (Jobson, 1991b):

$$\pi(l) = \frac{e^l}{1+e^l}. \quad (73)$$

Using the logit transformation $\ln[\pi/(1-\pi)]$ as the link function, a linear dependence is obtained between the input pattern \mathbf{x} and the variable π .

$$\ln \left[\frac{\pi(l)}{1-\pi(l)} \right] = \ln \left[\frac{e^l/(1+e^l)}{1/(1+e^l)} \right] = l. \quad (74)$$

Classification algorithms based on LR are usually applied to two-class problems. The LR model estimates the probability that a pattern \mathbf{x} correspond to

category represented by $t = 1$, which is given by the value of π . Therefore, the Bayes' decision rule can be applied since posterior probabilities are directly obtained from Eq. (72).

The maximum likelihood criterion is used to estimate the vector of model parameters \mathbf{w} from training data. According to Eq. (72), the likelihood for the n th sample in the training set is given by:

$$\pi_n^{t_n} (1 - \pi_n)^{(1-t_n)} = [\pi(\mathbf{w}\mathbf{x}_n)]^{t_n} [1 - \pi(\mathbf{w}\mathbf{x}_n)]^{(1-t_n)}, \quad (75)$$

where it has been assumed that pattern \mathbf{x} contains an additional term equal to unit in order to reflect the dependence of π on the product between \mathbf{w} and \mathbf{x} . Therefore, the likelihood (L_D) for the entire training set is given by:

$$L_D = \prod_{n=1}^N [\pi(\mathbf{w}\mathbf{x}_n)]^{t_n} [1 - \pi(\mathbf{w}\mathbf{x}_n)]^{(1-t_n)}, \quad (76)$$

where independency among samples has been assumed (Jobson, 1991b). Transforming this expression by the negative natural logarithm leads to:

$$-\ln L_D = -\sum_{n=1}^N \left\{ t_n \ln [\pi(\mathbf{w}\mathbf{x}_n)] \right\} + (1-t_n) \ln [1 - \pi(\mathbf{w}\mathbf{x}_n)], \quad (77)$$

which is the expression to be minimised. There is no analytical solution for this optimisation problem and an iterative process is required. The negative logarithm of the likelihood has a single minimum, i.e. the solution for the LR model is unique. It is possible to initialise \mathbf{w} close to this minimum. Usually, the iterated re-weighted least squares (IRLS) algorithm is used to find the parameters of the LR classifier (Nabney, 2002). It ensures a rapid optimisation process according to the maximum likelihood principle.

6.5. Probabilistic neural networks (PNN)

PNN implement the Bayes' decision rule for pattern classification. These networks estimate the probability density function associated to each class by applying the Parzen window procedure (Specht, 1990). To illustrate this technique, consider a region R limited by a hypercube with sides of length σ_p and centred on point \mathbf{x} . Its volume is given by:

$$V = \sigma_p^d. \quad (78)$$

The expression for K , the number of points that fall within R , can be obtained by defining a kernel function $H(\mathbf{u})$, also known as Parzen window, which is given by:

$$H(\mathbf{u}) = \begin{cases} 1 & |u_j| < 1/2 \quad j=1, \dots, d, \\ 0 & \text{otherwise} \end{cases}, \quad (79)$$

i.e., $H(\mathbf{u})$ corresponds to a unit hypercube centred at the origin. For a data point \mathbf{x} , the quantity $H\left[\frac{\mathbf{x} - \mathbf{x}^n}{\sigma_p}\right]$ is equal to unity if the point \mathbf{x} falls inside a hypercube of side σ_p centred on \mathbf{x}^n , and is zero otherwise. The total number of points falling inside the hypercube is then:

$$K = \sum_{n=1}^N H\left(\frac{\mathbf{x} - \mathbf{x}^n}{\sigma_p}\right). \quad (80)$$

Substituting this value of K into Eq. (57) to approximate the probability density function of the data leads to the following expression:

$$p(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N \frac{1}{\sigma_p^d} H\left(\frac{\mathbf{x} - \mathbf{x}^n}{\sigma_p}\right), \quad (81)$$

which can be viewed as the superposition of N cubes of side σ_p , with each cube centred on one of the data points in the training set.

The resulted function presents discontinuities due to the shape of $H(\mathbf{u})$. Different forms of the kernel function can be used to smooth out the estimation of $p(\mathbf{x})$. In general, the kernel must meet the following conditions (Bishop, 1995):

$$H(\mathbf{u}) \geq 0 \quad (82)$$

and

$$\int H(\mathbf{u}) d\mathbf{u} = 1. \quad (83)$$

As a result, the estimated $p(\mathbf{x})$ will satisfy $p(\mathbf{x}) \geq 0$ and $\int p(\mathbf{x}) d\mathbf{x} = 1$. A common choice is a multivariate Gaussian kernel, which is usually employed to implement PNN algorithms (Specht, 1990). These networks compute the Parzen window estimation for the probability density function of data in class ω_i from training samples corresponding to this class. As a result, the following expression is obtained:

$$p(\mathbf{x}|\omega_i) = \frac{1}{N_i} \sum_{n=1}^{N_i} \frac{1}{(2\pi\sigma_p^2)^{d/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}^n\|^2}{2\sigma_p^2}\right), \quad (84)$$

where a kernel function is centred on each training pattern corresponding to class ω_i and N_i denotes the number of training patterns in that class. Posterior probabilities can be directly obtained by applying the Bayes' theorem, as expressed in Eq. (60) for KNN classifiers. Similarly, prior probabilities $P(\omega_i)$ can be estimated as the proportion of samples in each class, Eq. (59). Finally, new input patterns are classified by the PNN network according to the Bayes' decision rule.

6.6. Multilayer perceptron (MLP) networks for classification

MLP is the most widely studied and used neural network. MLP networks are models for expressing knowledge using a connectionist paradigm inspired in the human brain. They present a massive interconnected architecture composed of simple units that imitate biological neurons. A perceptron is used to implement each of these

basic computational units (Haykin, 1996). The perceptron was conceived by Rosenblatt in 1958 (Rosenblatt, 1958). It is illustrated in Figure 12. Its input is a d -dimensional vector. Each element of the vector is fed into one of the input connections, each of which has associated a synaptic weight. The perceptron forms a weighted sum of the d components and adds a bias value. The result is then passed through a non-linearity to obtain the output or activation value, which is given by (Hush and Horne 1993):

$$z_j = g_{act} \left(\sum_{i=1}^d w_{ij} x_i + b_j \right), \quad (85)$$

where j denotes a perceptron in the MLP network, z_j is its activation value, w_{ij} is the weight associated to input connection i , b_j is the bias and $g_{act}(\cdot)$ denotes the non-linear function referred to as activation function. Commonly, a sigmoid function (S-shaped function) is employed for implementing the non-linearity.

Perceptrons in a MLP network are referred to as neurons, nodes or units. They are arranged in several interconnected layers. The input pattern feeds into each of the neurons in the first layer (i.e. the first hidden layer), which compute their output value as described. The outputs of this layer feeds into each of the second layer neurons, resulting in a fully connected architecture for which every node in a layer is connected to every node in the next layer. Finally, the set of output signals of nodes in the final layer (output layer) constitutes the response of the network to the input pattern (Haykin, 1999). A generic MLP network with one hidden layer composed of N_H hidden neurons and N_O neurons in the output layer is represented in Figure 13.

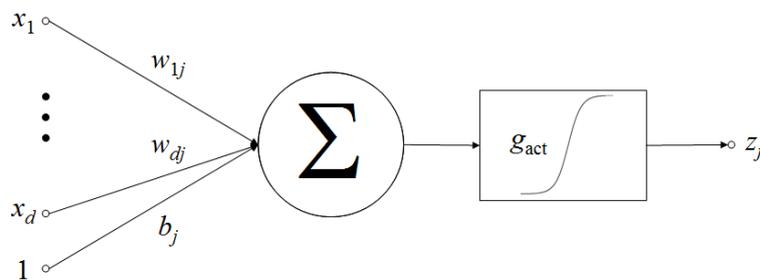


Figure 12. Scheme of a perceptron. Its output is obtained from the weighted sum of the input features, including a bias term, and a non-linear activation function.

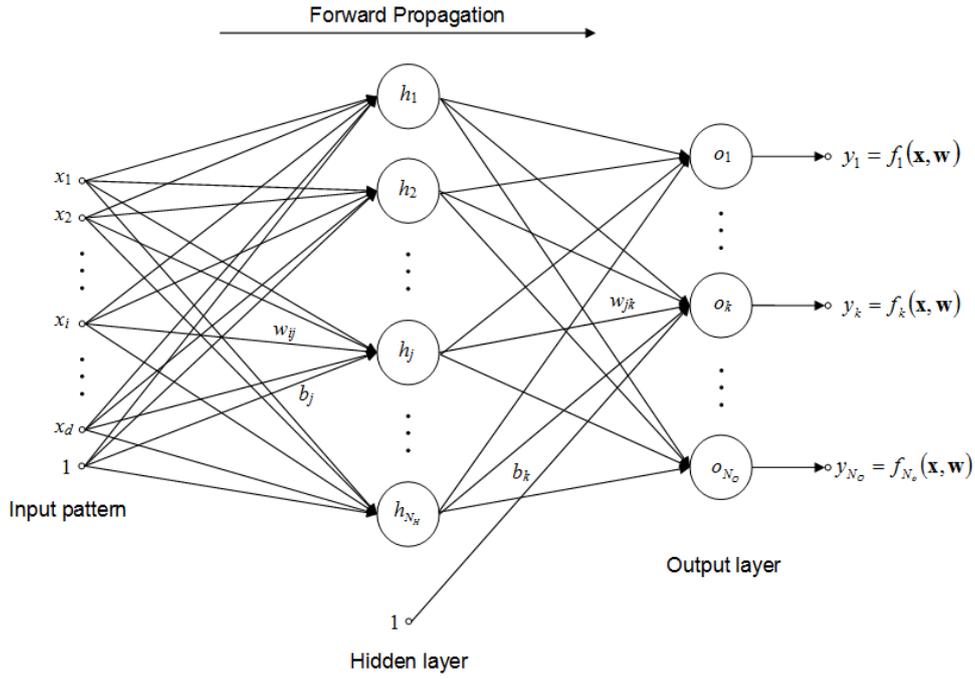


Figure 13. Scheme of a generic MLP network with a single hidden layer with N_H hidden neurons and an output layer with N_O output neurons.

Each network connection between two neurons is associated with a network adaptive parameter or weight (w_{ij}). The weight vector (\mathbf{w}), composed of all the adaptive parameters (weights and biases) in the network, is optimised during training. The statistical properties of the problem are inferred into network weights. As can be observed, there is no feedback loops in the network and the signal is propagated from the input layer to neurons in the output layer in a single direction. This ensures that the network outputs can be calculated as explicit functions of the inputs and the weights (Bishop, 1995). The value (y_k) provided by an output neuron (o_k) in the depicted network is given by:

$$y_k = f_k(\mathbf{x}, \mathbf{w}) = g_l \left\{ \sum_{j=1}^{N_H} \left[w_{jk} g_t \left(\sum_{i=1}^d w_{ij} x_i + b_j \right) + b_k \right] \right\}, \quad (86)$$

where d is the number of features in the input vector, N_H is the number of hidden units, w_{jk} is the weight connecting hidden unit h_j with output unit o_k , b_k is the bias associated with output unit o_k , w_{ij} is the weight connecting the feature i of the input pattern with hidden unit h_j , b_j is the bias associated with hidden unit h_j , $g_t(\cdot)$ is the activation function for units in the hidden layer and $g_l(\cdot)$ is the activation function for the output layer unit.

Network weights are adjusted during training from a data set D with pairs $(\mathbf{x}^n, \mathbf{t}^n)$ consisting of an input pattern and its associated output value. The aim of training is to model the underlying generator of the data (Bishop, 1995). Its complete description is given by the probability density $p(\mathbf{x}, \mathbf{t})$ in the joint input-target space. It can be expressed as:

$$p(\mathbf{x}, \mathbf{t}) = p(\mathbf{t}|\mathbf{x})p(\mathbf{x}), \quad (87)$$

where $p(\mathbf{t}|\mathbf{x})$ denotes the probability density of \mathbf{t} given that \mathbf{x} takes a particular value and $p(\mathbf{x})$ represents the unconditional density of \mathbf{x} . As explained later, the term $p(\mathbf{x})$ is not relevant for implementing MLP networks. These algorithms are used to model the conditional density $p(\mathbf{t}|\mathbf{x})$ in order to provide appropriate predictions for \mathbf{t} given the input \mathbf{x} . For classification problems, the target \mathbf{t} represents labels defining class membership $(\omega_1, \dots, \omega_c)$. MLP classifiers are capable of estimating posterior probabilities $p(\omega_i|\mathbf{x})$. Thus, the Bayes' decision rule can be applied to determine the membership of the input pattern in order to minimise the risk in the decision.

An appropriate design of the MLP classifier is required to provide the desired output representation. Considering mutually exclusive classes, an output neuron is used to represent each of the possible categories. Thus, the dimension of the output is c (i.e., $N_O = c$), the number of classes in the problem. A 1-of- c target coding scheme is adopted for training. To identify a given category, only one of the elements of the target vector \mathbf{t} , composed of c binary variables, is set to 1 while the others are set to 0. In order to be interpreted as probabilities, network outputs must range from 0 to 1 and must sum to unity. This can be achieved by using a sigmoid activation function known as softmax, which takes the form (Bishop, 1995):

$$y_k = \frac{\exp(a_k)}{\sum_{j=1}^c \exp(a_j)}, \quad (88)$$

where a_k is the activation value of output neuron o_k . In this study, the problem of SAHS diagnosis was modelled as a two-class classification task. Subjects are assigned to one

of two possible classes: SAHS-negative or SAHS-positive. According to the described procedure, the architecture of the MLP should contain two output neurons with the softmax activation function. This is equivalent to implement a single output neuron with a logistic activation function, which is a particular case of the softmax function. The logistic function takes the form:

$$y = \frac{1}{1 + \exp(-a)}, \quad (89)$$

which ranges between 0 and 1 for input values comprised between $-\infty$ and $+\infty$.

MLP network classifiers implemented in the Thesis were trained to estimate the probability that an input pattern \mathbf{x} correspond to a subject affected by SAHS, i.e. the output of the network (y) represents $p(\omega_1|\mathbf{x})$ where ω_1 is the value of the target t used to denote the SAHS-positive group. Therefore, $p(\omega_2|\mathbf{x}) = 1 - y$, where ω_2 is the value of the target t for patterns belonging to the SAHS-negative group. The coding scheme used to identify both groups was $t = 1$ for ω_1 and $t = 0$ for ω_2 . Thus, in a two-class classification context, t is a univariate dichotomous variable that can be modelled by a Bernoulli distribution (Bishop, 1995). The probability of observing the target value t when the input pattern \mathbf{x} is known can be modelled as:

$$p(t|\mathbf{x}) = y^t (1 - y)^{1-t}, \quad (90)$$

where the parameter of the distribution is given by the network output. Using this interpretation and assuming that training samples are statistically independent, the likelihood of observing the pairs of data in the training set D is given by:

$$L_D = \prod_{n=1}^N p(\mathbf{x}^n, t^n) = \prod_{n=1}^N p(t^n | \mathbf{x}^n) p(\mathbf{x}^n) = \prod_{n=1}^N (y^n)^{t^n} (1 - y^n)^{1-t^n} p(\mathbf{x}^n). \quad (91)$$

The maximum likelihood principle was applied to train MLP networks. Network weights are adjusted in order to maximise L_D . It is equivalent to minimise the negative logarithm of the likelihood, which leads to the error function:

$$E_D = -\ln L_D = -\sum_{n=1}^N \left[t^n \ln y^n + (1-t^n) \ln(1-y^n) \right] - \sum_{n=1}^N \ln p(\mathbf{x}^n). \quad (92)$$

The second term does not depend on the network parameters and can be removed. As a result, the error function to be minimised is expressed as:

$$E_D = -\sum_{n=1}^N \left[t^n \ln y^n + (1-t^n) \ln(1-y^n) \right], \quad (93)$$

which is known as the cross-entropy error function (Bishop, 1995).

There is no analytical solution for this optimisation problem due to the non-linear dependence of the error with respect to network weights. Their optimum value cannot be found directly and an iterative procedure must be used instead. Second-order non-linear optimisation algorithms are used for this purpose. Most of them are based on the backpropagation approach (Rumelhart and McClelland 1986), which provides an efficient means to compute the derivatives of the error with respect to network weights. Both the output (y) and the error (E_D) are differentiable functions of the weights when sigmoidal activation functions are used. For neurons in a hidden layer, the derivative of the error with respect to one of its weights is computed from derivatives related to weights acting on the output of such neuron, i.e. weights in the following layer. Thus, the term backpropagation denotes the propagation of the error from the output layer to units in previous layers.

Optimisation algorithms tend to present an iterative procedure for adjusting weights. At each step of the iteration process, two distinct operations can be distinguished. First, the derivatives of the error function with respect to the weights are evaluated. Second, these derivatives are used to compute the re-adjustment to be made for each of the weights (Bishop, 1995). The simplest technique is gradient descent (Rumelhart and McClelland 1986). The first stage is used to find the direction for which the variation of the error in the space defined by the network weights is the highest. The gradient of the error function is computed to determine this direction. In the second stage, weights are re-adjusted by descending on the error surface along the direction of the gradient.

A penalty term is usually added to the error function in order to apply regularisation. Excessively complex networks with respect to the training set (high variance) tend to overfit the data, resulting in poor generalisation. Regularisation techniques enable to control the effective complexity of the network. Smooth mappings are encouraged in order to avoid overfitting. Weight decay is a common form of regulariser (Bishop, 1995). The added penalty term is equal to the sum of the squares of the adaptive network parameters in order to favour small values for them. As a result, the new error function to be minimised is given by the following expression:

$$E = E_D + \nu \sum_i w_i^2 = - \sum_{n=1}^N \left[t^n \ln y^n + (1-t^n) \ln (1-y^n) \right] + \nu \sum_i w_i^2, \quad (94)$$

where the parameter ν , referred to as the regularisation parameter, controls the extent to which the penalty term influences the form of the solution (Moody, 1992).

6.7. Radial basis function (RBF) networks for classification

RBF networks represent a different neural network approach in which the activation of a hidden unit is determined by the distance between the input vector and a prototype vector. These methods have their origins in techniques for exact interpolation of a data set composed of points in a multi-dimensional space (Bishop, 1995). Initially, consider a mapping from a d -dimensional input space \mathbf{x} to a one-dimensional target space t . The data set D is composed of N input-output pairs of the form (\mathbf{x}^n, t^n) . The goal is to find a function $h(\mathbf{x})$ for which the following equality must hold:

$$h(\mathbf{x}^n) = t^n, \quad n = 1, \dots, N. \quad (95)$$

The RBF approach defines a set of N basis functions, one for each data point, which take the form $\psi(\|\mathbf{x} - \mathbf{x}^n\|)$, where $\psi(\cdot)$ is some non-linear function. Thus, the argument of each basis function depends on the distance between the input vector and the data point related to that function. Usually, the Euclidean distance is considered. The output of the mapping is then taken to be a linear combination of the basis functions (Bishop, 1995):

$$h(\mathbf{x}) = \sum_{n=1}^N w_n \psi(\|\mathbf{x} - \mathbf{x}^n\|), \quad (96)$$

where w_n denotes an adaptive model parameter or weight associated to the n th basis function. Several forms of the basis function can be considered. The most common one is the Gaussian function:

$$\psi(x) = \exp\left(-\frac{x^2}{2\sigma_r^2}\right), \quad (97)$$

where the variance σ_r is a parameter for controlling the smoothness properties of the interpolating function (Bishop, 1995).

Mapping functions based on exact interpolation have two main limitations. First, the exact interpolation for noisy data tends to be a highly oscillatory function, i.e. the noise present in the data is captured by the mapping function, resulting in poor generalisation. A smoother functional form is desired to obtain good predictions for new unseen data. Second, the number of basis functions is equal to the number of sample patterns. For large data sets, the mapping function could become very costly to evaluate. The RBF network is defined by introducing some modifications to the described exact interpolation procedure (Broomhead and Lowe 1988; Moody and Darken 1989). This provides a smooth interpolating function in which the number of basis functions is determined by the complexity of the mapping to be represented rather than by the size of the data set. These required modifications are as follows (Bishop, 1995):

1. The number of basis functions does not need to equal the number of data points, which is typically much higher.
2. The centres of the basis functions are not constrained to be given by input data vectors. Instead, the determination of suitable centres becomes part of the training process.

3. Instead of having a common width parameter σ_r , each basis function is given its own width σ_{rj} for which the value is determined during training.
4. Bias parameters are included in the sum to obtain the network output. They compensate for the difference between the average value over the data set of the basis functions and the corresponding average value of the targets.

When these changes are made to the exact interpolation formula, the mapping of the RBF network is expressed as:

$$y_k(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{N_B} w_{jk} \psi_j(\|\mathbf{x} - \boldsymbol{\mu}_j\|) + b_k, \quad (98)$$

where N_B is the number of basis functions, $\boldsymbol{\mu}_j$ is the centre of the basis function ψ_j , w_{jk} is the weight connecting basis function ψ_j with output neuron o_k and b_k is the bias parameter. The architecture of the network is similar to that of a MLP with a single hidden layer depicted in Figure 13. In RBF networks, the hidden units compute their output from the distance of the input pattern to a prototype pattern. Then, it is used as the argument of a Gaussian function. The output layer performs the same operation as in MLP networks.

In classification problems, RBF networks are used to model posterior probabilities. Using the Bayes' theorem, the expression of the posterior probability $p(\omega_i|\mathbf{x})$ for class ω_i is given by (Bishop, 1995):

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_i^c p(\mathbf{x}|\omega_i)P(\omega_i)}. \quad (99)$$

It can be viewed as a simple form of basis function network in which a single kernel $p(\mathbf{x}|\omega_i)$ was used to model data in class ω_i . Basis functions in this network would be expressed as:

$$\psi_i(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)}{\sum_{i'}^c p(\mathbf{x}|\omega_{i'})P(\omega_{i'})} \quad (100)$$

and second-layer connections would consist of one weight with value $P(\omega_i)$ from each hidden unit going to the corresponding output unit.

According to this modelling scheme, the outputs of the network estimate posterior probabilities. However, a good representation of class-conditional density $p(\omega_i|\mathbf{x})$ is not expected by using a single kernel function. Alternatively, a separate mixture model could be used to represent each of the conditional densities. This could lead to an excessive number of adjustable parameters. To avoid this limitation, a common set of N_B basis functions, labelled by an index j , are used to represent all conditional densities. Each of them is written as (Bishop, 1995):

$$p(\mathbf{x}|\omega_i) = \sum_{j=1}^{N_B} p(\mathbf{x}|j)p(j|\omega_i). \quad (101)$$

The unconditional density $p(\mathbf{x})$ can then be obtained by summing over all classes:

$$p(\mathbf{x}) = \sum_{i=1}^c p(\mathbf{x}|\omega_i)P(\omega_i) = \sum_{j=1}^{N_B} p(\mathbf{x}|j)P(j), \quad (102)$$

where $P(j)$ denotes the prior for the basis function ψ_j . It is defined as:

$$P(j) = \sum_{i=1}^c p(j|\omega_i)P(\omega_i). \quad (103)$$

Substituting expressions in Eq. (101) and (102) into the Bayes' theorem given by Eq. (99), the posterior probabilities of class membership are obtained as (Bishop, 1995):

$$p(\omega_i|\mathbf{x}) = \frac{\sum_{j=1}^{N_B} p(j|\omega_i) p(\mathbf{x}|j) P(\omega_i)}{\sum_{j'=1}^{N_B} p(\mathbf{x}|j') P(j')} \frac{P(j)}{P(j)} = \sum_{j=1}^{N_B} w_{ji} \psi_j(\mathbf{x}). \quad (104)$$

This expression represents a RBF network, in which the normalised basis functions are given by:

$$\psi_j(\mathbf{x}) = \frac{p(\mathbf{x}|j) P(j)}{\sum_{j'=1}^{N_B} p(\mathbf{x}|j') P(j')} = p(j|\mathbf{x}). \quad (105)$$

The weights connecting these functions with neurons in the output layer are given by:

$$w_{ji} = \frac{p(j|\omega_i) P(\omega_i)}{P(j)} = p(\omega_i|j). \quad (106)$$

Thus, the activation of a basis functions can be interpreted as the posterior probability for the presence of the feature represented by such function conditioned on the input to the unit. Similarly, the weights can be interpreted as the posterior probability of class membership given the presence of the feature represented by basis function ψ_j (Bishop, 1995).

One of the main advantages of RBF networks, when compared with MLP models, is the possibility of choosing appropriate parameters for the basis functions without having to perform a full non-linear optimisation of all the network parameters. The sum of the basis functions represents the unconditional density of the input data (Bishop, 1995; Nabney, 2002). Therefore, training RBF networks can be performed in two different phases. In the first one, an unsupervised learning procedure can be used to choose the basis function parameters, i.e. centres ($\boldsymbol{\mu}_j$) and widths (σ_{rj}) of the Gaussian kernels. The centres of the basis functions can be regarded as prototypes of the input patterns. Different approaches for setting the function centres are available. One simple procedure is to set them equal to a random subset of the training samples. Clustering techniques represent a more advanced approach. Algorithms such as k -means or fuzzy

c -means can be used to find a set of centres which more accurately reflect the distribution of the data (Marcos *et al.* 2008a). Finally, the estimation of the unconditional density of the data from the combination of several Gaussian kernels can be modelled as a mixture density estimation problem. The basis functions of the neural network can be regarded as the components of a mixture density model. Hence, the centres of the Gaussians can be obtained by maximum likelihood optimisation (Bishop, 1995). Once the centres are fixed, the width of each function must be selected. Usually, a different width is selected for each basis function. For instance, the value of the width can be determined from the average distance of each basis function to its K nearest neighbours.

The second phase of training focuses on the optimisation of weights in the output layer. It is performed from the projections of the input patterns on the space spanned by the basis functions and their corresponding target values. A single output neuron is required for two-class classification problems. This neuron must be characterised by a logistic activation function in order to interpret network outputs as probabilities (Nabney, 2004). As a result, adjusting output weights becomes a non-linear optimisation problem. General purpose optimisation algorithms as those used for MLP networks could be applied to RBF training. However, a considerable speed advantage could be obtained using techniques derived from generalised linear models. As described for LR models, weights in the output layer of RBF networks can be optimised according to the maximum likelihood principle by using the IRLS algorithm (Nabney, 2004). Indeed, the output layer of the RBF is equivalent to a LR model if the space spanned by the basis functions is viewed as its input space. As a result, the time required to train RBF networks is much less than that required for MLP networks.

6.8. Bayesian (BY) networks for classification

MLP networks described before are implemented according to the principle of maximum likelihood. Network weights are adjusted in order to minimise an error function. In this context, more complex models are typically better able to fit the training data. However, this does not necessarily imply good generalisation capability (Bishop, 1995). The Bayesian framework provides an alternative approach to implement MLP algorithms. It is based on the prior assumption that smooth mappings

provide better generalisation. Moreover, Bayesian techniques can be used to evaluate the relevance of each input variable for the network predictions (Neal, 1996; Nabney, 2002).

The Bayesian approach models the posterior probability density function of the weight vector rather than determining an optimum set of network weights (Bishop, 1995; Nabney, 2002). When the maximum likelihood approach is applied, different representative training sets lead to different network weights. Bayesian techniques aim to account for this uncertainty by representing the degrees of belief in the values of the weight vector with a probability distribution (Bishop, 1995). According to Bayes' theorem, the posterior distribution of the weights (\mathbf{w}) given the training set (D) is expressed as:

$$p(\mathbf{w}|D) = \frac{p(D|\mathbf{w})p(\mathbf{w})}{p(D)}, \quad (107)$$

where $p(\mathbf{w})$ is the prior probability function over weight space, $p(D|\mathbf{w})$ is the likelihood of the training data and $p(D)$ is a normalisation factor known as the evidence (Nabney, 2002). Once the posterior has been calculated, it can be used to infer the distribution of output values by computing the following expression:

$$p(t|\mathbf{x}, D) = \int p(t|\mathbf{x}, \mathbf{w})p(\mathbf{w}|D)d\mathbf{w}, \quad (108)$$

where $p(t|\mathbf{x}, \mathbf{w})$ is the model for the distribution of the noise on the target for a given weight vector. In pattern classification problems, it is given by the expression in Eq. (90). Therefore, the probability of membership of an input pattern to the SAHS-positive group is obtained as:

$$p(\omega_1|\mathbf{x}, D) = \int p(\omega_1|\mathbf{x}, \mathbf{w})p(\mathbf{w}|D)d\mathbf{w} = \int y(\mathbf{x}, \mathbf{w})p(\mathbf{w}|D)d\mathbf{w}. \quad (109)$$

The Gaussian approximation proposed by MacKay (1992a) was used to implement BY MLP networks. The posterior $p(\mathbf{w}|D)$ is modelled as a Gaussian function (Laplace approximation) used to solve integrals such as those in Eq. (108) and (109). Initially, the prior $p(\mathbf{w})$ and the likelihood function $p(D|\mathbf{w})$ must be defined. In the

absence of data, the prior probability distribution is chosen to favour small weights. Smooth mappings are preferred since they provide better generalisation by reducing variance and preventing overfitting (Bishop, 1995). Thus, the prior is modelled using the following zero-mean Gaussian function:

$$p(\mathbf{w}) = \frac{1}{Z_W} \exp(-\alpha E_W) = \frac{1}{Z_W} \exp\left(-\frac{\alpha}{2} \|\mathbf{w}\|^2\right), \quad (110)$$

where Z_W is a normalisation factor, $E_W = 1/2 \|\mathbf{w}\|^2$ and α , referred to as hyperparameter, is the inverse variance of the Gaussian. It controls the distribution of the network parameters, i.e., network weights and biases. The evaluation of the normalisation factor is straightforward and is given by (Bishop, 1995):

$$Z_W(\alpha) = \int \exp(-\alpha E_W) d\mathbf{w} = \left(\frac{2\pi}{\alpha}\right)^{W/2}, \quad (111)$$

where W is the number of adaptive parameters in the network. On the other hand, a binomial distribution is used for the target variable in classification problems. The likelihood function for the training data is given by the expression in Eq. (91). It can then be written as:

$$p(D|\mathbf{w}) = \prod_{n=1}^N y_n^{t_n} (1 - y_n)^{1-t_n} = \exp[-G(D|\mathbf{w})], \quad (112)$$

where $E_D = G(D|\mathbf{w})$ is the cross-entropy error function. According to Eq. (107), Bayes' theorem is applied to compute the posterior probability from $p(\mathbf{w})$ and $p(D|\mathbf{w})$. It is given by (Bishop, 1995):

$$p(\mathbf{w}|D) = \frac{1}{Z_S} \exp(-G - \alpha E_W) = \frac{1}{Z_S} \exp[-S(\mathbf{w})], \quad (113)$$

where Z_S is a normalisation factor and $S(\mathbf{w}) = G(D|\mathbf{w}) + \alpha E_W(\mathbf{w})$. MacKay (1992a) uses a Gaussian approximation for the posterior distribution. It is obtained by considering the

Taylor expansion of $S(\mathbf{w})$ around its minimum value (\mathbf{w}_{MP}), i.e., the weight vector that maximises the posterior $p(\mathbf{w}|D)$, and retaining terms up to second order centred:

$$S(\mathbf{w}) \approx S(\mathbf{w}_{MP}) + \frac{1}{2}(\mathbf{w} - \mathbf{w}_{MP})^T \mathbf{A}(\mathbf{w} - \mathbf{w}_{MP}), \quad (114)$$

where the linear term has vanished since the expansion was taken around a minimum of $S(\mathbf{w})$. The term \mathbf{A} is the Hessian matrix of the error function $S(\mathbf{w})$, with elements given by:

$$\mathbf{A} = \nabla \nabla S(\mathbf{w}_{MP}) = \nabla \nabla G(\mathbf{w}_{MP}) + \alpha \mathbf{I} = \mathbf{H} + \alpha \mathbf{I}, \quad (115)$$

where \mathbf{H} denotes the Hessian of the unregularised error function. The expansion given in Eq. (114) leads to a posterior distribution which is now a Gaussian function of the weights and is given by (Bishop, 1995):

$$p(\mathbf{w}|D) \approx \frac{1}{Z_S^*} \exp \left[-S(\mathbf{w}_{MP}) - \frac{1}{2} \Delta \mathbf{w}^T \mathbf{A} \Delta \mathbf{w} \right], \quad (116)$$

where $\Delta \mathbf{w} = \mathbf{w} - \mathbf{w}_{MP}$ and Z_S^* is the normalisation factor appropriate for the Gaussian. It is given by:

$$Z_S^*(\alpha) = \exp[-S(\mathbf{w}_{MP})] (2\pi)^{W/2} |\mathbf{A}|^{-1/2}. \quad (117)$$

At the moment, the hyperparameter α has been considered known. However, there is little idea about the optimum value of this hyperparameter. Additionally, such an optimum value that provides the best fit on the training set may result in poor generalisation. To make explicit its influence on the posterior distribution of the weights, it can be written as:

$$p(\mathbf{w}|D) = \int p(\mathbf{w}, \alpha|D) d\alpha = \int p(\mathbf{w}|\alpha, D) p(\alpha|D) d\alpha. \quad (118)$$

Two approaches for Bayesian treatment of hyperparameters have been proposed. One of them suggests performing the integral over α analytically. An

alternative approach is proposed by the evidence framework (MacKay, 1992b). Assuming that the probability density $p(\alpha|D)$ is sharply peaked around its most probable value α_{MP} , the posterior probability of the weights would be given by:

$$p(\mathbf{w}|D) \approx p(\mathbf{w}|\alpha_{MP}, D) \int p(\alpha|D) d\alpha = p(\mathbf{w}|\alpha_{MP}, D). \quad (119)$$

Once the value of α that maximises the posterior $p(\alpha|D)$ is found, the remaining calculations can be performed by keeping the hyperparameter fixed. The posterior distribution of α is given by:

$$p(\alpha|D) = \frac{p(D|\alpha)p(\alpha)}{p(D)}, \quad (120)$$

which requires a choice for the prior $p(\alpha)$. It is chosen to be very insensitive to the value of α to reflect little idea of suitable value for the hyperparameter (Bishop, 1995). Since the denominator in Eq. (120) is independent of α , the maximum-posterior value is found by maximising the likelihood term $p(D|\alpha)$. It is known as the evidence for α and can be written in the form (Bishop, 1995):

$$p(D|\alpha) = \int p(D|\mathbf{w}, \alpha) p(\mathbf{w}|\alpha) d\mathbf{w} = \int p(D|\mathbf{w}) p(\mathbf{w}|\alpha) d\mathbf{w}. \quad (121)$$

Using Eq. (110) and (112) for the prior and the likelihood together with Eq. (113) and (116), the evidence for α can be expressed as:

$$p(D|\alpha) = \frac{1}{Z_W(\alpha)} \int \exp[-S(\mathbf{w})] d\mathbf{w} = \frac{Z_S(\alpha)}{Z_W(\alpha)}. \quad (122)$$

The denominator for computing $p(D|\alpha)$ is given in Eq. (111) while the Gaussian approximation enables to estimate the value of $Z_S(\alpha)$ given in Eq. (117). The logarithm of the evidence yields the following expression:

$$\ln p(D|\alpha) = -\alpha E_W(\mathbf{w}_{MP}) - G(\mathbf{w}_{MP}) - \frac{1}{2} \ln |\mathbf{A}| + \frac{W}{2} \ln \alpha. \quad (123)$$

The derivative of this expression with respect to α is made equal to zero to find the maximum. It can be shown that the result can be approximated by (Bishop, 1995):

$$2\alpha E_W(\mathbf{w}_{MP}) = W - \sum_{i=1}^W \frac{\alpha}{\lambda_i + \alpha} = \gamma, \quad (124)$$

where λ_i denotes the eigenvalues of matrix \mathbf{H} and γ represents the number of well-determined parameters. It is given by:

$$\gamma \equiv \sum_{i=1}^W \frac{\lambda_i}{\lambda_i + \alpha}. \quad (125)$$

The practical implementation of the BY network requires to find optimum \mathbf{w}_{MP} and α_{MP} . The recommended approach is to define an initial value for α and use standard non-linear optimisation algorithms to determine \mathbf{w}_{MP} . Training is periodically halted for the hyperparameter to be updated using the actual value of \mathbf{w}_{MP} (Nabney, 2002). The new value of α is obtained as:

$$\alpha^{new} = \frac{\gamma}{2E_W(\mathbf{w}_{MP})}. \quad (126)$$

In addition, complex forms of the hyperparameter α can be specified, i.e., diverse values of α can be used to model the distribution of different groups of weights. Specifically, a different value of α can be used for the weights associated with each input feature in order to determine its relative influence on the network output. This procedure is known as automatic relevance determination (ARD) (Neal, 1996). Once the network has been trained, the importance of an input feature (x_i) can be evaluated by analysing its associated hyperparameter (α_i). A low value of α_i corresponds to a large variance of the prior, which allows weights of large magnitude. Thus, the feature x_i is very relevant for predicting the output. Conversely, a large value of α_i is interpreted as low influence of the input feature on network output. This procedure enables to implement feature selection in the network training process.

6.9. Support vector machines (SVM) for classification

SVM algorithms are inspired in the statistical learning theory developed by Vapnik (1999). They implement the method of structural risk minimisation. This induction principle assumes that the generalisation error is bounded by the sum of two terms: the error on the training set (empirical risk) and a term depending on the capacity (flexibility) of the learning technique. The latter is evaluated by means of the Vapnik-Chervonenkis (VC) dimension, which reflects the complexity of the model. According to the bias-variance trade-off, the structural risk minimisation principle suggests finding the optimum balance between the quality of the approximation and the complexity of the approximating function (Vapnik, 1999). SVM can provide good generalisation performance since both quantities are taken into account. As a result, SVM present lower sensitivity to the curse of dimensionality than other classification methods such as MLP or RBF networks. Only the error on training data is considered for training these networks while model complexity must be controlled by the designer and the use of regularisation techniques.

Consider the training set D composed of N input-output pairs (\mathbf{x}^n, t^n) , with each input $\mathbf{x} \in \mathfrak{R}^d$ and the corresponding output label t takes one of two possible values: ω_1 for identifying the SAHS-positive group and ω_2 for the SAHS-negative group. These are codified as $t = +1$ for ω_1 and $t = -1$ for ω_2 in order to implement SVM models. Assuming that training patterns are linearly separable, it is possible to find a hyperplane in the input space for which the following condition is satisfied:

$$t^n (\mathbf{w}^T \mathbf{x}^n + w_0) \geq 1, \quad (127)$$

where the vector \mathbf{w} and the bias w_0 defines the hyperplane. The optimum separating hyperplane (or maximal margin hyperplane) is the one that separates data without error and maximises the distance (r_o) to the closest point (i.e. maximises the distance between both groups) (Vapnik, 1999). It has been shown that this can be achieved by minimising the Euclidean norm of the vector \mathbf{w} , since r_o is given by (Haykin, 1999):

$$r_o = \frac{1}{\|\mathbf{w}\|}. \quad (128)$$

In the case of linearly non-separable data, the condition given in Eq. (127) is not met by all the patterns. A new set of non-negative slack variables $\xi^n \geq 0$ are introduced to measure the deviation of a data point from the ideal condition of pattern separability (Haykin, 1999). The optimum separating hyperplane is subject to:

$$t^n (\mathbf{w}^T \mathbf{x}^n + w_0) \geq 1 - \xi^n. \quad (129)$$

The misclassification error on the training set is modelled as the sum of the slack variables associated to patterns in the training set:

$$\Phi(\xi) = \sum_{n=1}^N \xi^n. \quad (130)$$

SVM are based on the definition of the optimum separating hyperplane for the classification problem at hand. However, the optimisation is not directly performed in the input feature space. SVM classifiers map input data into a higher dimensional space (\mathfrak{Z}) by means of a non-linear transformation $\mathbf{z} = \phi(\mathbf{x}) / \mathbf{z} \in \mathfrak{R}^{d_s}, d_s > d$. The optimum separating hyperplane is then built in the transformed space ($\mathbf{w}^T \mathbf{z} + w_0$) by minimising the functional (Vapnik, 1999)

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi^n \quad (131)$$

subject to the constraints:

$$t^n (\mathbf{w}^T \mathbf{z}^n + w_0) \geq 1 - \xi^n \text{ and } \xi^n \geq 0, \quad n = 1, \dots, N, \quad (132)$$

where C is a regularisation parameter. It controls the trade-off between the maximum margin of separation between classes and minimising the classification error (Vapnik, 1999). The functional of $\Phi(\mathbf{w}, \xi)$ represents an estimation of the structural risk to be minimised. The first term in Eq. (131) is related to the VC dimension of the SVM, which reflects the complexity of the algorithm. Thus, maximising the margin of

separation between classes leads to less complex models. The second term represents an upper bound on the number of training errors.

The above optimisation problem can be reformulated through a Lagrangian function (Cortes and Vapnik 1995). The Lagrange multipliers (η_1, \dots, η_N) can be found by maximising the expression

$$Q(\eta) = \sum_{n=1}^N \eta^n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \eta^n \eta^m t^n t^m K(\mathbf{x}^n, \mathbf{x}^m) \quad (133)$$

subject to the constraints:

$$\sum_{n=1}^N \eta^n t^n = 0 \text{ and } 0 \leq \eta^n \leq C, \quad n = 1, \dots, N. \quad (134)$$

The term denoted by $K(\cdot, \cdot)$ represents the inner-product kernel function in the transformed space, which is given by:

$$K(\mathbf{x}^n, \mathbf{x}^m) = (\mathbf{z}^n)^T \mathbf{z}^m = \boldsymbol{\Phi}^T(\mathbf{x}^n) \boldsymbol{\Phi}(\mathbf{x}^m) = \sum_{i=1}^d \phi_i(\mathbf{x}^n) \phi_i(\mathbf{x}^m), \quad (135)$$

where $\boldsymbol{\Phi}(\mathbf{x}) = [\phi_0(\mathbf{x}), \phi_1(\mathbf{x}), \dots, \phi_d(\mathbf{x})]$ is the image vector in the transformed space corresponding to pattern \mathbf{x} in the input feature space. The optimum solution for the weight vector is given by:

$$\mathbf{w} = \sum_{n=1}^N \eta^n t^n \boldsymbol{\Phi}(\mathbf{x}^n). \quad (136)$$

Therefore, only some of the training vectors, those for which the parameter η^n is non-zero, contribute to the definition of the decision boundary. These vectors are referred to as the support vectors. There are two types of support vectors (Vapnik and Chapelle, 2000). The first one consists of training samples with $\eta^n = C$ and satisfy $t^n(\mathbf{w}^T \mathbf{z}^n + w_0) < 1$. These are all the points that fall between the margins defined by the SVM classifier. Those training patterns that are misclassified are included in this group.

Support vectors of the second type correspond to those training patterns such that $0 < \eta^n < C$. These patterns satisfy $t^n(\mathbf{w}^T \mathbf{z}^n + w_0) = 1$, i.e., they fall on the margins defined by the SVM.

The value of the bias w_0 can be computed from the Karush-Kuhn-Tucker conditions (Haykin, 1999). For all support vectors with $0 < \eta^n < C$, the slack variable is zero and the following condition holds:

$$t^n (\mathbf{w}^T \mathbf{z}^n + w_0) - 1 = 0. \quad (137)$$

Averaging over all patterns satisfying this condition yields a numerically stable solution. Finally, the output of the SVM classifier is expressed as a function of data in the original space:

$$y = \sum_{n \in S} \eta^n t^n K(\mathbf{x}^n, \mathbf{x}) + w_0, \quad (138)$$

where S is a subset of the indices $\{1, \dots, N\}$ corresponding to the nonzero Lagrange multipliers η^n associated to the support vectors (Vapnik, 1999). The sign of the output value is used to determine the group of the input pattern. As can be observed, classifiers based on SVM divide the transformed space into two different subspaces by using a hyperplane. However, it is worth noting that it corresponds to a non-linear decision boundary in the input space due to the applied transformation. Therefore, SVM can be effectively used for modelling tasks in which a binary output is required. The formulated SAHS diagnosis problem suitably adapts to this situation.

The requirement on the kernel $K(\cdot, \cdot)$ is to satisfy the Mercer's theorem (Vapnik, 1999). Within this requirement, the choice of the kernel function corresponds to the designer. Several functions have been proposed for implementing SVM. Indeed, the kernel is used to identify the type of SVM algorithm. The most common ones are the polynomial learning machine, the two-layer perceptron and the RBF network (Haykin, 1999). The latter was considered in this study. The Gaussian kernel is given by:

$$K(\mathbf{x}^n, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}^n\|^2}{2\sigma_s^2}\right), \quad (139)$$

where the width parameter σ_s is specified a priori by the user. Unlike RBF networks, it is common to all the kernels.

SVM possesses some useful properties (Vapnik, 1999). Irrespective of how a SVM is implemented, the optimisation problem for constructing it is very fast and has a unique solution. Moreover, SVM represent a different solution for implementing a learning machine by automatically controlling model complexity (VC dimension). They differ from the conventional empirical risk minimisation principle used in the design of other methods such as MLP or RBF networks. Unless regularisation techniques are used, the complexity of these networks is controlled by keeping the number of free parameters (i.e. hidden units and input features) small. Therefore, it is required to define an appropriate network architecture and to find in that network the function that minimises the training error. Conversely, network architecture is automatically found in SVM models.

7. Pattern recognition methods for regression

- 7.1. Introduction
- 7.2. Multiple linear regression (MLR)
- 7.3. Generalised regression neural networks (GRNN)
- 7.4. Multilayer perceptron (MLP) networks for regression
- 7.5. Radial basis function (RBF) networks for regression
- 7.6. Bayesian (BY) networks for regression
- 7.7. Multivariate adaptive regression splines (MARS)
- 7.8. Support vector machines (SVM) for regression

7.1. Introduction

The second approach to model SAHS diagnosis was based on regression techniques. Diagnosis of SAHS is determined from AHI derived from nocturnal PSG. Regression methods can be used to model the dependence of AHI with respect to the features extracted from oximetry signals. In addition to provide automated diagnosis about SAHS, the estimated AHI supplies valuable information about the degree of severity of the disease. Several regression methods have been evaluated in the Thesis. They are described in the following subsections.

7.2. Multiple linear regression (MLR)

MLR is the most commonly applied statistical technique for modelling the dependence of a variable as a function of other variables. For the MLR model, the regression function is assumed to be linear, i.e. the mapping implemented by the algorithm takes the form:

$$y = f(\mathbf{x}, \mathbf{w}) = w_0 + w_1x_1 + \dots + w_dx_d = \mathbf{w}^T \mathbf{x}, \quad (140)$$

where $\mathbf{w} = (w_0, \dots, w_d)^T$ is a set of adaptive parameters. In order to approximate the regression function, \mathbf{w} is determined by minimising the sum-of-squares error according to the maximum likelihood criterion (Jobson, 1991a). Using matrix notation, the resulting equation for \mathbf{w} is expressed as:

$$\mathbf{X}^T \mathbf{t} = \mathbf{X}^T \mathbf{X} \mathbf{w} \quad \mathbf{X}^T \mathbf{t} = \mathbf{X}^T \mathbf{X} \mathbf{w}, \quad (141)$$

where rows of matrix \mathbf{X} are samples \mathbf{x} in the training set and elements of the column vector $\mathbf{t} = (t_1, \dots, t_N)^T$ are their corresponding target values. The solution is given by:

$$\mathbf{w} = \mathbf{X}^+ \mathbf{t}, \quad (142)$$

where $\mathbf{X}^+ = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is the pseudoinverse matrix of \mathbf{X} . A necessary and sufficient condition for finding a solution for \mathbf{w} is that training vectors \mathbf{x} , i.e. the rows of matrix \mathbf{X} , must be linearly independent (Jobson, 1991a).

7.3. Generalised regression neural networks (GRNN)

Similarly to PNN used for classification, GRNN models are based on the Parzen window procedure for probability density estimation. Usually, the regression function is modelled by assuming a specific functional form with a set of adaptive parameters. Then, it is fit to the training data by minimising the sum-of-squares error. GRNN models implement a different approach to approximate the regression function. They allow the appropriate form to be expressed as a probability density function that is empirically determined from the observed data using Parzen window estimation (Specht, 1991). The definition of the regression function is given by:

$$\langle t|\mathbf{x} \rangle = \int tp(t|\mathbf{x})dt = \frac{\int tp(\mathbf{x},t)dt}{p(\mathbf{x})} = \frac{\int tp(\mathbf{x},t)dt}{\int p(\mathbf{x},t)dt}. \quad (143)$$

Therefore, the joint probability function of variables \mathbf{x} and t enables to compute the conditional probability density function $p(t|\mathbf{x})$ and its expected value. The training set D composed of pairs (\mathbf{x}^n, t^n) is used to estimate $p(\mathbf{x},t)$ using the Parzen window procedure. The following approximation for the joint probability density function is considered (Specht, 1991):

$$p(\mathbf{x},t) \approx \frac{1}{(2\pi)^{(d+1)/2} \sigma_g^{(d+1)}} \frac{1}{N} \sum_{n=1}^N \exp \left[-\frac{(\mathbf{x} - \mathbf{x}^n)^T (\mathbf{x} - \mathbf{x}^n)}{2\sigma_g^2} \right] \exp \left[-\frac{(t - t^n)^2}{2\sigma_g^2} \right]. \quad (144)$$

Substituting this approximation into the expression of the desired regression function in Eq. (143) leads to the following result:

$$\langle t|\mathbf{x} \rangle \approx \frac{\sum_{n=1}^N \exp \left[-\frac{(\mathbf{x} - \mathbf{x}^n)^T (\mathbf{x} - \mathbf{x}^n)}{2\sigma_g^2} \right] \int t \exp \left[-\frac{(t - t^n)^2}{2\sigma_g^2} \right] dt}{\sum_{n=1}^N \exp \left[-\frac{(\mathbf{x} - \mathbf{x}^n)^T (\mathbf{x} - \mathbf{x}^n)}{2\sigma_g^2} \right] \int \exp \left[-\frac{(t - t^n)^2}{2\sigma_g^2} \right] dt}. \quad (145)$$

The integral in the numerator is the mean value for Gaussian variable t centred at t^n , while the integral in the denominator reduces to unity. Defining the scalar function $(\theta^n)^2$ as

$$(\theta^n)^2 = (\mathbf{x} - \mathbf{x}^n)^T (\mathbf{x} - \mathbf{x}^n) \quad (146)$$

yields the following expression:

$$\langle t | x \rangle \approx \frac{\sum_{n=1}^N t^n \exp \left[-\frac{(\theta^n)^2}{2\sigma_g^2} \right]}{\sum_{n=1}^N \exp \left[-\frac{(\theta^n)^2}{2\sigma_g^2} \right]}. \quad (147)$$

For an input pattern \mathbf{x} , the output provided by the model can be viewed as a weighted average of all of the target values contained in the training set. The weight for t^n is computed according to the Euclidean distance of the input to its corresponding pattern \mathbf{x}^n .

The variance σ_g acts as a smoothing parameter and its value must be specified by the designer. Parzen (1962) and Cacoullos (1966) showed that the estimator used for $p(\mathbf{x}, t)$ is consistent, i.e., asymptotically converges to the underlying probability density function $p(\mathbf{x}, t)$, at all points (\mathbf{x}, t) at which the density function is continuous, provided that $\sigma_g = \sigma_g(N)$ is chosen as a decreasing function of N . The variance must meet the following requirements:

$$\lim_{N \rightarrow \infty} \sigma_g(N) = 0 \quad (148)$$

and

$$\lim_{N \rightarrow \infty} N \sigma_g^d(N) = \infty. \quad (149)$$

When the parameter σ_g is large, the estimated density is forced to be smooth and in the limit becomes a multivariate Gaussian with covariance matrix $\sigma_g^2 \mathbf{I}$. In this case, the output of the model tends to be the mean value of the target samples in the training set. As σ_g goes to 0, the output assumes the value of the target t^n associated with the observation \mathbf{x}^n closest to the input \mathbf{x} . For intermediate values of σ_g , all values of t^n are taken into account, but those corresponding to points closer to \mathbf{x} are given heavier weights.

7.4. Multilayer perceptron (MLP) networks for regression

Previously, MLP networks were presented from a classification point of view. Due to their flexible architecture, MLP networks are also powerful tools for regression analysis. Indeed, both classification and regression can be seen as particular cases of function approximation. For regression problems, it is the regression function that is approximated. In a classification context, the probabilities of membership of the different classes are approximated as a function of the input variables (Bishop, 1995).

A continuous univariate target variable was used to model the AHI in the proposed regression problem. As a result, different network architecture and modelling principles characterise MLP networks for regression analysis in comparison with MLP classifiers. Initially, a linear activation function is required for the output neuron (a single output neuron is used since the target variable represents a scalar quantity). The output of this neuron represents the network output value, which is given by:

$$y = f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{N_H} \left[w_j g_t \left(\sum_{i=1}^d w_{ij} x_i + b_j \right) + b \right], \quad (150)$$

where d is the number of features in the input vector, N_H is the number of hidden units, w_j is the weight connecting hidden unit h_j with the output unit, b is the bias associated with the output unit, w_{ij} is the weight connecting the feature i of the input pattern with hidden unit h_j , b_j is the bias associated with hidden unit h_j and $g_t(\cdot)$ is the activation function for units in the hidden layer.

As indicated before, the learning process focuses on modelling the statistical properties of data in D , which requires to know the distribution $p(\mathbf{x}, t) = p(t|\mathbf{x})p(\mathbf{x})$. Specifically, the term $p(t|\mathbf{x})$ provides useful information about D , since it represents the distribution of target values given a pattern \mathbf{x} . In regression analysis, a Gaussian model is used for this distribution. Its mean value is given by the network output $y(\mathbf{x}, \mathbf{w})$, as detailed in Eq. (9), since it approximates the regression function. Therefore, the likelihood can be written as:

$$L_D = \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{n=1}^N \exp \left\{ -\frac{[y(\mathbf{x}, \mathbf{w}) - t]^2}{2\sigma^2} \right\}. \quad (151)$$

The negative logarithm transformation is used to obtain the expression for the error function:

$$E_D = -\log L_D = \frac{1}{2} \sum_{i=1}^N [y(\mathbf{x}, \mathbf{w}) - t]^2, \quad (152)$$

where terms independent of \mathbf{w} have been omitted. This yields the sum-of-squares error function. According to the maximum likelihood principle, network weights \mathbf{w} are optimised by error minimisation. As a result, the output of the network can be interpreted as an estimation of the regression function (Bishop, 1995).

Non-linear optimisation algorithms are required for network training since a non-linear dependence relates the error with network weights. These algorithms are the same used for MLP classifiers, which are based on the backpropagation procedure. In addition, regularisation techniques can be used during training to control the complexity of the network in order to increase its generalisation capability. Weight decay generalisation can be used by adding a term to the error function. Weights with a high value are penalised, reducing the effective number of adaptive parameters in the model.

7.5. Radial basis function (RBF) networks for regression

As explained in the preceding chapter, RBF networks were originally motivated from the perspective of function approximation. Similarly to MLP networks,

some modifications are required to adapt RBF networks to regression analysis in comparison with their use for classification purposes. Regression modelling by means of RBF networks requires to use a linear activation function for neurons in the output layer. A single output neuron will be used since a univariate target variable is considered. Therefore, the output of the network is given by the linear combination of the basis functions in the hidden layer:

$$y = f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{N_B} w_j \psi_j(\|\mathbf{x} - \boldsymbol{\mu}_j\|) + b, \quad (153)$$

where N_B is the number of basis functions, $\boldsymbol{\mu}_j$ is the centre of the basis function ψ_j , w_j is the weight connecting basis function ψ_j with the output neuron and b is the bias parameter for this neuron. A Gaussian kernel is commonly used for the basis function ψ_j (Bishop, 1995).

As explained for RBF classifiers, the set of basis functions should be chosen to form a representation of the probability density of the input data, which does not require to use information about the target variable (Bishop, 1995; Nabney, 2002). Therefore, network training can be performed in two different stages. In the first one, only input patterns \mathbf{x} in the training set are used to determine the parameters of the basis functions. These are centres ($\boldsymbol{\mu}_j$) and widths (σ_{ij}) of each Gaussian kernel. The basis functions are then kept fixed while weights in the output layer are found in the second phase of training. Using matrix notation, the output of the RBF network can be expressed as:

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}\boldsymbol{\Psi}(\mathbf{x}), \quad (154)$$

where vector \mathbf{w} denotes the set of weights in the output layer and vector $\boldsymbol{\Psi}^n(\mathbf{x}^n) = (1, \psi_1^n, \dots, \psi_{N_B}^n)$ is composed of the activation values of the basis functions. The first element of this vector is set to unity and represents the activation value associated to the bias term. It should be noted that a linear activation function is used for the output neuron since the target represents a continuous variable. In order to approximate the regression function, the statistical model of Eq. (9) is used for the

distribution of target values t given the input pattern \mathbf{x} . As shown for MLP networks, the maximum likelihood principle requires to optimise weights by minimising the sum-of-squares error. This optimisation problem corresponds to that described for MLR models. The error is expressed as a quadratic function of the weights and its minimum can be found from the solution of a set of linear equations. Therefore, weights are given by:

$$\mathbf{w}^T = \Psi^+ \mathbf{t}, \quad (155)$$

where matrix Ψ is composed of vectors ψ^n containing the activation value of the basis functions for pattern \mathbf{x}^n in the training set and Ψ^+ denotes the pseudoinverse matrix.

7.6. Bayesian (BY) networks for regression

BY networks were previously introduced in a classification context. In addition, these algorithms can be applied to regression analysis. As explained above, the Bayesian framework represents an alternative method for model optimisation different to the maximum likelihood principle. It suggests modelling the probability function over weight space rather than finding an optimum value for network weights (Bishop, 1995). The Bayes' theorem, as expressed in Eq. (107), is used to compute the posterior distribution of weights once training data have been observed. This function is then used to infer the distribution of output values according to Eq. (108). In regression analysis, a Gaussian function, as expressed in Eq. (9), is used to model the distribution of the noise on the target. It can be written as:

$$p(t|\mathbf{x}, \mathbf{w}) \propto \exp \left\{ -\frac{\beta}{2} [y(\mathbf{x}, \mathbf{w}) - t]^2 \right\}, \quad (156)$$

where factor β controls the shape of the distribution and is referred to as an hyperparameter and the mapping function $y(\mathbf{x}, \mathbf{w})$ represents the mean of the distribution.

Initially, a prior distribution $p(\mathbf{w})$ is defined to model the probability density function of the weights in the absence of data. The discussion about the bias-variance

trade-off indicates that smooth mappings will typically have better generalisation ability than one which is overfitted to the training data. Such mappings can be achieved by favouring small values for the weights. Therefore, the function given in Eq. (110) is assumed for this distribution. On the other hand, the likelihood of the training data is obtained from the noise model for the target variable. Provided training samples are drawn independently from this distribution, the likelihood is given by:

$$\begin{aligned} p(D|\mathbf{w}) &= \prod_{n=1}^N p(t^n | \mathbf{x}^n, \mathbf{w}) = \\ &= \frac{1}{Z_D(\beta)} \exp \left\{ -\frac{\beta}{2} \sum_{n=1}^N [y(\mathbf{x}^n, \mathbf{w}) - t^n]^2 \right\} = \frac{1}{Z_D(\beta)} \exp(-\beta E_D), \end{aligned} \quad (157)$$

where E_D is the error term and $Z_D(\beta)$ is a normalisation factor. This expression of the likelihood is different to that proposed for classification problems. The normalisation factor can be computed as:

$$Z_D(\beta) = \int \exp(-\beta E_D) dD = \left(\frac{2\pi}{\beta} \right)^{N/2}. \quad (158)$$

Finally, the posterior distribution $p(\mathbf{w}|D)$ is derived from the Bayes' theorem. The product of $p(\mathbf{w})$ and $p(D|\mathbf{w})$ is computed for this purpose. The factor $p(D)$ in the denominator is not considered since it does not depend on network weights and only acts as a normalisation term. Thus, the posterior is expressed as:

$$p(\mathbf{w}|D) = \frac{1}{Z_S} \exp(-\beta E_D - \alpha E_W) = \frac{1}{Z_S} \exp[-S(\mathbf{w})], \quad (159)$$

where Z_S ensures that the integral sums to unity and $S(\mathbf{w}) = \beta E_D(\mathbf{w}) + \alpha E_W(\mathbf{w})$. In order to use the Gaussian approximation, the term $S(\mathbf{w})$ is estimated by the Taylor expansion around its minimum value (\mathbf{w}_{MP}), retaining terms up to second order. As a result, the posterior distribution for the weight vector is Gaussian and is given by the expression in Eq. (116). It can be used to evaluate the normalisation factor (MacKay, 1992a):

$$Z_S^*(\alpha, \beta) = \exp[-S(\mathbf{w}_{MP})] (2\pi)^{W/2} |\mathbf{A}|^{-1/2}. \quad (160)$$

Suitable values must be chosen for the hyperparameters α and β . The evidence framework is applied to optimise them (MacKay, 1992b). The procedure is similar to that described for the classification approach. Parameter β is only included in the analysis when regression problems are modelled. In this context, the evidence for α and β is given by:

$$p(D|\alpha, \beta) = \int p(D|\mathbf{w}, \alpha, \beta) p(\mathbf{w}|\alpha, \beta) d\mathbf{w} = \int p(D|\mathbf{w}, \beta) p(\mathbf{w}|\alpha) d\mathbf{w}, \quad (161)$$

where the likelihood function is independent of α as well as the prior is independent of β . The exponential forms for the prior and the likelihood distributions can be used together with Eq. (159) and (160) to obtain the expression of the evidence:

$$p(D|\alpha, \beta) = \frac{1}{Z_D(\beta)} \frac{1}{Z_W(\alpha)} \int \exp[-S(\mathbf{w})] d\mathbf{w} = \frac{Z_S(\alpha, \beta)}{Z_D(\beta) Z_W(\alpha)}. \quad (162)$$

Again, it is assumed that the evidence is sharply peaked around α and β . The hyperparameters are fixed to the values for which the evidence is maximised. The natural logarithm of the evidence is given by:

$$\ln p(D|\alpha) = -\alpha E_W(\mathbf{w}_{MP}) - \beta E_D(\mathbf{w}_{MP}) - \frac{1}{2} \ln |A| + \frac{W}{2} \ln \alpha + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi). \quad (163)$$

Computing the derivative of this expression with respect to α and making it equal to zero yields the result shown in Eq. (124). The same procedure is applied to obtain an expression for the hyperparameter β . The following condition is satisfied at the maximum point (Bishop, 1995):

$$2\beta E_D(\mathbf{w}_{MP}) = N - \gamma. \quad (164)$$

For the practical implementation of Bayesian learning, hyperparameters α and β are periodically updated during training. Once \mathbf{w}_{MP} has been determined by minimising $S(\mathbf{w})$ (non-linear optimisation algorithms are required), α is updated as detailed in Eq. (126) while the new value of β is computed according to (Bishop, 1995):

$$\beta^{new} = \frac{N - \gamma}{2E_D(\mathbf{w}_{MP})}. \quad (165)$$

As described for the classification approach, the influence of each input feature on the network output can be measured by using ARD (Neal, 1996). For this purpose, a different hyperparameter α is defined for controlling the weights associated to a given input feature.

Once the Gaussian approximation to the posterior $p(\mathbf{w}|D)$ was computed, Eq. (108) and (156) are applied to obtain the following result for the distribution of network outputs:

$$p(t|\mathbf{x}, D) \propto \int \exp\left\{-\frac{\beta}{2}[t - y(\mathbf{x}, \mathbf{w})]^2\right\} \exp\left(-\frac{1}{2}\Delta\mathbf{w}^T \mathbf{A} \Delta\mathbf{w}\right) d\mathbf{w}, \quad (166)$$

where constant factors have been dropped. It is assumed that the width of the posterior distribution is sufficiently narrow that the network function $y(\mathbf{x}, \mathbf{w})$ may be approximated by its linear expansion around \mathbf{w}_{MP} :

$$y(\mathbf{x}, \mathbf{w}) = y(\mathbf{x}, \mathbf{w}_{MP}) + \mathbf{g}^T \Delta\mathbf{w}. \quad (167)$$

The variable \mathbf{g} is the gradient of $y(\mathbf{x}, \mathbf{w})$ evaluated at \mathbf{w}_{MP} :

$$\mathbf{g} \equiv \nabla y \Big|_{\mathbf{w}=\mathbf{w}_{MP}}. \quad (168)$$

Substituting Eq. (167) into Eq. (166) and computing the integral leads to a Gaussian distribution of the outputs (Bishop, 1995):

$$p(t|\mathbf{x}, D) = \frac{1}{(2\pi\sigma_t^2)^{1/2}} \exp\left[-\frac{(t - y_{MP})^2}{2\sigma_t^2}\right]. \quad (169)$$

The variance of the distribution is given by:

$$\sigma_t^2 = \frac{1}{\beta} + \mathbf{g}^T \mathbf{A}^{-1} \mathbf{g}. \quad (170)$$

This variance can be viewed as an error bar on the mean value of the distribution, which is represented by the network output y_{MP} . As can be observed, two different terms contribute to the error in the prediction. The first one arises from the intrinsic noise on the target data and is represented by β . The second term is due to the width of the posterior distribution of the weights and is reflected by the Hessian matrix \mathbf{A} (Bishop, 1995).

7.7. Multivariate adaptive regression splines (MARS)

MARS is a non-linear method for regression analysis. It was developed as a generalisation of recursive partitioning regression, which takes the form (Friedman, 1991):

$$\text{if } \mathbf{x} \in R_m, \text{ then } y = g_m(\mathbf{x}, w_m), \quad (171)$$

where R_m ($m = 1, \dots, M$) are disjoint subregions representing the input feature space. The functions $g_m(\cdot)$ are generally taken to be of quite simple parametric form. The most common is a constant function:

$$g_m(\mathbf{x}, w_m) = w_m. \quad (172)$$

Recursive partitioning uses training data to simultaneously estimate a good set of subregions and the parameters (w_m) associated with the separate functions in each subregion. The partitioning is accomplished through the recursive splitting of previous subregions. The starting region is the entire feature space. At each stage of the process, a feature x_v is selected for the next partition of the region R_m as indicated in the following expression (Friedman, 1991):

$$\text{if } \mathbf{x} \in R_m, \text{ then } \begin{cases} \mathbf{x} \in R_l, x_v \leq u \\ \mathbf{x} \in R_r, x_v > u \end{cases}, \quad (173)$$

where v is used to label one of the input features and u is a value on that feature. The split is jointly optimised over $1 \leq v \leq d$ and $-\infty \leq u \leq \infty$ by minimising an error function. For instance, the sum-of-squares error can be used for this purpose. In order to minimise the effect of the curse of dimensionality, the subregions are then recombined in a reverse manner until an optimum set is reached. An additional term that penalises an increasing number of regions is used together with the error measure.

However, recursive partitioning suffers from severe restrictions that limit its effectiveness. The first fundamental drawback is the lack of continuity of the solution. It severely limits the accuracy of the approximation, especially when the true underlying function is continuous. Imposing continuity only on the function, i.e. derivatives of low order are not required to be continuous, may significantly improve approximation accuracy. Second, recursive partitioning often results in a poor predictive ability for certain types of simple functions. For instance, linear functions with more than a few non-zero coefficients and additive functions involving several variables. The difficulties of the method arise when the dominant interactions involve a small fraction of the total number of variables (Friedman, 1991; Crino and Brown 2007).

MARS was proposed as a generalisation of recursive partitioning to overcome these problems. The starting point is to write the expressions in Eq. (171) and (172) according to the form of an expansion in a set of basis functions:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{m=1}^M w_m B_m(\mathbf{x}), \quad (174)$$

where $\mathbf{w} = (w_1, \dots, w_m)$ is a set of adaptive weights. The basis functions $B_m(\cdot)$ take non-zero values for points $\mathbf{x} \in R_m$. Functions of this type are expressed as the product of several step functions, resulting in the following expression:

$$B_m(\mathbf{x}) = \prod_{k=1}^{K_m} H[s_k(x_v - u_k)], \quad (175)$$

where H denotes the step function, K_m is the number of step functions (number of splits) that compose the basis function B_m , s_k takes the values +1 or -1, x_v is one of the input

features and u_k is a value on this variable representing the location associated to the step function. At each iteration of the forward phase of the algorithm, a basis function B_m is removed and is replaced by two new basis functions. These are the product of B_m with the step function defined for the input variable x_v , centred at u' and its reflected form. In the backward phase, those basis functions that no longer contribute to the accuracy of the fit are removed. The strategy is to overfit the data with an excessively large model. Then, the backward procedure is used to reduce it to its proper size (Friedman, 1991).

To overcome the lack of continuity, the step function is replaced by a continuous function of the same argument everywhere it appears. The two-sided truncated power basis functions for representing q th order splines are used for this purpose (Friedman, 1991). They are given by the following expression:

$$b_q(x-u) = [\pm(x-u)]_+^q, \quad (176)$$

where u indicates the knot location, q is the order of the spline and the subscript (+) indicates the positive part of the argument. For $q > 0$, the spline approximation is continuous and has $q - 1$ continuous derivatives. Using these splines, the basis functions take the form:

$$B_m(\mathbf{x}) = \prod_{k=1}^{K_m} [s_k(x_v - u_k)]_+^q. \quad (177)$$

The resulting basis functions will not reduce to a set of tensor product spline basis functions, as observed for step functions, since the algorithm could produce multiple splitting on the same variable. Therefore, the obtained basis function could present dependencies of higher order than q on individual variables. To avoid this, the algorithm is enforced to produce functions composed of the product of univariate splines of different variables. The aim is to preserve the properties of splines for function approximation in the recursive partitioning approach (de Boor, 1978).

The second limitation of recursive partitioning to be addressed is its inability to provide accurate approximations to certain types of simple functions. These are functions that either have no strong interaction effects or interactions involve at

most a few of the predictor variables. The problem of recursive partitioning is that basis functions tend to involve more than a few variables, resulting in higher order interactions. Each iteration of the algorithm removes a basis function of lower interaction order and replaces it by two functions, each with interaction order one level higher. The proposed modification involves not removing the parent basis function, resulting in a double number of basis functions obtained from each iteration. Since no restrictions are placed on the choice of a parent term, the modified procedure is able to produce models involving either high or low order interactions or both (Friedman, 1991).

The MARS algorithm is obtained from the recursive partitioning approach by applying the modifications described before. The forward phase of the algorithm produces M_{max} tensor product spline basis functions. This basis set is then subjected to a backwards deletion strategy to obtain a final set of basis functions. The output of the algorithm is computed according to expression in Eq. (174).

Selection of basis functions in both phases is carried out in order to minimise an error measure. The sum-of-squares error is used for this purpose. In addition, a penalty term to account for the increased variance associated with higher model complexity is included in the error function, which takes the form (Friedman and Silverman, 1989):

$$E = \frac{1}{N} \sum_{n=1}^N \frac{[y^n(\mathbf{x}^n, \mathbf{w}) - t^n]^2}{\left[1 - \frac{C(M)}{N}\right]^2}, \quad (178)$$

where $C(M)$ reflects the number of parameters that are fit to the data. If the value of the basis function parameters such as the number of factors K_m , variables x_v and knot locations u were determined independently of the target values, the complexity cost function would be given by:

$$C(M) = \text{trace} \left[\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \right] + 1 \quad (179)$$

where \mathbf{B} is the $(M \times N)$ data matrix composed of elements $B_{ij} = B_i(\mathbf{x}_j)$. The value of $C(M)$ represents the number of linearly independent basis functions, i.e., the number of parameters w_m being fit. In order to reflect the remaining parameters that are fit to the data, an increased cost complexity function is used. Such function is expressed as:

$$\tilde{C}(M) = C(M) + pM \quad (180)$$

where p represents a cost for each basis function optimisation and is a smoothing parameter of the procedure. Larger values of p lead to fewer knots being placed and thereby smoother function estimates. As suggested by Friedman (1991), the best value for p is in the range $2 \leq p \leq 4$.

Another choice for the MARS model is related to the value of q , the power of the univariate spline functions. The continuity properties of the approximation are governed by this parameter. Imposing continuity of the first derivative could result in additional accuracy if the function presents marked variations. Thus, the value of q would be set to 2. However, the points with a higher contribution to the error are those located near the boundaries of the domain. Fitting high degree polynomials (associated with high degree regression splines) in these regions leads to high variance of the function estimate there. One approach is to modify the spline basis in order to present a linear behaviour near the ends of the data interval (Friedman, 1991). The expression is given by:

$$b(x, s = +1, u_-, u, u_+) = \begin{cases} 0 & x \leq u_- \\ p_+ (x - u_-)^2 + r_+ (x - u_-)^3 & u_- < x < u_+ \\ x - u & x \geq u_+ \end{cases}, \quad (181)$$

$$b(x, s = -1, u_-, u, u_+) = \begin{cases} -(x - u) & x \leq u_- \\ p_- (x - u_+)^2 + r_- (x - u_+)^3 & u_- < x < u_+ \\ 0 & x \geq u_+ \end{cases}$$

with $u_- < u < u_+$. Setting

$$\begin{aligned}
p_+ &= (2u_+ + u_- - 3u) / (u_+ - u_-)^2 \\
r_+ &= (2u - u_+ - u_-) / (u_+ - u_-)^3 \\
p_- &= (3u - 2u_- - u_+) / (u_- - u_+)^2 \\
r_- &= (u_- + u_+ - 2u) / (u_- - u_+)^3
\end{aligned} \tag{182}$$

causes $b(x, s, u_-, u, u_+)$ to be continuous and have continuous first derivatives. Each factor in every basis function in the approximation produced by the original algorithm, i.e., using $q = 1$, is replaced by a corresponding cubic factor as described. The central knot u is placed at the same location. Side knots u_- and u_+ are located in order to reduce the number of second derivative discontinuities (Friedman, 1991).

Finally, it is interesting for the designer to control the order of the interactions between variables in the basis functions, represented by the parameter mi . The value $mi = 1$ represents a pure additive model. Its efficiency can be compared with that of other models with higher values of mi . In addition, an appropriate value of M_{max} , the maximum number of basis functions in the model, must be chosen. A high value of this parameter is required to approximate complex functions. The value of M_{max} should be considerably larger than the optimum number of basis functions. Friedman (1991) found that setting M_{max} to approximately twice the number of basis functions selected for the final model is sufficient.

7.8. Support vector machines (SVM) for regression

Initially, SVM were developed to perform classification tasks. Additionally, SVM algorithms based on the principles of statistical learning theory have been adapted to regression analysis (Vapnik, 1999). As described for the classification approach, SVM regression algorithms are trained according to the principle of structural risk minimisation. High generalisation ability is obtained by achieving high performance on training data (i.e. reducing empirical risk) while avoiding overfitting.

Given a training set D composed of N input-output pairs (\mathbf{x}^n, t^n) , with $\mathbf{x} \in \mathfrak{R}^d$ and $t \in \mathfrak{R}$, the goal in SVM regression is to find a function $y(\mathbf{x}, \mathbf{w})$ that has at most ε deviation from the samples in D . Errors are not considered if they are less than ε but a

deviation greater than this will not be accepted. The problem would consist in finding the following function:

$$y = f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{x} + w_0. \quad (183)$$

A smooth mapping is desired in order to reduce overfitting effects. For the function detailed above, this is achieved by defining a small \mathbf{w} . Therefore, the optimisation problem would focus on minimising (Smola and Schölkopf 2004)

$$E_w = \frac{1}{2} \|\mathbf{w}\|^2 \quad (184)$$

subject to the constraints:

$$\begin{aligned} t^n - \mathbf{w}^T \mathbf{x}^n - w_0 &\leq \varepsilon \\ \mathbf{w}^T \mathbf{x}^n + w_0 - t^n &\leq \varepsilon \end{aligned} \quad (185)$$

It has been assumed that the function $y(\mathbf{x}, \mathbf{w})$ that approximates all pairs (\mathbf{x}^n, t^n) with ε precision actually exists. However, this sometimes may not be the case and possible errors must be allowed. Equally to SVM classifiers, non-negative slack variables ξ^n and ξ_*^n are introduced to account for deviations from the target value higher than ε . These variables define a new type of loss function called ε -insensitive loss, which is adopted for SVM regression (Smola and Schölkopf 2004). It takes into account the accuracy in the approximation required for the SVM algorithm and is given by the following expression:

$$\xi_{(*)}^n = L_\varepsilon(t^n, y(\mathbf{x}^n, \mathbf{w})) = \begin{cases} 0 & \text{if } |t^n - y(\mathbf{x}^n, \mathbf{w})| \leq \varepsilon \\ |t^n - y(\mathbf{x}^n, \mathbf{w})| & \text{otherwise} \end{cases}. \quad (186)$$

Therefore, the error on the training set is evaluated as:

$$E_D = \sum_{n=1}^N L_\varepsilon(t^n, y(\mathbf{x}^n, \mathbf{w})) = \sum_{n=1}^N (\xi^n + \xi_*^n). \quad (187)$$

The optimisation problem is then defined as the minimisation of the functional

$$E = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi^n + \xi_*^n) \quad (188)$$

subject to the constraints:

$$\begin{aligned} t^n - y(\mathbf{x}^n, \mathbf{w}) - w_0 &\leq \varepsilon + \xi_*^n \\ y(\mathbf{x}^n, \mathbf{w}) + w_0 - t^n &\leq \varepsilon + \xi^n, \\ \xi^n, \xi_*^n &\geq 0 \end{aligned} \quad (189)$$

where C is a positive regularisation constant. The functional E implements the structural risk minimisation principle. The parameter C determines the trade-off between the smoothness of the approximation and the amount up to which deviations larger than ε are tolerated.

Similarly to classification problems, the optimisation is not directly carried out from in the space defined by the original features. Instead, SVM regressors map input data (\mathbf{x}) onto a higher dimensional space (\mathfrak{Z}) of dimension $d_s > d$ using a non-linear transformation $\mathbf{z} = \boldsymbol{\varphi}(\mathbf{x}) / \mathbf{z} \in \mathfrak{R}^{d_s}, d_s > d$. The representation of the data in the transformed space facilitates regression analysis. The optimisation is then performed from the set of transformed patterns obtained from samples in the training set. A linear model in the transformed space is obtained. It is expressed as:

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{z} + w_0. \quad (190)$$

Ideally, solving this optimisation problem requires to find the values of \mathbf{w} and w_0 once the non-linear transformation has been specified. The key idea is to construct a Lagrange function from the objective function E , leading to a dual optimisation problem. It requires to maximise the expression (Smola and Schölkopf 2004)

$$\begin{aligned}
Q(\eta^n, \eta_*^n) = & -\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N (\eta^n - \eta_*^n)(\eta^m - \eta_*^m) K(\mathbf{x}^n, \mathbf{x}^m) - \\
& -\varepsilon \sum_{n=1}^N (\eta^n + \eta_*^n) + \sum_{n=1}^N t^n (\eta^n - \eta_*^n)
\end{aligned} \tag{191}$$

subject to the constraints:

$$\sum_{n=1}^N (\eta^n - \eta_*^n) = 0 \text{ and } 0 \leq \eta^n, \eta_*^n \leq C. \tag{192}$$

As in the classification case, the term $K(\cdot, \cdot)$ represents the inner-product kernel function in the transformed space and is given by expression in Eq. (135). The set of adaptable weights \mathbf{w} is obtained as a linear combination of the transformed training patterns:

$$\mathbf{w} = \sum_{n=1}^N (\eta^n - \eta_*^n) \boldsymbol{\varphi}(\mathbf{x}^n). \tag{193}$$

Those training patterns for which either η^n or η_*^n is distinct to zero are identified as support vectors. Only one of both parameters can take a non-zero value. Those patterns for which $\eta_{(*)}^n \neq 0$ are involved in the computation of the output and are referred to as support vectors. Again, two different types of support vector can be identified. Those training patterns with corresponding $\eta_{(*)}^n = C$ lie outside the ε -insensitive tube and are associated to $\xi_{(*)}^n > 0$. A different type of support vectors are those with $0 < \eta_{(*)}^n < C$, which are associated to $\xi_{(*)}^n = 0$, i.e., they lie on the margin of the tube (Vapnik and Chapelle 2000).

The value of w_0 can be computed by exploiting the Kuhn-Tucker conditions (Haykin, 1999). At any point with $0 < \eta_{(*)}^n < C$ and $\xi_{(*)}^n = 0$ the following equality holds:

$$|t^n - \mathbf{w}^T \mathbf{z}^n - w_0| - \varepsilon = 0. \tag{194}$$

All support vectors satisfying this equation can be evaluated and the average over the different results for w_0 can be used for the final SVM algorithm. Its output is finally obtained as:

$$y = \sum_{n \in S} (\eta^n - \eta_*^n) K(\mathbf{x}^n, \mathbf{x}) + w_0, \quad (195)$$

where S denotes the subset of indices corresponding to the support vectors (Smola and Schölkopf 2004). The complexity of a function approximated by SVM is independent of the dimensionality of the input space and depends only on the number of support vectors.

As can be observed, it is not required to know the non-linear transformation applied on data in the original input space and only the inner product kernel function is needed to compute the output of the algorithm. This function must satisfy Mercer's theorem. Several forms can be considered for that function. The Gaussian kernel given in Eq. (139) is one of the most common forms and will be considered for SVM regression algorithms implemented in this study.

8. Results

- 8.1. Introduction
- 8.2. Performance assessment
 - 8.2.1. Performance measures for classification
 - 8.2.2. Performance measures for regression
- 8.3. Model selection and performance estimation
- 8.4. Feature extraction
- 8.5. Feature normalisation and dimensionality reduction
- 8.6. Classification results
 - 8.6.1. Design and training of classification algorithms
 - 8.6.1.1. *Design of KNN classifiers*
 - 8.6.1.2. *Design of FLD classifiers*
 - 8.6.1.3. *Design of LR classifiers*
 - 8.6.1.4. *Design of PNN classifiers*
 - 8.6.1.5. *Design of MLP classifiers*
 - 8.6.1.6. *Design of RBF classifiers*
 - 8.6.1.7. *Design of BY network classifiers*
 - 8.6.1.8. *Design of SVM classifiers*
 - 8.6.2. Testing classification algorithms
- 8.7. Regression results
 - 8.7.1. Design and training of regression algorithms
 - 8.7.1.1. *Design of MLR algorithms*
 - 8.7.1.2. *Design of GRNN algorithms*
 - 8.7.1.3. *Design of MLP algorithms for regression*
 - 8.7.1.4. *Design of RBF algorithms for regression*
 - 8.7.1.5. *Design of BY network algorithms for regression*
 - 8.7.1.6. *Design of MARS algorithms*
 - 8.7.1.7. *Design of SVM algorithms for regression*
 - 8.7.2. Testing regression algorithms
- 8.8. Conventional analysis of oximetry signals

8.1. Introduction

This chapter shows the results achieved in the Thesis for the proposed classification and regression problems. Initially, the measures used to evaluate classification and regression performance are presented. Subsequently, a quantitative description of the features extracted from SaO₂ signals is provided. Their individual capability to discriminate between SAHS-negative and SAHS-positive subjects as well as to predict the AHI value is evaluated. Afterwards, the transformation used for dimensionality reduction is analysed. In the following points, a detailed description of the design of classification and regression methods is provided. Usually, model design involves adjusting one or more smoothing parameters that determine model complexity and generalisation ability. The algorithms with the selected design were subsequently evaluated on the test set to measure their ability in classification and regression tasks. Finally, the individual utility of conventional oximetry indices for solving both problems was also assessed.

8.2. Performance assessment

8.2.1. Performance measures for classification

Accuracy (*Acc*) is the critical performance measure for a classifier (Jain, 2000). It was used to evaluate classification algorithms implemented in the Thesis. *Acc* is defined as the probability of correct response and its estimation is computed as the percentage of subjects correctly classified:

$$Acc = \frac{Tp + Tn}{Tp + Fn + Tn + Fp}, \quad (196)$$

where *Tp* (true positives) is the number of SAHS-positive subjects correctly classified, *Tn* (true negatives) is the number of SAHS-negative subjects correctly classified, *Fn* (false negatives) is the number of misclassified SAHS-positive subjects and *Fp* (false positives) is the number of misclassified SAHS-negative subjects. In addition, *Acc* can be expressed in terms of sensitivity (*Se*) and specificity (*Sp*). They indicate the number of SAHS-positive and SAHS-negative subjects correctly classified, respectively. These statistics can be respectively computed as:

$$Se = \frac{Tp}{Tp + Fn} \quad (197)$$

and

$$Sp = \frac{Tn}{Tn + Fp}. \quad (198)$$

Confusion matrices can be used to reflect the quantity of signals used in the study as well as the number of correct and incorrect decisions. A confusion matrix is a form of contingency table showing the difference between the true and predicted classes for a set of labelled examples, as shown in Table 2 (Bradley, 1997). Acc , Se and Sp can be directly obtained from the information summarised in this matrix.

Additionally, other performance measures can be considered for classification problems. Receiver operating characteristic (ROC) analysis has been proposed as a robust method to evaluate classification performance (Bradley, 1997). Indeed, its use is especially common in medical applications (Zweig and Campbell 1993; Schwaighofer *et al.* 2003). Performance measures such as Acc , Se and Sp are obtained by comparing the output of the classifier with a given threshold. It corresponds with a surface in the input space that defines the decision boundary. The ROC analysis suppresses the requirement for a threshold by appraising the performance of a classifier over its whole range of possible values (Zweig and Campbell 1993). A plot of sensitivity versus $1 - \text{specificity}$ is made over the evaluated thresholds to obtain the ROC curve. Different theoretically based measures have been proposed to reduce an entire ROC curve to a single quantitative index. The most popular of them is the estimation of the area under the ROC curve (AUC). This index varies from 0.5 (no apparent accuracy) to 1.0 (perfect accuracy) as the ROC curve moves towards the left

True Class	Predicted Class	
	-	+
-	Tn	Fp
+	Fn	Tp

Table 2. Structure of a confusion matrix for a two-class classification problem.

and top boundaries of the graph. *AUC* represents the probability of correct classification for a randomly chosen pairs of samples and it has been suggested as an appropriate performance measure for machine learning algorithms (Hanley and McNeil 1982).

A different measure for evaluating classification performance is the Cohen's Kappa index (κ) (Cohen, 1960). It is a measure of observer agreement for categorical data. This index can be used to quantify the reliability or consistency of a classifier by evaluating the agreement with its reference technique. In the SAHS diagnosis problem, diagnostics provided by PSG are used as reference. Index κ has been long studied in medicine (Di Eugenio and Glass 2004). It is computed as:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}, \quad (199)$$

where $P(A)$ is the observed agreement among the observers and $P(E)$ is the expected agreement, i.e. $P(E)$ represents the probability that the observers agree by chance. The values of κ are constrained to the interval between -1 and 1 : $\kappa = 1$ means perfect agreement, $\kappa = 0$ means that agreement is equal to chance and $\kappa = -1$ means absolute disagreement.

8.2.2. Performance measures for regression

Regression methods were evaluated using the intraclass correlation coefficient (*ICC*). *ICC* is a reliability index used to evaluate the agreement between quantitative scores from different observers (Fisher, 1925). Historically, the Pearson correlation coefficient has been used as an index to measure the association for continuous outcomes. However, this correlation coefficient only measures the linear relationship between two quantitative variables and not the agreement (Chen and Barnhart 2008). *ICC* takes systematic error into account in assessing agreement.

For a group of measurements, the total variance in the data (σ_T^2) can be thought of as being due to true score variance (σ_{tr}^2) and error variance (σ_e^2). A

reliability coefficient can be obtained as the ratio between the true variance and the total variance (Weir, 2005):

$$\text{Reliability} = \frac{\sigma_{tr}^2}{\sigma_{tr}^2 + \sigma_e^2}. \quad (200)$$

An index of the true variance in the data is obtained by measuring between-subjects variability, i.e., the variance due to how subjects differ from each other. In this context, regarding the expression in Eq. (200), reliability is formally defined as:

$$\text{Reliability} = \frac{\text{between subjects variability}}{\text{between subjects variability} + \text{error}}. \quad (201)$$

There exist several versions of *ICC* to quantify reliability as expressed in the previous equation. The necessary variance estimations are derived from analysis of variance (ANOVA). Specifically, the various *ICC* can be calculated from mean square values derived from a within-subjects, single factor ANOVA (Weir, 2005).

Shrout and Fleiss (1979) presented three general models to compute *ICC*. Each of them can be calculated in two different ways, resulting in a total of six different expressions for *ICC*. If the scores in the analysis correspond to single scores for each subject from each observer, the *ICC* is given a second designation of 1. If the scores in the analysis represent the average of k scores for each subject, then the *ICC* is given a second designation of k . Using this nomenclature, a designation of 2,1 indicates an *ICC* calculated using model 2 with single scores. The use of one of the three different models depends on the problem under analysis and must be determined by the user. For model 1, each subject is assumed to be assessed by a different set of observers than other subjects, and these observers are assumed to be randomly sampled from the population of possible observers, i.e. observers represent a random effect. It is called a one-way model since the variability between subjects is the only source of variance. Model 2 assumes each subject was assessed by the same group of observers. In addition, these observers were randomly sampled from the population of possible observers (random effects). Model 3 assumes each subject was assessed by the same group of observers, but these particular observers are the only ones of interest (fixed

effects). In this case, the analysis attempts to determine the reliability of the observers used by that particular study and observers are considered a fixed effect (Weir, 2005). Models 2 and 3 are called two-way models since variance is due to variability between subjects as well as between observers.

In order to select the appropriate estimation of *ICC*, some issues must be considered. The first decision must determine if a one-way or a two-way model should be applied. In a one-way model, the effect of observers is not crossed with subjects, meaning that it allows for situations where all observers do not score all subjects. In contrast, in two-way models it is assumed that a fixed number of observers provides scores for all the subjects. On the other hand, fixed effects should be considered if there is no attempt at generalisation of the reliability analysis. However, for two-way models, the choice does not only involve the decision regarding fixed or random effects. Models 2 and 3 also differ in their treatment of systematic error. Model 3 only considers random error while model 2 considers both random and systematic error (Weir, 2005).

For the SAHS diagnosis problem, two different scores of the AHI value are associated to each subject. One of the scores is provided by the regression algorithm under analysis while the other is the reference value obtained through PSG. *ICC* was used to evaluate the reliability of the algorithm regarding the gold-standard method. Thus, there are two observers and each of them provided a single score for all the subjects in the study. In addition, it would be desirable to take into account the systematic error. The goal is to detect situations where the regression algorithm tends to provide overestimated or underestimated values of the AHI. Therefore, the *ICC*(2,1) model described by Shrout and Fleiss (1979) was used. It is given by the following expression:

$$ICC(2,1) = \frac{MS_S - MS_E}{MS_S + (n_o - 1)MS_E + \frac{n_o(MS_T - MS_E)}{n_s}}, \quad (202)$$

where n_o is the number of observers, n_s is the number of subjects, MS_S is the subjects mean square, MS_E is the error mean square and MS_T is the observers mean square.

The *ICC* can be interpreted as the proportion of relevant variance that is associated with differences among measured subjects. Its value ranges from -1 to 1 . A negative *ICC* indicates that more differences are observed within (error in the approximation) than between subjects. *ICC* values close to one reflect good reliability of the method under study. However, it is worth noting that its value depends on the variability between subjects. Low levels of between-subjects variability will reduce the *ICC* even if the differences between subjects' scores across observers are small (Weir, 2005).

The standard error of the approximation measured as the root mean square error (E_{RMS}) is another performance index to evaluate the accuracy of a regression algorithm. It is computed as:

$$E_{RMS} = \sqrt{\frac{\sum_{n=1}^N (y^n - t^n)^2}{N}}. \quad (203)$$

However, it does not provide information about the size of the error in comparison with the target AHI value. In order to quantify the relevance of the error, the relative error (E_{rel}) per subject was used. It evaluates the deviation between the estimation and the true value of AHI and is compared with the latter:

$$E_{rel} = \frac{1}{N} \sum_{n=1}^N \frac{|y^n - t^n|}{t^n}. \quad (204)$$

8.3. Model selection and performance estimation

Most of the evaluated pattern recognition techniques require the user to specify some design or smoothing parameters. They determine the complexity of the mapping function $y(\mathbf{x}, \mathbf{w})$ and, thus, have strong effects on the generalisation ability of the algorithm. The goal is to find models that give good predictions for new data. The bias-variance trade-off showed that there is an optimum degree of complexity for a given training set. Models with too little flexibility as well as those with a high degree of complexity will lead to poor generalisation ability due to underfitting and overfitting, respectively (Bishop, 1995). Thus, a model selection process must be carried out to

properly design pattern recognition algorithms, i.e. to find the optimum balance between bias and variance.

Model selection requires to compare the performance of several algorithms with different configurations (i.e. with different values of the smoothing parameters). The hold-out method can be used for this purpose (Jain, 2000). It consists in using independent sets of data for training and validation. Results achieved on the validation set determine the optimum algorithm for the problem under study. However, this procedure could lead to some overfitting on the validation set and the selected algorithm should be evaluated on an independent test set composed of previously unseen data (Bishop, 1995).

In the hold-out method the initial group of signals is divided into three different sets. However, it is not a feasible procedure for the proposed SAHS diagnosis problem since a limited quantity of data is available. As previously indicated, a training set with 96 SaO₂ signals and a test set with 144 SaO₂ signals were used to carry out this study. In order to get an unbiased estimation of the results, the training set was used for both training and model selection. Subsequently, the selected algorithm was evaluated on the test set to estimate its performance in the proposed problem. Model selection was carried out by using leave-one-out cross-validation on training data (Jain, 2000). For a given configuration, a model is implemented by using $N - 1$ training samples and evaluated on the one remaining sample. This is repeated N times with every training set of size $N - 1$. As a result, a prediction is obtained for each of the signals. The predicted values can be used to estimate the performance of the configuration defined for the N models implemented during the process. The optimum configuration is selected by comparing the performance of different model configurations through the leave-one-out procedure. Once model selection is completed, a new model with the selected configuration is trained by using the whole training set. Subsequently, it is evaluated on the test set to estimate its performance.

8.4. Feature extraction

Oximetry signals in the database were preprocessed to remove artefacts that could falsely influence the value of the extracted features. The obtained signals were

analysed in time and frequency domains to perform feature extraction. A total of 14 features (7 time-domain features and 7 frequency-domain features) were computed for classification and regression tasks.

Two properties of oximetry data must be considered for time-domain analysis. First, they are extremely long signals corresponding to several hours of sleep monitoring, which leads to thousands of samples per signal. Second, they present a non-stationary behaviour with marked fluctuations during apnoea episodes and a stable profile in phases characterised by normal breathing. Thus, the following procedure was applied to estimate the value of each time-domain feature from a SaO₂ signal (Marcos *et al.* 2010):

1. Signal was divided into epochs of 512 samples. Samples at the end of the recording not included in the last epoch were removed.
2. Each feature was separately computed on each of the epochs.
3. The definitive estimation for a given feature was computed as the average of the values from all the epochs.

Conventional statistics and non-linear methods were proposed to analyse SaO₂ signals in the time domain. The distribution of SaO₂ values tends to have a different behaviour for SAHS-negative and SAHS-positive groups. The first four statistical moments (μ , σ , γ and δ) were used to characterise the random variable representing the SaO₂ amplitude. For each signal, the relative frequency of each SaO₂ value was used to approximate the probability density function of this variable. To illustrate its behaviour, Figure 14 depicts average densities for both groups computed from SaO₂ signals in the training set. As can be observed, a narrow and peaked distribution centred approximately at SaO₂ = 95% was obtained for control subjects. In contrast, the distribution for subjects affected by SAHS reflects a larger spread, assigning more probability to lower SaO₂ values. In this case, probability is approximately concentrated around SaO₂ = 94%.

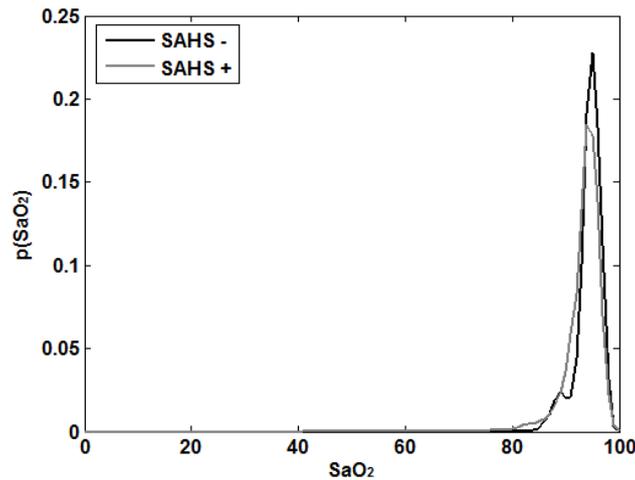


Figure 14. Average probability density function of the variable representing the SaO_2 amplitude computed from SAHS-positive and SAHS-negative subjects in the training set.

On the other hand, non-linear methods (*ApEn*, *CTM* and *LZC*) were used to assess irregularity, variability and complexity of SaO_2 signals. These properties are reflected in the evolution of recordings along time and are related to the repetition of apnoea events. Each of these methods requires the user to specify the value of one or more design parameters. Previous studies analysed the influence of these parameters by evaluating a wide range of values for them (Álvarez *et al.* 2006; Álvarez *et al.* 2007; Hornero *et al.* 2007). The optimum configurations determined in these preceding studies were used to compute *ApEn*, *CTM* and *LZC*. In the case of *ApEn*, the sequence length was set to $m = 1$ whereas the optimum width of the tolerance window r was fixed at 0.25 times the standard deviation of the samples in each signal epoch (Hornero *et al.* 2007). To compute *CTM*, a radius $\rho_{CTM} = 1$ was selected as optimum (Álvarez *et al.* 2007). Finally, *LZC* was computed by converting SaO_2 signals into 0–1 sequences. Each SaO_2 value was compared with the median of the samples in the corresponding epoch to transform the data (Álvarez *et al.* 2006).

To perform frequency-domain analysis, the non-parametric Welch's method (Welch, 1967) was used to estimate the PSD of SaO_2 signals. A Hanning window with a length of 512 samples and an overlapping of 50% were used. For each signal segment, the Fast Fourier Transform was computed for a length of 1024 samples. The estimated PSD was calculated as the average of the PSD obtained for all the segments. Figure 15 depicts the average PSD for the SAHS-positive and the SAHS-negative group from

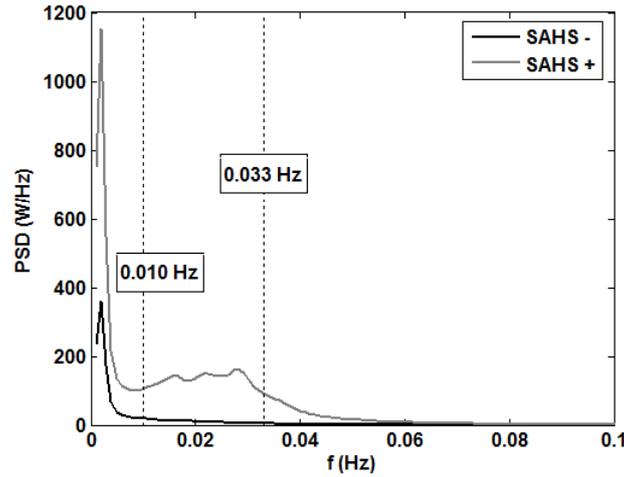


Figure 15. Average PSD function computed from SAHS-positive and SAHS-negative subjects in the training set.

signals in the training set. It reflects that signal power is concentrated in lower frequencies. However, signals from subjects affected by SAHS tend to present more power in the band between 0.010 and 0.033 Hz than signals from control subjects (Zamarrón *et al.* 2003). It indicates that these frequency components have a relevant influence in signal dynamics. In order to characterise the behaviour of SaO₂ signals in the frequency domain, the random variable representing the frequency component was analysed using conventional statistics. The first four statistical moments (μ_f , σ_f , γ_f and δ_f) were computed for this variable. The normalised PSD was used as its probability density function. In addition, three spectral properties of oximetry recordings were evaluated. These are the power of the signal (S_T), computed as the area under the PSD function, the power in the band between 0.010 and 0.033 Hz (S_B), which corresponds to the area under the PSD in that region, and the peak amplitude of the PSD in this range of frequencies (PA).

The proposed time-domain and frequency-domain features were computed for the database of SaO₂ signals. Their mean value from subjects in the training set is indicated in Table 3. As expected, signals from SAHS-negative subjects were associated with higher values of μ_t , γ_t , δ_t , CTM , γ_f and δ_f . The opposite result was observed for σ_t , LZC , μ_f , σ_f , S_T , S_B and PA , which tended to be higher in SaO₂ recordings corresponding to subjects affected by SAHS. A particular result was obtained for $ApEn$. The initial hypothesis explained that more irregularity is expected for recordings from SAHS-

TRAINING SET			
	All	SAHS-positive	SAHS-negative
μ_t	93.7378 ± 2.3079	93.4099 ± 2.3346	94.3936 ± 2.1401
σ_t	3.7093 ± 10.7092	5.2464 ± 12.8711	0.6350 ± 0.3884
γ_t	-0.0363 ± 0.4149	-0.1176 ± 0.3167	0.1263 ± 0.5317
δ_t	4.0549 ± 0.9569	3.7544 ± 0.7212	4.6560 ± 1.0906
$ApEn$	0.4491 ± 0.0678	0.4497 ± 0.0678	0.4480 ± 0.0689
CTM	0.9660 ± 0.0802	0.9503 ± 0.0946	0.9974 ± 0.0026
LZC	0.2522 ± 0.0338	0.2596 ± 0.0353	0.2373 ± 0.0252
μ_f	0.0120 ± 0.0050	0.0137 ± 0.0049	0.0084 ± 0.0029
σ_f	0.0004 ± 0.0002	0.0004 ± 0.0001	0.0003 ± 0.0002
γ_f	8.9779 ± 2.0335	8.3282 ± 1.4624	10.2773 ± 2.3899
δ_f	172.5797 ± 77.8868	152.1343 ± 55.1621	213.4704 ± 99.0837
S_T	2.4987 ± 6.6300	3.4651 ± 7.9611	0.5658 ± 0.3456
S_B	0.9409 ± 3.6388	1.3734 ± 4.4040	0.0758 ± 0.0436
PA	163.6669 ± 568.7297	239.5683 ± 710.6847	11.8642 ± 7.0759

Table 3. Mean values of the time-domain and frequency-domain features extracted from oximetry recordings in the training set.

positive subjects due to the repetition of apnoea during sleep. However, the values obtained for $ApEn$ tend to be similar in both populations.

The individual ability of each feature in classification and regression problems proposed in the Thesis was evaluated. For classification, ROC analysis from data in the training set was performed to determine the decision threshold for every feature. The value that provided the highest classification accuracy on training data was selected as optimum. Subsequently, it was used to classify signals in the test set. Table 4 shows the decision threshold determined for the 14 features under analysis. Classification results are summarised in Table 5.

The utility of each feature to approximate the AHI was also assessed. A linear regression model was built from each feature to estimate the AHI. The form of the regression equation is given by:

$$y = w_1 x_i + w_0, \quad (205)$$

where x_i represents one of the input features while w_1 and w_0 are the adaptive parameters of the model. Only training data was used to define the regression equation using sum-of-squares error minimisation. The utility of each model in regression analysis was evaluated using the *ICC*. Table 6 summarises the coefficients of the models. Results for AHI prediction are shown in Table 7.

Feature	Threshold
μ_i	95.7088
σ_i	0.8914
γ_i	0.3459
δ_i	4.1923
<i>ApEn</i>	0.3559
<i>CTM</i>	0.9946
<i>LZC</i>	0.2258
μ_f	0.0095
σ_f	0.0002
γ_f	8.8848
δ_f	202.2964
S_T	0.6164
S_B	0.1140
<i>PA</i>	17.7680

Table 4. Optimum decision threshold determined from the training set for each of the extracted features.

Feature	<i>Tn</i>	<i>Fp</i>	<i>Tp</i>	<i>Fn</i>	<i>Se</i> (%)	<i>Sp</i> (%)	<i>Acc</i> (%)	<i>AUC</i>	κ
μ_i	18	30	88	8	91.67	37.50	73.61	0.814	0.33
σ_i	40	8	81	15	84.38	83.33	84.03	0.932	0.65
γ_i	12	36	92	4	95.83	25.00	72.22	0.700	0.25
δ_i	29	19	73	23	76.04	60.42	70.83	0.774	0.36
<i>ApEn</i>	1	47	87	9	90.63	2.08	61.11	0.371	-0.09
<i>CTM</i>	44	4	86	10	89.58	91.67	90.28	0.951	0.79
<i>LZC</i>	15	33	87	9	90.63	31.25	70.83	0.671	0.25
μ_f	34	14	82	14	85.42	70.83	80.56	0.862	0.56
σ_f	13	35	90	6	93.75	27.08	71.53	0.679	0.25
γ_f	35	13	72	24	75.00	72.92	74.31	0.851	0.45
δ_f	23	25	88	8	91.67	47.92	77.08	0.795	0.43
S_T	38	10	84	12	87.50	79.17	84.72	0.912	0.66
S_B	44	4	88	8	91.67	91.67	91.67	0.968	0.82
<i>PA</i>	43	5	84	12	87.50	89.58	88.19	0.970	0.74

Table 5. Classification results individually achieved by each of the extracted features. *Tn*: true negatives; *Fp*: false positives; *Tp*: true positives; *Fn*: false negatives; *Se*: sensitivity; *Sp*: specificity; *Acc*: accuracy; *AUC*: area under the ROC curve; κ : Cohen's kappa coefficient.

Feature	w_0	w_1
μ_t	407.20	-4.08
σ_t	19.97	1.29
γ_t	24.02	-20.05
δ_t	86.77	-15.30
$ApEn$	-1.27	57.94
CTM	250.32	-233.52
LZC	-73.56	389.82
μ_f	-22.70	3969.70
σ_f	-3.00	76268.00
γ_f	72.15	-5.28
δ_f	47.33	-0.13
S_T	19.65	2.04
S_B	21.40	3.56
PA	21.11	0.02

Table 6. Coefficients of the linear regression models developed for each of the input features.

Feature	Tn	Fp	Tp	Fn	Se (%)	Sp (%)	Acc (%)	ICC	E_{RMS}	$E_{rel} (-/+)$
μ_t	1	47	96	0	100.00	2.08	67.36	0.48	21.96	(8.35/0.48)
σ_t	0	48	96	0	100.00	0.00	66.67	0.54	22.08	(11.30/0.46)
γ_t	3	45	95	1	98.96	6.25	68.06	0.19	24.92	(12.13/0.60)
δ_t	12	36	91	5	94.79	25.00	71.53	0.49	20.95	(10.19/0.55)
$ApEn$	0	48	96	0	100.00	0.00	66.67	0.02	26.76	(15.69/0.59)
CTM	0	48	96	0	100.00	0.00	66.67	0.77	16.90	(9.23/0.36)
LZC	7	41	89	7	92.71	14.58	66.67	0.40	22.82	(12.07/0.57)
μ_f	24	24	87	9	90.63	50.00	77.08	0.68	19.42	(6.59/0.57)
σ_f	5	43	96	0	100.00	10.42	70.14	0.32	23.65	(13.07/0.55)
γ_f	10	38	95	1	98.96	20.83	72.92	0.34	23.19	(9.52/0.61)
δ_f	6	42	96	0	100.00	1.25	70.83	0.27	23.85	(11.55/0.61)
S_T	0	48	96	0	100.00	0.00	66.67	0.51	25.56	(11.22/0.47)
S_B	0	48	96	0	100.00	0.00	66.67	0.46	21.75	(11.92/0.47)
PA	0	48	96	0	100.00	0.00	66.67	0.46	23.80	(11.73/0.49)

Table 7. Regression results individually achieved by each of the extracted features. Tn : true negatives; Fp : false positives; Tp : true positives; Fn : false negatives; Se : sensitivity; Sp : specificity; Acc : accuracy; ICC : intra-class correlation coefficient; E_{RMS} : root mean square error; $E_{rel} (-/+)$: relative error per subject (SAHS-negative group/SAHS-positive group).

The results show that the proposed input features can provide useful information for classification purposes. Specifically, the discrimination capability of frequency-domain features was higher than that of time-domain features. The classification accuracy achieved by S_T , S_B and PA was 84.72%, 91.67% and 88.19%,

respectively. It confirms that spectral analysis in the band between 0.010 and 0.033 Hz is a useful approach for identifying SAHS, as suggested by Zamarrón *et al.* (2003). In addition, time-domain features such as σ_t and *CTM* provided significant results in the classification problem. In particular, *CTM* achieved an accuracy of 90.28%. In contrast, *ApEn* provided 61.11% accuracy, which was the smallest value of all the extracted features. It was observed that *ApEn* had no classification ability by itself. On the other hand, features exhibited lower performance in the regression problem. The results show that there is not a clear linear relationship between the AHI and any of the features. The best linear estimator was built from *CTM*, with an *ICC* of 0.77. Feature μ_f also achieved significant performance in regression analysis (*ICC* of 0.68) among the extracted features. It should be noted that diagnostic results provided by the estimated AHI were very poor, reflecting a low reliability of the individual methods. Similarly to the classification problem, the smallest value of *ICC* (0.02) was obtained from *ApEn*.

8.5. Feature normalisation and dimensionality reduction

The range of variation differs from one feature to another, as can be observed from data shown in Table 3. Thus, each feature was linearly transformed to have a zero mean and unit variance distribution. As a result, differences between the magnitudes of the input features were removed (Bishop, 1995). Once features are normalised, they can be used as input to the pattern recognition algorithm. Additionally, techniques for feature selection or dimensionality reduction can be considered before pattern analysis.

PCA has been commonly used for dimensionality reduction in the design of pattern recognition algorithms. As suggested by the curse of dimensionality, there might be some benefit of using a smaller input feature set since model complexity, and hence variance, is reduced. For each pattern recognition method proposed in the Thesis, the effect of PCA was analysed by comparing the results obtained with and without this dimensionality reduction technique.

A new set of 14 variables or principal components was obtained from PCA analysis of the normalised features. Figure 16 depicts the results from the PCA transformation. The percentage of variance explained for each of the components is

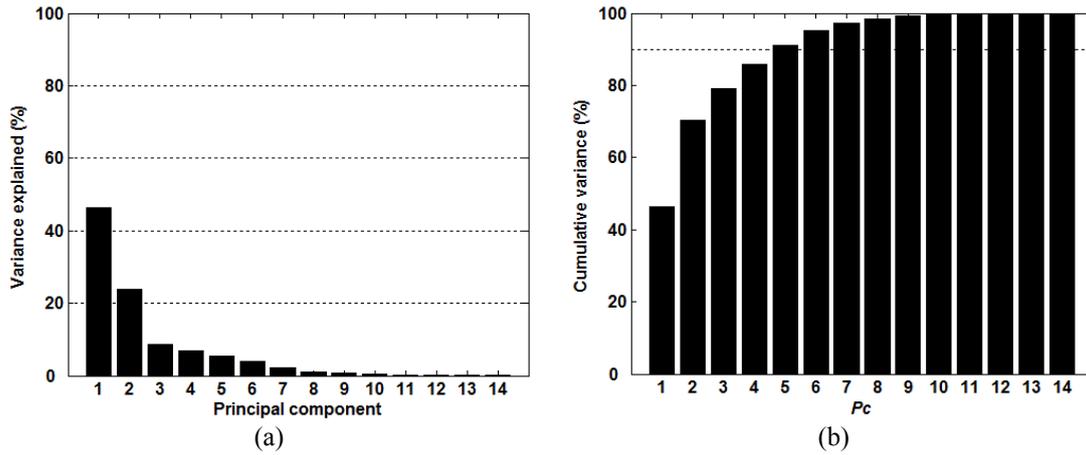


Figure 16. Results from PCA analysis of the normalised features: (a) variance explained for each of the components and (b) cumulative variance for the first P_c components.

shown in order to evaluate their individual relevance. In addition, the variance explained by the first P_c components is provided. The order to compute the percentage of cumulative variance was set according to the individual percentage of explained variance. A threshold of 90% on the total variance of the data was used to determine the number of components to be retained. According to this rule, the first 5 components of PCA ($P_c = 5$) were selected since they explained 91.19% of the total variance. Thus, the dimension of the input pattern for classification and regression algorithms using PCA decreased from 14 to 5 features.

8.6. Classification results

8.6.1. Design and training of classification algorithms

Initially, model selection was performed to determine the optimum configuration for each pattern classification method. Design or smoothing parameters influence model complexity. They are problem-specific and must be properly adjusted to achieve high generalisation capability. The results of the model selection phase for each classification method are detailed in the following subsections. The numerical value of these results was not considered and they were reported using graphical representations. The aim was to compare several configurations of a given method with each other to identify the most appropriate one. It was preferred in order to show the effect of smoothing parameters involved in the design. In addition, it enables to evaluate a wide range of values for these parameters.

Two different input patterns were assessed for classifying SaO₂ signals: the pattern composed of the 14 normalised features and the pattern composed of the components retained from PCA. Thus, two algorithms were designed for each classification method. The form “*All-Method_c*” was used to identify the algorithm using the complete feature set for a given classification method. Similarly, “*PCA-Method_c*” was used to denote the classification algorithm fed with the components from PCA.

8.6.1.1. Design of KNN classifiers

The only design parameter for KNN classifiers was the number of nearest neighbours (K) used to estimate posterior probabilities. Its value influences the complexity of the decision boundary defined by the classifier, with high bias (smooth boundaries) corresponding to high values of K (Bishop, 1995). The minimum K is 1 while its maximum value is given by the number of samples in the training set. Its influence on classification performance of *All-KNN_c* and *PCA-KNN_c* algorithms was analysed by setting K from 1 to 90. Figure 17 shows the value of classification accuracy versus K for both approaches.

Parameter K had a similar effect on both algorithms. The highest classification performance was observed for smaller values of K . Indeed, accuracy reduced as K was progressively increased. The optimum value of K was found to be 3

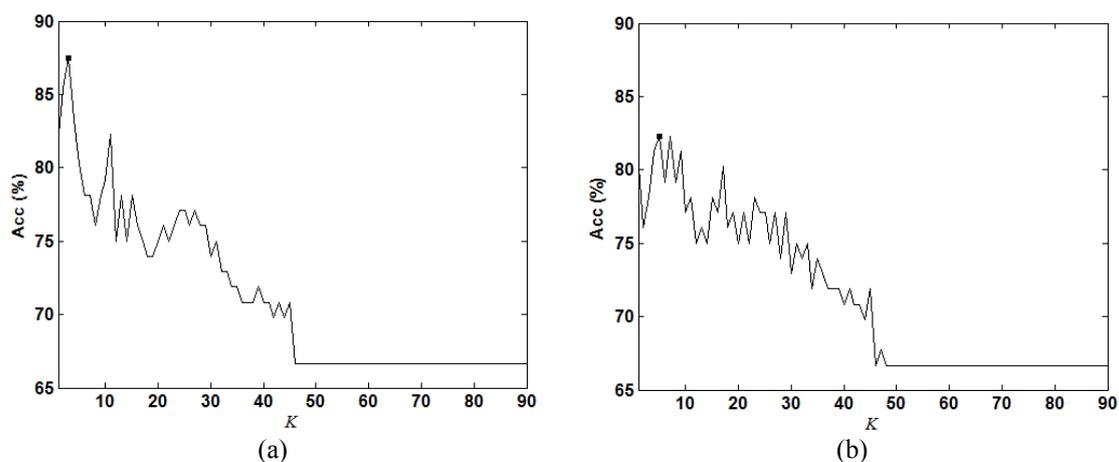


Figure 17. Influence of the number of nearest neighbours (K) on classification accuracy for (a) the *All-KNN_c* classifier and (b) the *PCA-KNN_c* classifier.

for the $All-KNN_c$ algorithm while $K = 5$ and $K = 7$ provided the highest accuracy for $PCA-KNN_c$. Finally, the configuration with $K = 5$ was selected for this algorithm since classification accuracy tended to decrease with the number of nearest neighbours.

8.6.1.2. Design of FLD classifiers

A model selection phase was not required for FLD classifiers since any design parameter must be adjusted. Therefore, a unique FLD algorithm is derived from data in the training set using the optimisation process described above. The coefficients obtained for $All-FLD_c$ and $PCA-FLD_c$ algorithms are given in Tables 8 and 9. It should be noted that these coefficients do not directly apply to time-domain and frequency-domain features in the $All-FLD_c$ algorithm, since they were previously normalised.

New input patterns are classified by assigning it to the closest class in the

Feature	Coefficient
μ_t	0.0006
σ_t	-0.0242
γ_t	0.0064
δ_t	0.0192
$ApEn$	0.0060
CTM	-0.0303
LZC	0.0017
μ_f	-0.0277
σ_f	0.0094
γ_f	0.0849
δ_f	-0.0649
S_T	-0.0829
S_B	0.0476
PA	0.0374

Table 8. Coefficients of the linear equation for the $All-FLD_c$ algorithm.

Feature	Coefficient
PC_1	0.0006
PC_2	-0.0242
PC_3	0.0064
PC_4	0.0192
PC_5	0.0060

Table 9. Coefficients of the linear equation for the $PCA-FLD_c$ algorithm.

projected space. It requires to evaluate the centroid associated to each of the classes, i.e. the projected value for the mean vector of each class. Therefore, the decision threshold in the projected space is given by the middle point between the centroids of both classes.

The magnitude and sign of the coefficients in FLD equations are often used to interpret the influence of the input variables on the classifier output. However, high degrees of collinearity (high correlation among input features) may be obtained, making it difficult to interpret the contribution of each independent feature (Amick and Walberg 1975). Thus, the coefficients shown in the tables above must be interpreted together with all other coefficients since they do not provide useful information by themselves.

8.6.1.3. Design of LR classifiers

Similarly to FLD, any smoothing parameter must be adjusted to design LR algorithms. Once the training set is fixed, the LR model is unique for a given set of input features. Therefore, *All-LR_c* and *PCA-LR_c* algorithms were derived from training data without a previous model selection process. The coefficients of the equations obtained for these algorithms are provided in Tables 10 and 11, respectively.

Both LR-based algorithms were trained to approximate the posterior probability for the SAHS-positive group. Thus, diagnostic decisions are made according to the Bayes' rule in order to minimise the probability of misclassification.

As indicated for FLD equations, correlation among input features may lead to apparently incoherent values of the coefficients with respect to the individual behaviour of each feature. For instance, coefficients associated to σ_t , S_B or PA reflect an opposite influence on the decision of the *All-LR_c* algorithm regarding the individual correlation of these features with SAHS. The coefficients listed above should be interpreted as the weight of each feature to the output value in the presence of the other features (Rencher, 1992). Thus, the individual analysis of each coefficient does not provide useful knowledge about the classification rule implemented by the algorithm. Previous studies analysed the influence of correlation between input features on their respective coefficients in LR models (Abbot and Carrol 1984).

Feature	Coefficient
w_0	4.55
μ_t	1.18
σ_t	-37.01
γ_t	15.48
δ_t	-2.70
<i>ApEn</i>	-2.07
<i>CTM</i>	-7.16
<i>LZC</i>	-0.23
μ_f	8.48
σ_f	-5.12
γ_f	-13.16
δ_f	8.78
S_T	52.94
S_B	-8.97
<i>PA</i>	-11.34

Table 10. Coefficients of the equation derived for the *All-LR_c* algorithm.

Feature	Coefficient
w_0	4.93
PC_1	4.67
PC_2	2.12
PC_3	1.17
PC_4	2.20
PC_5	-0.75

Table 11. Coefficients of the equation derived for the *PCA-LR_c* algorithm.

8.6.1.4. Design of PNN classifiers

PNN classifiers are based on the Parzen window procedure to approximate class-conditional probabilities. Spherical Gaussian kernels were assumed. The accuracy of the approximation highly depends on the width of the Gaussian kernels, which must be defined by the user. In addition, the same width σ_p was considered for all of them since the goal is to evaluate the contribution of each kernel under the same conditions (Specht, 1990). It has been shown that the decision boundary of PNN classifiers varies continuously from a hyperplane when σ_p tends to infinity to a very non-linear boundary representing the 1-nearest neighbour classifier when it is close to 0 (Specht, 1990). The influence of this parameter in the proposed classification problem was assessed by

varying its value in a wide range. The spread (Sp), defined as the distance from the centre for which the output of the Gaussian kernel is 0.5, was used to define the range of σ_p values since it represents a more intuitive measurement. The expression relating σ_p and Sp is given by:

$$Sp = \sqrt{\ln(4)}\sigma_p, \quad (206)$$

where the normalisation factor for the Gaussian has been omitted. For each training pattern \mathbf{x}^n , the mean distance (r^n) to its K -nearest neighbours ($K = 3$) was computed. The average (r_{avg}) of all these distances was taken as reference. A value proportional to r_{avg} was used to define Sp :

$$Sp = \tau r_{avg}, \quad (207)$$

where τ acts as a smoothing parameter that must be empirically determined. The value of τ was varied from 0.1 to 5 to assess its influence on generalisation performance. This range is large enough to cover PNN models with narrow and wide Gaussian kernels. This parameter determines the shape of the approximation to class-conditional probabilities $p(\mathbf{x}|\omega_i)$. It becomes smoother and tends to be Gaussian for higher values of σ_p . In contrast, a small σ_p causes the estimated density to have distinct modes corresponding to the locations of the training samples. In general, neither of these limiting cases will provide good generalisation ability. The effect of τ on classification performance was evaluated using leave-one-out cross-validation on training data. The results obtained for *All-PNN_c* and *PCA-PNN_c* algorithms are depicted in Figure 18.

A similar effect was observed for both models. Small τ resulted in an excessively complex decision boundary due to overfitted estimation of the class-conditional probability density functions. Similarly, underfitting was produced for high values of τ greater than 1.5, leading to oversmoothed approximations of the density functions. The balance between these situations was obtained with $\tau = 0.9$ for both models (*All-PNN_c* and *PCA-PNN_c*).

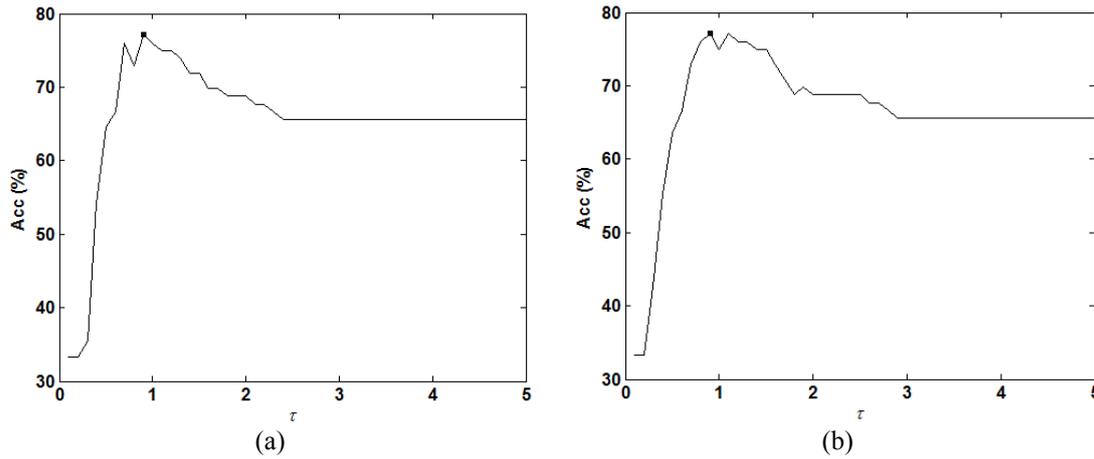


Figure 18. Influence of the spread parameter (controlled by τ) on classification accuracy for (a) the *All-PNN_c* classifier and (b) the *PCA-PNN_c* classifier.

8.6.1.5. Design of MLP classifiers

The design of MLP classifiers requires the choice of appropriate values for the regularisation parameter (ν) and the number of nodes in the hidden layer (N_H). MLP networks with a single hidden layer were considered since they are capable of universal approximation, i.e. they can approximate any continuous input-output mapping to any desired degree of accuracy, given a sufficient number of hidden units is available (Hornik, 1991; Haykin, 1996). Parameters ν and N_H affect the complexity of the network and hence its generalisation capability. Defining a high value of N_H increases the number of network weights and the complexity of the model. On the other hand, weight decay regularisation represented by parameter ν aims to control network complexity by favouring small weights. As indicated by Moody (1992), the effective complexity of the network depends on the value of this parameter. Increasing the value of ν involves a reduction in network complexity, leading to smoother mapping functions. However, making ν too high may lead to excessively simple networks. Thus, a suitable choice of ν and N_H must be made in order to avoid poor generalisation due to overfitting or underfitting.

In addition, other decisions are needed to complete the design of MLP classifiers. As explained, MLP training requires to minimise an error function that presents a non-linear dependence with respect to network weights. Conventional gradient descent can be used for this purpose. However, faster convergence can be

gained using more advanced algorithms. Specifically, the scaled conjugate gradient (SCG) was used for weight optimisation (Moller, 1993). It does not require critical user-dependant parameters and is considerably faster than gradient descent. The maximum number of iterations during training (i.e. the number of times that the whole training set is presented to the MLP network) was set to 100. Other MLP-based classifiers were successfully trained in previous studies using such a configuration of the SCG algorithm for a similar number of training samples (Marcos *et al.* 2008b; Marcos *et al.* 2010a). A moderate value must be chosen for this parameter in order to reduce training time and preventing overfitting. Differentiable activation functions must be selected for all the network nodes. This condition enables error propagation from output to hidden layer weights (backpropagation). The hyperbolic tangent function was used for nodes in the hidden layer since it provides a fast convergence of training algorithms (Haykin, 1999). It is a sigmoidal function with an output range between -1 and $+1$. The logistic function was used for the single output node of the MLP classifiers evaluated in the Thesis. This function maps any real value into the interval between 0 and 1, which enables to interpret network outputs as probabilities. Finally, weight initialisation must be addressed. As suggested by Bishop (1995), weights and biases were initially generated from a spherically symmetric zero-mean Gaussian distribution. The variance of the Gaussian was set to $1/(d + 1)$ and $1/(N_H + 1)$ for weights and biases in the hidden and the output layers, respectively. As a result, the summed inputs to the sigmoidal functions are close to one, which ensures fast convergence of training.

An exhaustive model selection process was carried out by evaluating multiple combinations of parameters N_H and ν . The former was varied from 2 up to 50 nodes. A large collection of values for α was also assessed by varying this parameter between 0.01 and 5. The aim was to observe the trend in classification performance. Due to random initialisation of weights, a trained MLP network is not a unique solution even if the training set, N_H , ν and the remaining design parameters keep fixed. Thus, the complete leave-one-out cross-validation process was repeated 10 times for every pair of N_H and ν . The average accuracy was used to determine the performance of a given configuration. The effect of N_H and ν on classification performance for *All-MLP_c* and *PCA-MLP_c* algorithms is shown in Figure 19.

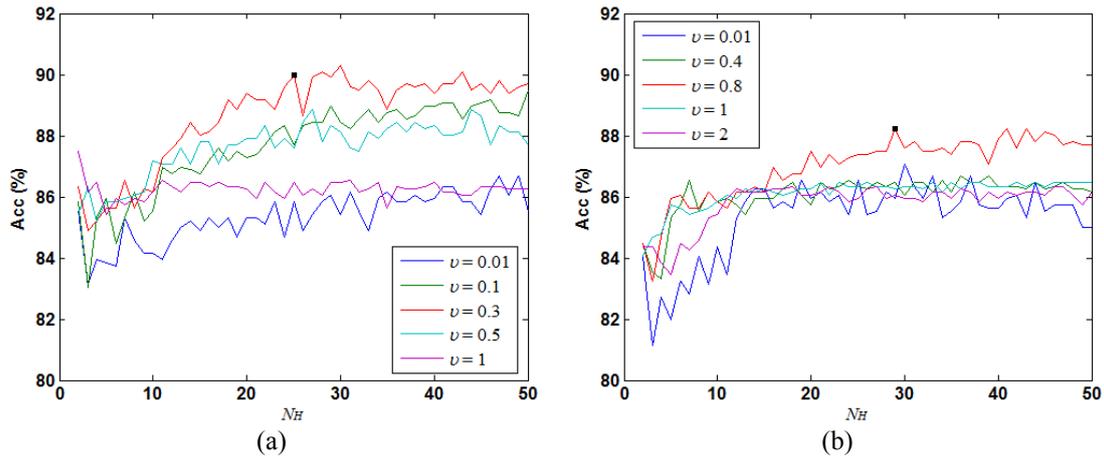


Figure 19. Influence of the number of hidden nodes (N_H) and the regularisation parameter (ν) on classification accuracy for (a) the *All-MLP_c* classifier and (b) the *PCA-MLP_c* classifier.

The results show that weight decay regularisation improved classification performance. MLP classifiers trained with ν small were affected by overfitting and provided poor classification accuracy. Network complexity was reduced by progressively increasing ν , resulting in improved accuracy. This behaviour was observed when ν varied from 0.01 to 0.3 for *All-MLP_c* classifiers and from 0.01 to 0.8 for *PCA-MLP_c* classifiers. For both models, higher values of ν provided lower classification accuracy due to an excessive reduction of network complexity. Therefore, $\nu = 0.3$ and $\nu = 0.8$ were selected for *All-MLP_c* and *PCA-MLP_c*, respectively.

Once the regularisation parameter was fixed, the effect of varying N_H was analysed. Performance curves were similar for both classifiers. Initially, accuracy was improved by gradually increasing N_H , which indicates that some degree of complexity is required to model the problem. However, this tendency stopped for N_H larger than a given value. From this point, classification performance did not significantly vary by setting N_H higher. Therefore, $N_H = 25$ was selected for the *All-MLP_c* classifier while $N_H = 29$ was chosen for the *PCA-MLP_c* algorithm. As can be appreciated in Figure 19, these values of N_H approximately correspond with the starting point of saturated performance.

Network configurations with higher N_H showed similar performance to that finally selected since regularisation prevented them to overfit the data. Model complexity was controlled by reducing the magnitude of the weights. To illustrate the

effect of weight decay regularisation, the size of the weights was analysed for different network configurations of the *All-MLP_c* and *PCA-MLP_c* classifiers. Figure 20 shows the percentage of network weights under a threshold value (u_w) for several configurations of the *All-MLP_c* classifier in two different cases. In the first one, ν was fixed to the value previously selected and networks with different N_H were analysed. In the second case, the optimum N_H was considered while ν was varied. The same analysis was performed for the *PCA-MLP_c* classifier. The results are depicted in Figure 21. As expected, the weights became smaller as ν was increased. In addition, smaller weights were obtained when more hidden nodes were added in order to prevent an increment of network

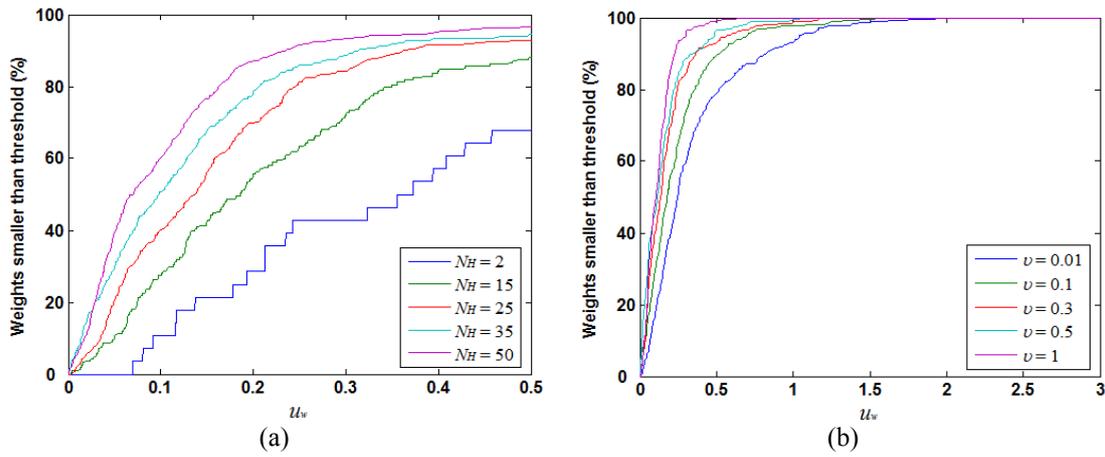


Figure 20. Effect of weight decay regularisation on the size of network weights: (a) *All-MLP_c* algorithms with $\nu = 0.3$ and different N_H ; (b) *All-MLP_c* algorithms with $N_H = 25$ and different ν .

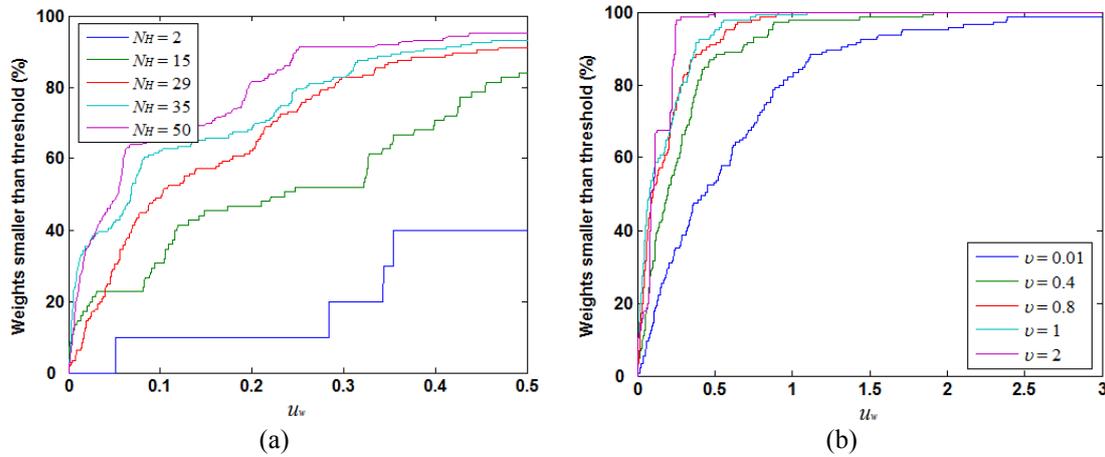


Figure 21. Effect of weight decay regularisation on the size of network weights: (a) *PCA-MLP_c* algorithms with $\nu = 0.8$ and different N_H ; (b) *PCA-MLP_c* algorithms with $N_H = 29$ and different ν .

complexity. These results show the link between the magnitude of the weights and network complexity. Large networks trained with weight decay regularisation can provide high classification performance as their weights are kept small after training (Bartlett, 1998).

8.6.1.6. Design of RBF classifiers

The first design decision for RBF networks is the selection of the kernel function. RBF networks with Gaussian kernels were evaluated in the Thesis. The k -means algorithm was used to determine their centres (Schwenker *et al.* 2001). The number of Gaussian kernels (N_B) and the width of each of them (σ_{r_j}) were defined in the design of RBF networks. As explained before, the combination of the basis functions represents an estimation for the unconditional probability of the input patterns. Therefore, both parameters must be chosen in order to cover the entire region defined by samples \mathbf{x} in the training set. The spread value (Sp_j), as defined in Eq. (206), was used instead of σ_{r_j} to calibrate the width of each Gaussian kernel. The mean distance (r_i) from each kernel centre to its K -nearest centres ($K = 3$) was used as a reference to determine its width according to the following expression (Schwenker *et al.* 2001):

$$Sp_j = \tau r_j, \quad (208)$$

where the smoothing parameter τ must be optimised empirically. Its influence on classification performance was analysed by varying it between 0.1 and 5, i.e. a large range of possible values for the width of the basis functions was explored. It should be noted that a different width parameter was defined for each kernel function. Initially, it is expected that the non-linear behaviour of the RBF classifier will become more pronounced for smaller τ . In this case, the overlapping between Gaussian kernels is reduced and the estimated distribution of patterns \mathbf{x} tends to present an irregular shape. Increasing τ causes the Gaussian kernels to be overlapped and the estimated density tends to be Gaussian as well. As a result, the decision boundary defined by the RBF classifier becomes smoother. Network bias and variance are also affected by the choice of N_B . A reduced number of units may be insufficient to provide an accurate representation of the target distribution. Conversely, overfitting may arise when N_B is too high due to an excessive adaptation to training data. The value of N_B ranged from 2

to 50 in the experiments with RBF classifiers, which enables to compare simple and highly complex networks. Due to random initialisation of the k -means algorithm for centre selection, each network configuration was evaluated in 10 different runs of the leave-one-out process. Average accuracy was measured to perform model selection. The effect of σ_{ij} and N_B on $All-RBF_c$ and $PCA-RBF_c$ algorithms is depicted in Figure 22.

There were little differences between the performance curves obtained for $All-RBF_c$ and $PCA-RBF_c$ algorithms by varying τ and N_B . As expected, too small values of τ resulted in an excessively complex form for the estimated distribution of the input data, leading to poor generalisation. Classification accuracy was improved by progressively increasing τ , i.e. reducing the complexity of the mapping. It was observed that high classification performance was obtained with τ greater than 0.5 for $All-RBF_c$ and $PCA-RBF_c$ models. It ensures some degree of overlapping between Gaussian kernels and smooth decision boundaries.

The analysis of N_B shows that networks with a small number of units are not complex enough to model the problem and suffer from underfitting. Classification accuracy progressively improved when this parameter varied from 2 until a value between 7 and 10. However, more complex networks (i.e. those with higher N_B) reached lower accuracy due to overfitting. Finally, a configuration with $\tau = 3$ and $N_B = 10$ provided the highest accuracy for the $All-RBF_c$ classifier and was selected as optimum. For the $PCA-RBF_c$ algorithm, $\tau = 1.5$ and $N_B = 7$ were chosen.

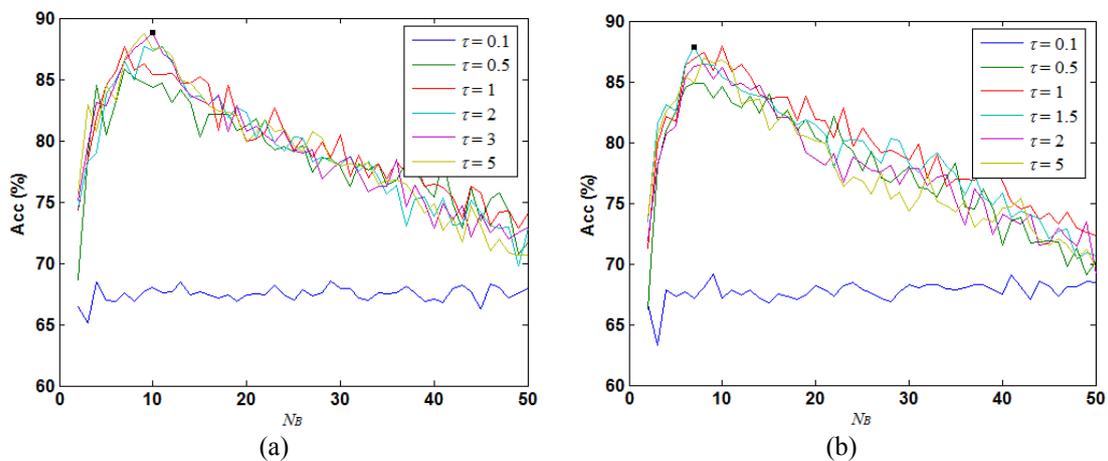


Figure 22. Influence of the number of basis functions (N_B) and the kernel width parameter (controlled by τ) on classification accuracy for (a) the $All-RBF_c$ classifier and (b) the $PCA-RBF_c$ classifier.

8.6.1.7. Design of BY network classifiers

The architecture and design of BY networks was similar to that used for MLP classifiers trained according to the maximum likelihood principle. The same decisions were adopted for the number of layers, the number of neurons in the output layer or the type of activation functions. As detailed before, the Gaussian approximation to $p(\mathbf{w}|D)$ was used to train BY networks. The regularisation parameter has not to be specified since it is optimised during training according to the evidence procedure (MacKay, 1992b). ARD was used to identify the most relevant features for classification. Network weights were allocated into different groups and a specific weight decay parameter α_k was used for each of them. Weights associated to each of the input features were separately considered. Additional groups were those composed of the biases parameters in the hidden layer as well as weights and bias associated to the output neuron. A zero-mean Gaussian prior distribution with a variance equal to $1/\alpha_k$ is assumed for network weights and biases in each group. The scaled conjugate gradient was used to minimise $S(\mathbf{w})$ to find \mathbf{w}_{MP} in order to obtain the Gaussian approximation of $p(\mathbf{w}|D)$. A training cycle consisted of weight optimisation during 10 training epochs followed by the re-estimation of the hyperparameters. A total of 10 cycles were carried out to implement BY classifiers. The initial value of all the hyperparameters was set to 0.1. Weights were initialised as described for MLP network classifiers instead of using the prior distribution $p(\mathbf{w})$ associated to the initial α_k . Usually, a small starting value is used for the hyperparameters in order to reflect little prior knowledge about the weights. Using the initial prior distribution $p(\mathbf{w})$ for network initialisation may lead to large magnitudes of the weights. As a result, it is likely that the final weights will also be large due to the local nature of the optimisation process. In this case, the final solution is expected to be unsatisfactory (Nabney, 2002).

The size of the hidden layer (N_H) is the only design parameter to be optimised. The effect of model complexity on generalisation performance was analysed by varying N_H from 2 to 50, i.e. a wide variety of network architectures were compared, including simple and complex models. The classification accuracy of each configuration was used to determine the optimum network size. Due to random initialisation of the adaptive network weights, each network configuration was validated 10 times using

leave-one-out cross-validation and the average accuracy was obtained. Design results for *All-BY_c* and *PCA-BY_c* classifiers are represented in Figure 23.

For both algorithms, there were no significant differences in classification accuracy when N_H ranged from 2 to 10, approximately. Subsequently, accuracy decreased by gradually adding more hidden nodes. A similar behaviour was observed in a previous study using BY network classifiers for SAHS diagnosis from a different database of SaO₂ signals (Marcos *et al.* 2010a). The results point out that the evidence procedure is not properly working for large networks. As indicated by MacKay (1992a), the Gaussian approximation breaks down when the number of adaptive parameters is too large in comparison with the number of training samples. Specifically, for the practical problems analysed by MacKay, it was found that the ratio (r_w) between the number of training samples and the number of adaptive parameters should be greater than 3 ± 1 for a valid Gaussian approximation. Figure 24 depicts the evolution of this ratio as a function of N_H for BY network architectures for an input pattern with the complete set of features and in the case of only using those inputs selected from PCA. According to the results shown in Figure 23, the method seems to be effective for models with N_H up to 10, which corresponds with r_w equal to 0.60 samples per weight for *All-BY_c* and 1.35 samples per weight for *PCA-BY_c* networks. The highest accuracy was achieved with $N_H = 8$ and $N_H = 2$ for *All-BY_c* and *PCA-BY_c* algorithms, respectively.

Finally, the relative influence of each input feature on the network output

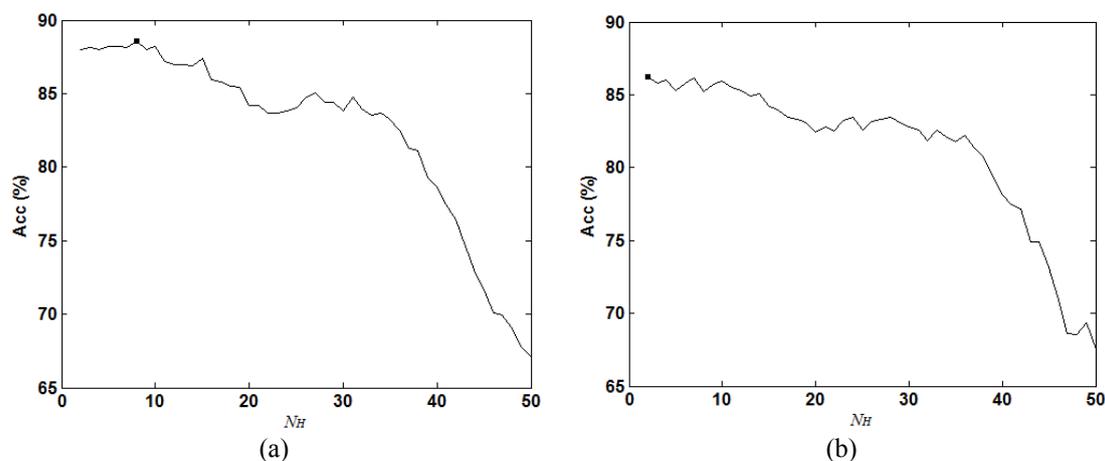


Figure 23. Influence of the number of hidden nodes (N_H) on classification accuracy for (a) the *All-BY_c* classifier and (b) the *PCA-BY_c* classifier.

value was assessed by using the evidence procedure and ARD. Figure 25 shows the hyperparameters α_k for the selected configuration of the $BY\text{-}MLP_c$ network. The results show that classification is mainly carried out from information in S_T and σ_t features. In addition, the first three standard moments of the frequency component (μ_f , σ_f and γ_f) as well as kurtosis computed in the time domain (δ_t) have a relevant influence on the output decision since they are associated to small α_k .

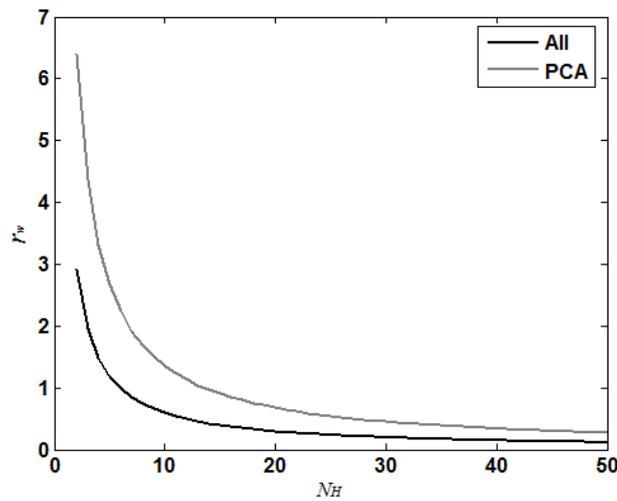


Figure 24. Evolution of the ratio (r_w) between the number of training samples and the number of adaptive parameters in a BY network architecture using all the input features and a BY network architecture using the components selected from PCA.

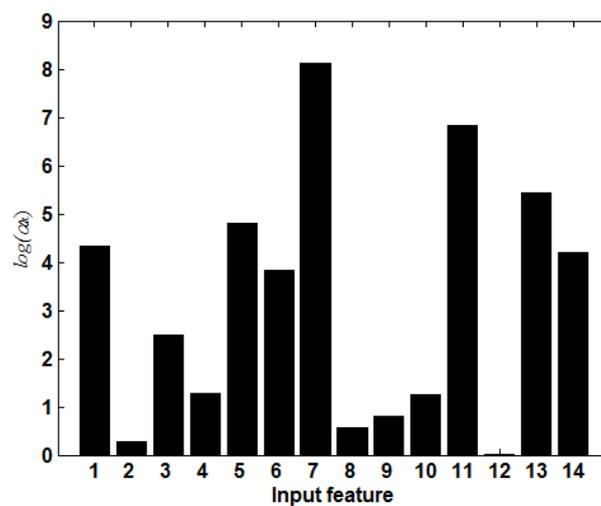


Figure 25. Results from ARD for the $All\text{-}BY_c$ classification algorithm. The logarithm (base 10) of α_k is represented.

8.6.1.8. Design of SVM classifiers

Different kernel functions can be used to implement SVM classifiers. In this Thesis, Gaussian kernels were used. Despite their architecture is the same as RBF networks, SVM represent a different implementation of machine learning algorithms (Vapnik, 1999). Model selection was carried out in order to find optimum values for the regularisation parameter C and the kernel width σ_s . As C decreases, the margin maximisation term dominates: model complexity is reduced and less emphasis is placed on reducing classification errors on the training set. Due to margin maximisation, more training patterns fall between the margins defined by the SVM classifier. Beyond a certain point, all of them will be selected as support vectors with $\alpha_i = C$ (i.e. any of the support vectors is on the margins). Thus, too small values of C will lead to very low complex classifiers that could be affected by underfitting. Conversely, model complexity is increased for higher values of C since minimisation of the training error becomes more relevant. If C tends to infinity, the margin between classes is reduced as well as the number of support vectors. Training samples selected as support vectors will be of the type $0 < \alpha_i < C$ and $\xi_i = 0$, i.e. all of them are on the margins. The resulted SVM algorithms will be highly complex and may overfit training data. Parameter σ_s also has a direct influence on model complexity. If it is set too small, the non-linear behaviour of the kernel function is enhanced. For σ_s very close to zero, all training vectors are selected as support vectors, i.e. the classifier is completely adapted to the data and will suffer from overfitting. In contrast, for σ_s very large, all points in the input space are treated as the same point, showing a complete lack of flexibility, i.e. underfitting is produced.

There is no prior knowledge about optimum values for C and σ_s . Therefore, an exhaustive search on the range of possible values is required. An exponentially growing sequence was assigned to both parameters, with $C \in \{10^{-4}, 10^{-3}, \dots, 10^8\}$ and $\sigma_s \in \{10^{-4}, 10^{-3}, \dots, 10^8\}$ (Hsu and Lin 2002). Every pair of C and σ_s values was evaluated. Performance curves obtained for *All-SVM_c* and *PCA-SVM_c* classifiers are depicted in Figure 26.

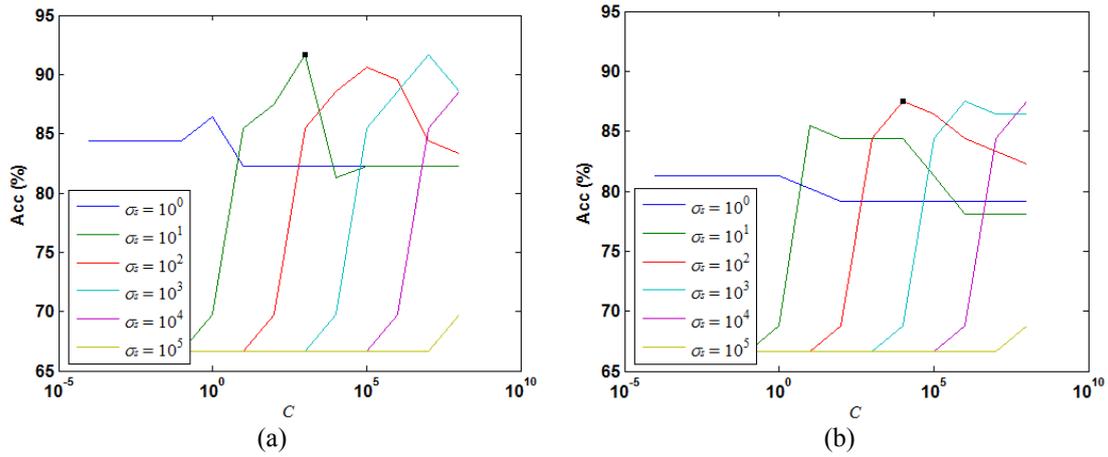


Figure 26. Influence of the regularisation parameter (C) and the kernel width parameter (σ_s) on classification accuracy for (a) the *All-SVM_c* classifier and (b) the *PCA-SVM_c* classifier.

Accuracy values differ from *All-SVM* to *PCA-SVM* models. However, the influence of C and σ_s (determined by the profiles of the curves) was quite similar. High classification performance was achieved with σ_s ranging from 10^1 to 10^4 for both models. A parallel evolution can be observed for accuracy curves corresponding to σ_s values in that range. It reflects that higher C (increased complexity) must be chosen when σ_s is increased (reduced complexity) in order to achieve similar performance. Therefore, several combinations of parameters C and σ_s can be selected to provide high generalisation ability, i.e. to achieve an optimum balance between bias and variance. Finally, the configuration with $C = 10^3$ and $\sigma_s = 10$ was chosen for the *All-SVM_c* classifier. These parameters were set to $C = 10^4$ and $\sigma_s = 10^2$ for the *PCA-SVM_c* algorithm.

8.6.2. Testing classification algorithms

The entire training set with 96 subjects was used to implement the definitive classification algorithms for each method using the optimum configurations determined in the model selection process. A total of 16 different algorithms were obtained since each classification method was implemented with and without PCA preprocessing. The generalisation ability of these algorithms was estimated by using an independent test set composed of 144 previously unseen patterns. Classification accuracy was the performance measure used for algorithm comparison. Table 12 summarises the results obtained in the test set.

Algorithm	Tn	Fp	Tp	Fn	Se (%)	Sp (%)	Acc (%)	AUC	κ
<i>All-KNN_c</i>	41	7	85	11	88.54	85.42	87.50	0.895	0.73
<i>PCA-KNN_c</i>	36	12	86	10	89.58	75.00	84.72	0.920	0.65
<i>All-FLD_c</i>	44	4	82	14	85.42	91.67	87.50	0.947	0.73
<i>PCA-FLD_c</i>	45	3	84	12	87.50	93.75	89.58	0.950	0.78
<i>All-LR_c</i>	45	3	83	13	86.46	93.75	88.89	0.935	0.76
<i>PCA-LR_c</i>	40	8	86	10	89.58	83.33	87.50	0.953	0.72
<i>All-PNN_c</i>	23	25	82	14	85.42	47.92	72.92	0.808	0.35
<i>PCA-PNN_c</i>	29	19	80	16	83.33	60.42	75.69	0.840	0.44
<i>All-MLP_c</i>	42	6	91	5	94.79	87.50	92.36	0.958	0.83
<i>PCA-MLP_c</i>	39	9	85	11	88.54	81.25	86.11	0.944	0.69
<i>All-RBF_c</i>	40	8	85	11	88.54	83.33	86.81	0.953	0.71
<i>PCA-RBF_c</i>	41	7	82	14	85.42	85.42	85.42	0.948	0.68
<i>All-BY_c</i>	42	6	87	9	90.63	87.50	89.58	0.855	0.77
<i>PCA-BY_c</i>	39	9	85	11	88.54	81.25	86.11	0.949	0.69
<i>All-SVM_c</i>	42	6	87	9	90.63	87.50	89.58	0.951	0.77
<i>PCA-SVM_c</i>	41	7	85	11	88.54	85.42	87.50	0.928	0.73

Table 12. Results achieved on the test set by the 16 classification algorithms implemented in the study. Tn : true negatives; Fp : false positives; Tp : true positives; Fn : false negatives; Se : sensitivity; Sp : specificity; Acc : accuracy; AUC : area under the ROC curve; κ : Cohen's kappa coefficient.

The *All-MLP_c* algorithm achieved the highest classification performance with an accuracy of 92.36% (94.79% sensitivity and 87.50% specificity). The values obtained for AUC (0.958) and κ (0.83) statistics confirm its high classification ability. In addition, other classifiers also achieved significant results with diagnostic accuracy around or higher than 85%. Indeed, only *All-PNN_c* and *PCA-PNN_c* classifiers provided significantly lower results with an accuracy of 72.92% and 75.69%, respectively. It should be noted that the classification performance of linear algorithms based on FLD and LR methods was similar or even better than that of other more complex classifiers based on BY, SVM or RBF networks. Moreover, there were no substantial differences between the algorithms built from these methods and the *All-KNN_c* classifier. Even, it outperformed the *All-RBF_c* algorithm. Therefore, simple techniques such as FLD, LR or KNN can be considered for developing effective algorithms for the proposed classification problem. Their main advantage is that model selection is not required for FLD and LR or, in the case of KNN, it is a straightforward process. However, the algorithms built from these simple methods did not reach the accuracy value provided

by the *All-MLP_c* algorithm. This result suggests that a non-linear decision boundary is required to achieve the optimum performance for the proposed classification problem.

On the other hand, the results reveal that dimensionality reduction using PCA did not improve classification performance for most of the evaluated methods. Only classification algorithms based on FLD and PNN were outperformed by their respective versions using PCA. This technique was suggested as a measure to prevent overfitting due to the curse of dimensionality. However, complex methods for which this phenomenon is more problematic achieved high generalisation capability. For instance, *All-MLP_c* and *All-BY_c* algorithms performed substantially better than *PCA-MLP_c* and *PCA-BY_c*, respectively. The results indicate that controlling model complexity by selecting appropriate design parameters, using regularisation techniques such as weight decay or applying training techniques such as Bayesian learning or SVM principles was more effective to overcome this problem. The discarded components may contain useful information for the classification problem.

Figure 27 shows the number of algorithms that misclassified each of the signals in the test set. It shows that all the algorithms did not tend to fail on the same subjects. Only three subjects were not correctly diagnosed by all the classification algorithms. They correspond to the SAHS-positive group and had AHI values of 11.5, 14.3 and 19.8 h⁻¹ (i.e. two mild and one moderate SAHS patients) (Qureshi and Ballard 2003). Their BMI were 23.19, 27.43 and 20.20 kg/m², respectively. These subjects had

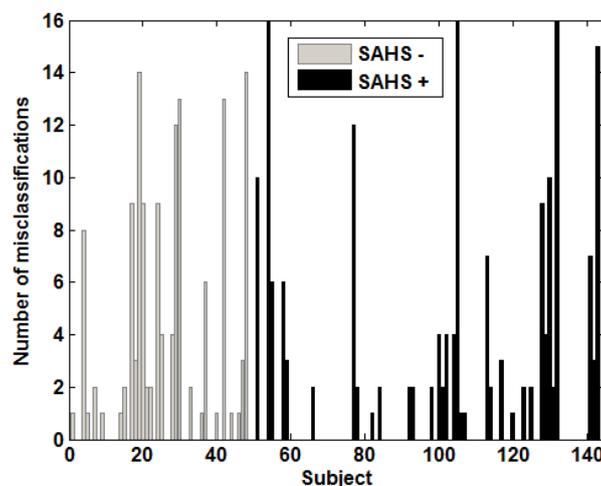


Figure 27. Number of classification algorithms that provided an incorrect decision for each of the subjects in the test set.

CT90 equal to zero and their minimum value of saturation during sleep was 92.1%, 90.5% and 91.7%, respectively. Specifically, if the results of the *All-MLP_c* algorithm (the classifier with the best performance) are analysed, it was observed that it misclassified 11 subjects in the test set: 6 SAHS-negative and 5 SAHS-positive. Their corresponding AHI values as well as the output provided by the algorithm are summarised in Table 13. Most of the subjects with an incorrect diagnosis suffer from mild SAHS. i.e. they are obvious doubtful cases. Three SAHS-negative and four SAHS-positive subjects misclassified by the algorithm have $5 \text{ h}^{-1} \leq \text{AHI} < 15 \text{ h}^{-1}$. This condition is reflected by the network output value (probability of having SAHS) assigned to some of them. For some of the misdiagnosed subjects, the output was around the decision threshold placed at 0.5, which reflects uncertainty about the diagnostic decision of the network. However, it does not represent a tendency to fail for border-line subjects. The algorithm provided a correct diagnosis for most of the mild SAHS cases in the test set: only 7 out of 32 were misclassified.

8.7. Regression results

8.7.1. Design and training of regression algorithms

As explained for classification techniques, model selection was performed in order to determine appropriate values for the design or smoothing parameters of the proposed regression methods. Different configurations were assessed by varying these parameters in a wide range. Leave-one-out cross-validation from data in the training set

<i>All-MLP_c</i>		
Subject	AHI	y
4	4.1	0.8164
17	7	0.5751
19	8.7	0.6091
29	7.6	0.9483
42	4.1	0.5015
48	0.6	0.5419
54	14.3	0.0025
105	19.8	0.0031
132	11.5	0.0024
141	10.0	0.2791
143	13.5	0.0200

Table 13. Subjects in the test set misclassified by the *All-MLP_c* algorithm. AHI: apnoea-hypopnoea index; y : output value provided by the *All-MLP_c* algorithm.

was carried out to compare their performance. The *ICC* was used as the performance measure for model comparison.

A total of 7 regression methods were proposed. Two different algorithms were implemented for each of them depending on the use or not of PCA for dimensionality reduction. A similar nomenclature to that used for classification methods was used to refer regression algorithms. The form “*All-Method_r*” denotes the regression algorithm based on a given method and using the 14 normalised features as the input pattern. The form “*PCA-Method_r*” was used to refer the regression algorithm with the components selected from PCA as the input pattern.

8.7.1.1. Design of MLR algorithms

Any smoothing parameter must be adjusted to implement MLR models. They have a unique solution given a fixed training set. The coefficients of the linear equation of *All-MLR_r* and *PCA-MLR_r* algorithms are summarised in Tables 14 and 15, respectively.

Similarly to FLD and LR models, the interpretation of coefficients in MLR can be problematic due to variable interactions and multicollinearity. Therefore, the

Feature	Coefficient
w_0	24.7490
μ_t	-2.3882
σ_t	-6.7425
γ_t	-0.7288
δ_t	-5.5454
<i>ApEn</i>	-4.5412
<i>CTM</i>	-11.4900
<i>LZC</i>	6.0803
μ_f	14.4140
σ_f	0.9456
γ_f	3.2283
δ_f	0.8954
S_T	13.4510
S_B	-27.0770
<i>PA</i>	14.3820

Table 14. Coefficients of the linear equation of the *All-MLR_r* algorithm.

Feature	Coefficient
w_0	24.7490
PC_1	7.2354
PC_2	-0.3956
PC_3	1.1005
PC_4	12.3990
PC_5	-3.3373

Table 15. Coefficients of the linear equation for the $PCA-MLR_r$ algorithm.

influence of each input feature on the output cannot be directly evaluated from the magnitude and sign of its corresponding coefficient.

8.7.1.2. Design of GRNN algorithms

GRNN algorithms are based on the estimation of the density $p(\mathbf{x},t)$ using the Parzen window method. Gaussian kernels are used for this purpose. The width parameter (σ_g), which is common for all the kernel functions, must be defined to implement these algorithms. The spread value (Sp) was used instead of σ_g to analyse the effect of the width of the Gaussians on generalisation capability. A Gaussian kernel centred on each training sample was used to compute the approximated density. Similarly to PNN algorithms, the mean distance (r^n) of each centre to its K -nearest centres ($K = 3$) was computed. The average (r_{avg}) for the set of r^n values was used to set Sp . A proportional constant τ was used as indicated in Eq. (207).

The kernel width can be considered as a smoothing parameter that controls the complexity of the mapping. As σ_g is increased, the approximated density becomes smoother and tends to a Gaussian profile. Conversely, it is more irregular as σ_g is smaller. A wide range of σ_g values were evaluated by varying τ from 0.1 to 5. It was defined in order to compare GRNN models with different complexity. Figure 28 shows ICC as a function of the kernel width parameter for $All-GRNN_r$ and $PCA-GRNN_r$ algorithms.

The highest ICC was achieved with $\tau = 0.5$ and $\tau = 0.4$ for the algorithms $All-GRNN_r$ and $PCA-GRNN_r$, respectively. The optimum configurations for both algorithms were given by a small kernel width. Thus, the estimated AHI will be heavily

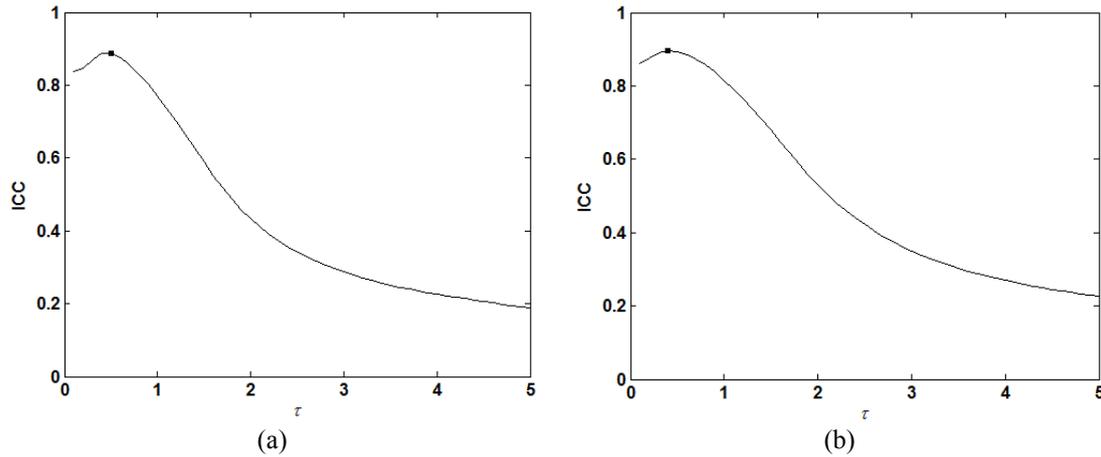


Figure 28. Influence of the spread parameter (controlled by τ) on the *ICC* for (a) the *All-GRNN_r* and (b) the *PCA-GRNN_r* regression algorithms.

weighted by the target value associated to the kernel centre that is closest to the input pattern.

8.7.1.3. Design of MLP algorithms for regression

The design of MLP networks for regression analysis was similar to that described for the classification approach. Decisions involving the optimisation algorithm, its maximum number of iterations or weight initialisation were the same adopted for MLP network classifiers. In addition, a single output neuron was used to implement MLP regression algorithms. However, its activation function must be linear since the target variable approximated by the network is continuous.

A single hidden layer of nodes was used to implement MLP regression algorithms. As previously mentioned, MLP networks with this architecture and a sufficient number of hidden nodes are capable of universal approximation (Hornik, 1991). Therefore, model selection was performed by evaluating different combinations of the regularisation parameter (ν) and the number of nodes in the hidden layer (N_H). Both influence model complexity and generalisation ability as explained for MLP classifiers.

The performance of multiple network configurations was compared. Parameter ν was varied from 0 to 100 while N_H ranged from 2 to 50. The intervals

defined for both design parameters were chosen in order to observe the trend of performance with respect to them. The value of ICC was computed from leave-one-out cross-validation from training data. This process was repeated 10 times and the average ICC was derived. The performance curves for $All-MLP_r$ and $PCA-MLP_r$ regression algorithms are depicted in Figure 29.

Initially, the effect of parameter ν was analysed. This parameter controls network complexity and it must be chosen in order to find the best balance between bias and variance. The optimum value of ν was set to 50 and 15 for the $All-MLP_r$ and $PCA-MLP_r$ regression algorithms, respectively. Networks with ν smaller than these values were affected by overfitting due to excessive model complexity. Larger ν led to the opposite situation and the ICC decreased as the flexibility of the networks was reduced.

The performance curves obtained for the selected ν were similar to those observed for MLP classifiers. Initially, ICC gradually increased as more hidden nodes were added, showing that too much simple network architectures cannot provide a proper representation of the problem. However, there was no substantial improvement beyond a given value of N_H . This point was selected as optimum for both MLP regression models. It was set to $N_H = 40$ for the $All-MLP_r$ algorithm and $N_H = 22$ for the $PCA-MLP_r$ algorithm. Similar performance was achieved by networks with higher N_H since overfitting was prevented by weight decay regularisation.

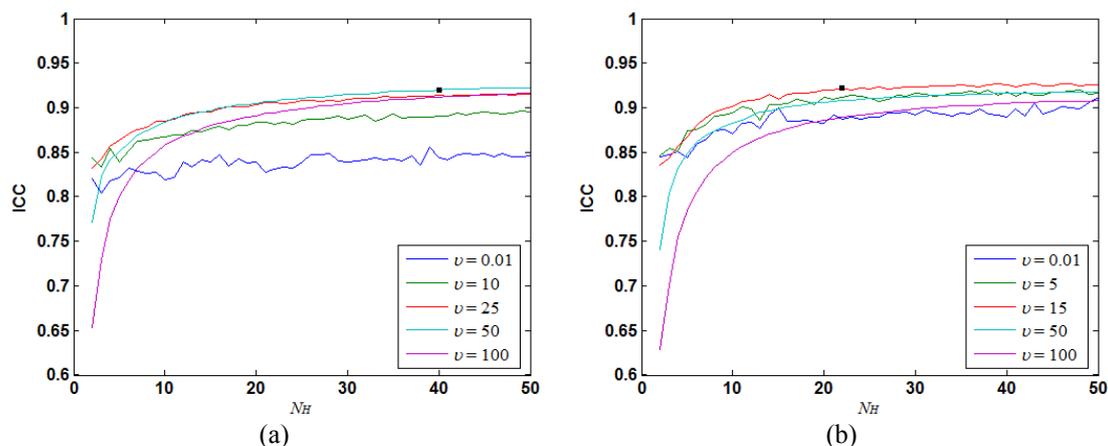


Figure 29. Influence of the number of hidden nodes (N_H) and the regularisation parameter (ν) on the ICC for (a) the $All-MLP_r$ and (b) the $PCA-MLP_r$ regression algorithms.

Figure 30 enables to compare the magnitude of the weights for different configurations of the $All-MLP_r$ regression algorithm. Given the value of ν , regularisation causes weights to be smaller as N_H is larger. As a result, the effective complexity of the model is not increased and overfitting effects are avoided. If N_H is fixed, a larger ν provided networks with a higher percentage of weight close to zero. An excessive reduction of model complexity could be produced by setting ν too large, resulting in underfitting. An analogous exercise is shown in Figure 31 using several configurations of the $PCA-MLP_r$ regression algorithm. The analysis provided similar conclusions to those drawn from MLP classifiers, i.e. even networks with large N_H can

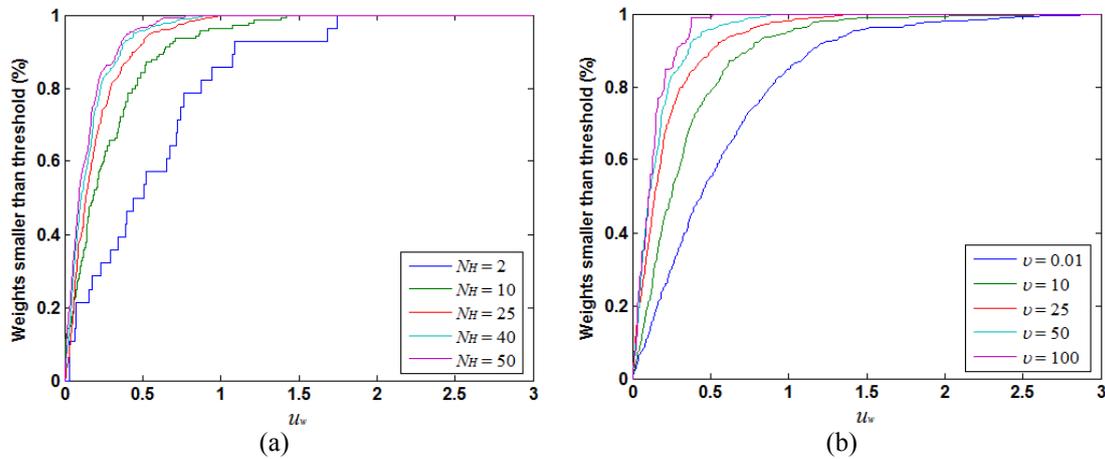


Figure 30. Effect of weight decay regularisation on the size of the weights: (a) $All-MLP_r$ regression algorithms with $\nu = 50$ and different N_H ; (b) $All-MLP_r$ regression algorithms with $N_H = 40$ and different ν .

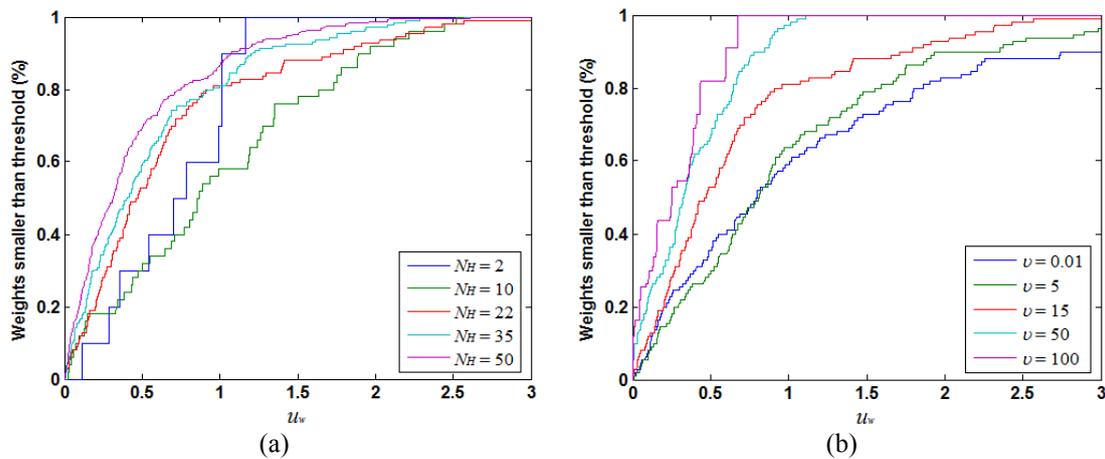


Figure 31. Effect of weight decay regularisation on the size of the weights: (a) $PCA-MLP_r$ regression algorithms with $\nu = 15$ and different N_H ; (b) $PCA-MLP_r$ regression algorithms with $N_H = 22$ and different ν .

reach high generalisation ability due to weight decay regularisation. Network complexity was not increased by adding more hidden neurons as the magnitude of the weights became smaller.

8.7.1.4. Design of RBF algorithms for regression

RBF networks for regression analysis were implemented with Gaussian kernels. Their centres were determined by means of the k -means algorithm. A single output neuron characterised by a linear activation function was used. Output layer weights were adjusted by minimising the sum-of-squares error function. Therefore, the size of the hidden layer (N_B) and the variance of each Gaussian kernel (σ_{rj}) were the parameters to be optimised through model selection. The latter was specified by using its corresponding spread (Sp_j) value. For each kernel, Sp_j was set to a quantity proportional to the mean distance of the kernel centre to its K nearest centres ($K = 3$). The tuning parameter τ was used to determine the value of Sp_j as expressed in Eq. (208). In order to find the optimum configuration for *All-RBF_r* and *PCA-RBF_r* regression algorithms, τ was varied from 0.1 to 5. In addition, N_H ranged between 2 and 50. RBF networks with a different degree of complexity were evaluated to detect the influence of both parameters on generalisation capability. The performance curves obtained for both algorithms are shown in Figure 32.

As expected, the performance was poor when kernel widths were set too small. *ICC* was improved by increasing the degree of overlapping among kernels. However, setting kernel widths to excessively large values resulted in decreased performance due to oversmoothed mappings. The balance between both situations was found for $\tau = 1.5$ and $\tau = 1.8$ for *All-RBF_r* and *PCA-RBF_r* algorithms, respectively. The performance curves corresponding to these τ were analysed to determine the optimum N_B . Their profile reveals a progressive increment of *ICC* towards a global maximum. Subsequently, performance slowly decreased as more basis functions were added. Such maximum point was given by $N_B = 16$ and $N_B = 12$ for *All-RBF_r* and *PCA-RBF_r* algorithms, respectively.

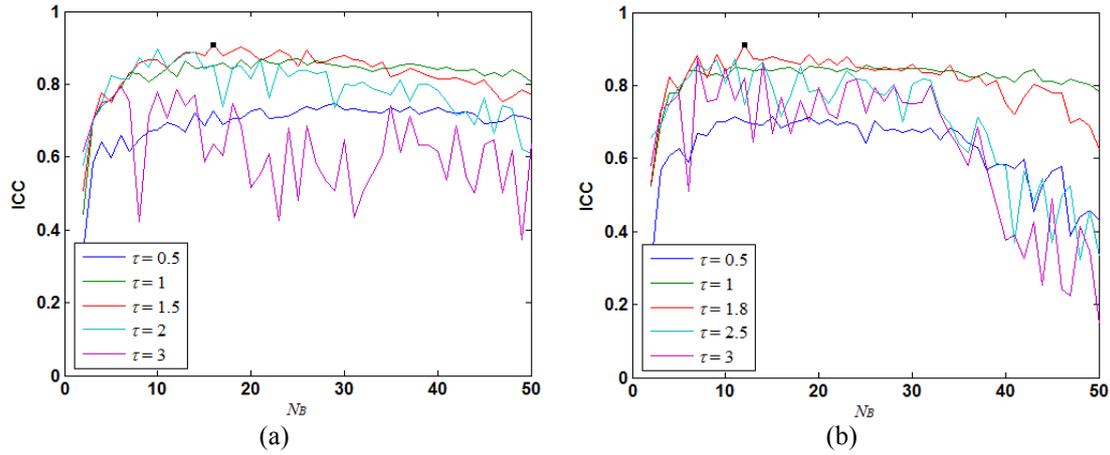


Figure 32. Influence of the number of basis functions (N_B) and the kernel width parameter (controlled by τ) on the ICC for (a) the *All-RBF_r*, and (b) the *PCA-RBF_r*, regression algorithms.

8.7.1.5. Design of BY network algorithms for regression

BY networks for regression analysis had the same architecture as MLP regression algorithms. The Gaussian approximation to the posterior distribution of network weights and the evidence framework were used for implementing BY networks. ARD was used to evaluate the relative importance of each input feature in the regression algorithm. As described for BY classifiers, network weights were split into different groups and a hyperparameter α_k was assigned to each of them. An additional hyperparameter β must be considered for regression modelling. The distribution of target values is modelled by a Gaussian function in regression problems. The variance of this distribution is given by the inverse of β . Hyperparameters α_k and β were periodically updated during training as stated by the evidence framework. Initially, all α_k were set to 0.1 to reflect the lack of prior knowledge about the network weights. However, the prior distribution $p(\mathbf{w})$ defined by the initial value of α_k was not used for weight initialisation as explained for BY classifiers. On the other hand, hyperparameter β was set to 50, i.e a small variance was used to model the distribution of target values.

The optimum number of neurons in the hidden layer (N_H) must be defined. Networks ranging from 2 to 50 hidden neurons were evaluated for analysing the effect of this parameter on generalisation ability. This range is sufficient to cover simple and highly complex network architectures. For each N_H , leave-one-out cross-validation was repeated 10 times in order to obtain a reliable estimation of ICC . The average

performance curves for $All-BY_r$ and $PCA-BY_r$ regression algorithms are shown in Figure 33.

Common properties can be appreciated in the profiles of both performance curves. High ICC values were provided by networks with N_H lower than a certain limit. In the case of $All-BY_r$ algorithms, the ICC is close to zero for N_H higher than 8. The same occurs for $PCA-BY_r$ models with a network size greater than 21 hidden neurons. These limits correspond to r_w equal to 0.74 samples per weight for $All-BY_r$ and 0.65 samples per weight for $PCA-BY_r$ networks. As observed for BY classifiers, the Bayesian learning provided valid models for the proposed problem even for low values of r_w . However, the Gaussian approximation is not valid for excessively complex networks with a large number of adaptive parameters. The selected N_H for $All-BY_r$ and $PCA-BY_r$ regression algorithms were 5 and 8, respectively. These configurations provided the highest ICC values.

The complete training set was used to train the definitive algorithms using the selected configurations. The hyperparameters α_k associated to the input features were evaluated to perform ARD analysis. The values of the hyperparameters after training the $All-BY_r$ algorithms are depicted in Figure 34. The results show that the spectral features S_B and PA were the most relevant ones in the mapping function defined by the network. In addition, the variance of SaO_2 samples (σ_i) and the expectation of the variable representing the frequency component (μ_f) have significant influence on the estimated AHI.

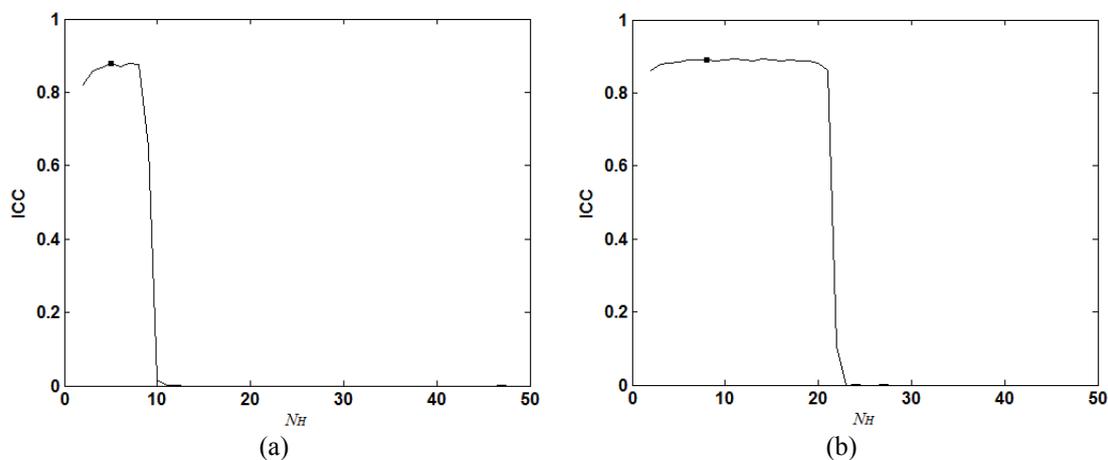


Figure 33. Influence of the number of hidden nodes (N_H) on the ICC for (a) the $All-BY_r$ and (b) the $PCA-BY_r$ regression algorithms.

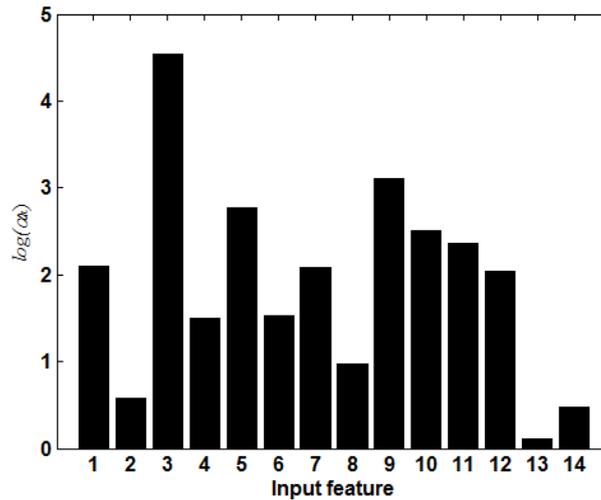


Figure 34. Results from ARD for the *All-BY_r* regression algorithm. The logarithm (base 10) of α_k was represented.

8.7.1.6. Design of MARS algorithms

Several design parameters must be adjusted to design MARS algorithms. Friedman (1991) found that the best value for the smoothing parameter p is in the range between 2 and 4, with an optimum value equal to 2 in the case of additive models ($mi = 1$). These findings were considered for determining the value of p , which was set to 3 for all models allowing interactions between different variables and to 2 for additive models. Two additional parameters must be fixed: the allowable level of interaction between variables (mi) and the maximum number of basis functions (M_{max}). The former is limited by the number of input features. Low values of mi restrict the flexibility of the MARS model. Setting $mi = 1$ forces the algorithm to provide an approximation based on the sum of univariate functions. A completely unconstrained model can be implemented by defining $mi = d$. Also, the influence of M_{max} on generalisation capability was analysed. According to Friedman (1991), a suitable strategy is to let the forward phase of the algorithm define a large number of basis functions and then reduce the model to its proper size during the backward phase. Thus, M_{max} was varied from its minimum possible value ($M_{max} = 1$) up to a sufficiently large number of basis functions ($M_{max} = 100$). The results obtained for *All-MARS_r* and *PCA-MARS_r* algorithms for different configurations of these parameters are depicted in Figure 35.

Smooth mappings characterised by small mi provided better generalisation performance. Increasing model complexity by choosing high mi led to inaccurate algorithms. The additive model ($mi = 1$) achieved the highest performance for the *All-MARS_r* algorithm. On the other hand, some benefit is obtained by allowing interaction effects in *PCA-MARS_r* algorithms. The configurations with $mi = 2$ achieved the highest *ICC*. Larger mi resulted in decreased performance due to excessive variance. On the other hand, the profiles of the curves show the irregular dependence of regression performance on parameter M_{max} . The higher this parameter is, the larger the number of basis function candidates. This parameter should be large enough to properly fit training data. The configurations of the *All-MARS_r* and *PCA-MARS_r* algorithms that achieved the highest *ICC* were selected as optimum. These corresponded to $M_{max} = 38$ and $M_{max} = 78$, respectively.

The regression function implemented by a MARS model can be expressed by means of its ANOVA decomposition. It provides useful information about the predictive relationship between the response variable (i.e. the estimated AHI) and the input features. The idea is to collect together all basis functions that involve identical feature sets (Friedman, 1991). The model can be expressed as:

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{K_m=1} f_i(x_i, w_i) + \sum_{K_m=2} f_{ij}(x_i, x_j, w_{ij}) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k, w_{ijk}) + \dots \quad (209)$$

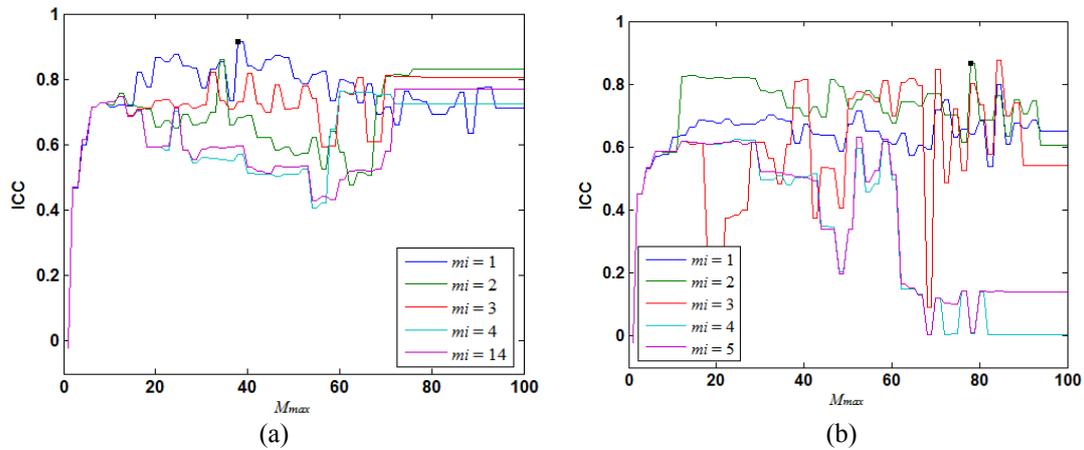


Figure 35. Influence of the maximum number of basis functions (M_{max}) and the allowable level of interaction between variables (mi) on the *ICC* for (a) the *All-MARS_r* and (b) the *PCA-MARS_r* regression algorithms.

Each function f_i in the first term is given by the sum of the basis functions, weighted by their corresponding coefficients w_i , involving the variable x_i . Similarly, each function f_{ij} associated to coefficient w_{ij} is given by the sum of all the basis functions involving variables x_i and x_j . The ANOVA decomposition was used to analyse the *All-MARS_r* algorithm involving the complete set of 14 normalised input features. The entire training set was used to build the algorithm using the configuration selected ($m_i = 1$ and $M_{max} = 38$) in the design phase. Table 16 summarises the approximating function implemented by the algorithm. A total of 12 basis functions were defined, yielding 9 different univariate ANOVA functions. From this set, ANOVA functions 2, 6 and 7 were composed of the sum of 2 basis functions while the others involved a single basis function. The relative importance of each function to the overall model can be evaluated by comparing its standard deviation (σ_{af}), with higher standard deviation being interpreted as more relevant (Friedman, 1991). Using this rule, the analysis of ANOVA functions reveals that functions related to S_T , S_B , LZC and μ_f have greater influence on the estimated AHI.

8.7.1.7. Design of SVM algorithms for regression

Regression algorithms based on SVM requires the designer to specify the kernel function to be used, the width of the ε -insensitive zone and the regularisation parameter (C). As previously indicated, SVM with Gaussian kernels (RBF architecture) were considered for regression analysis. This kernel function adds an additional design parameter represented by the width of the Gaussians (σ_s). In the case of SVM models, it

ANOVA function	# of basis	Features	σ_{af}
1	1	S_B	7.20
2	2	μ_f	6.48
3	1	δ_t	4.99
4	1	LZC	6.96
5	1	$ApEn$	4.70
6	2	S_T	9.00
7	2	σ_f	3.65
8	1	CTM	0.73
9	1	σ_t	0.57

Table 16. ANOVA function decomposition for the *All-MARS_r* algorithm. The standard deviation of each function reflects its influence on the output of the algorithm.

is the same for all the kernels in the algorithm. Parameter ε was not considered for optimisation and was fixed to 1, i.e. a maximum error of 1 unit was allowed for the estimated AHI. It represents a reasonable error for the scale of AHI values, which approximately ranges from 0 to 150 h⁻¹. Therefore, model selection was focused on the analysis of parameters C and σ_s .

The influence of both parameters on the behaviour of SVM regression algorithms is similar to that described previously for the classification approach. Increasing the value of C encourages the algorithm to reduce the error on the training set rather than minimising the norm of the vector \mathbf{w} . As a result, model complexity increases and a lower number of training patterns with $\xi_i > 0$ can be observed. In addition, the number of support vectors becomes smaller since less training patterns lie outside the ε -insensitive tube. If C is set too large, the algorithm may be extremely adapted to training data, suffering from overfitting. The width of the Gaussian kernels also affects the profile of the approximating function implemented by the SVM algorithm. As σ_s is set larger, the mapping defined by the algorithm becomes smoother. Underfitting may arise if σ_s is too large. In the limit of σ_s close to infinity, there is no difference in the output obtained for different input patterns. Conversely, setting σ_s too small may lead to select all the training patterns as support vectors, indicating an excessive adaptation of the algorithm to the training data.

Initially, there is no knowledge about the optimum values of C and σ_s . Therefore, both parameters were varied in a wide interval of possible values ranging from 10^{-4} to 10^8 (Hsu and Lin 2002). The *ICC* curves obtained for *All-SVM_r* and *PCA-SVM_r* algorithms are shown in Figure 36.

Several combinations of C and σ_s provided high *ICC* values for *All-SVM_r* and *PCA-SVM_r* algorithms. As observed in the classification approach, the results show that C and σ_s complement themselves in order to achieve a degree of complexity close to the optimum. For a range of σ_s values between 10^1 and 10^2 , increasing this parameter can be compensated by selecting a larger value of C . The configuration with the highest *ICC* for the *All-SVM_r* model was that with $C = 10^5$ and $\sigma_s = 10^2$. The parameters

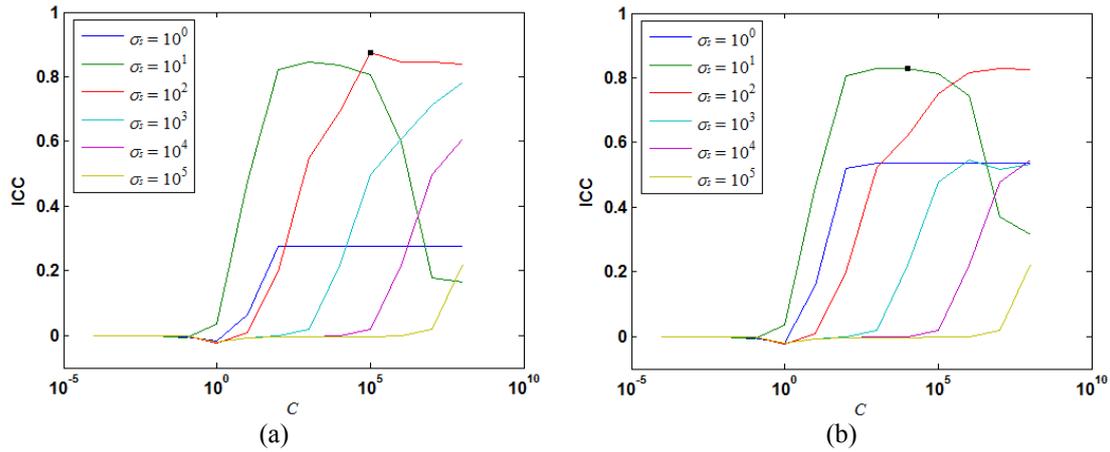


Figure 36. Influence of the regularisation parameter (C) and the kernel width parameter (σ_s) on the ICC for (a) the $All-SVM_r$ and (b) the $PCA-SVM_r$ regression algorithms.

selected for the $PCA-SVM_r$ algorithm were $C = 10^4$ and $\sigma_s = 10$ since this configuration achieved the highest performance in the leave-one-out evaluation.

8.7.2. Testing regression algorithms

The entire training set was used to implement the regression algorithms using the configuration selected in the design phase. The number of algorithms developed for regression analysis was 14 (7 pattern recognition methods were assessed, each of them with and without PCA preprocessing). They were evaluated on data in the test set to estimate their generalisation ability. The obtained results are summarised in Table 17.

In general, the evaluated methods provided satisfactory performance for the estimation of the AHI. Four of the algorithms implemented in the study ($All-MLP_r$, $PCA-MLP_r$, $PCA-BY_r$ and $All-MARS_r$) achieved $ICC \geq 0.9$. They also provided the smallest average error (measured by E_{RMS}) in the estimation. Bland-Altman analysis was carried out to assess the performance of these algorithms in detail. The difference between actual AHI and predicted AHI (d_{AHI}) for each subject is depicted as a function of the mean value of both quantities (μ_{AHI}). The mean (μ_d) and the standard deviation (σ_d) of the differences provide useful information about the quality of the approximation. Two parallel lines placed at $\mu_d + 2\sigma_d$ and $\mu_d - 2\sigma_d$ are plotted to

Algorithm	Tn	Fp	Tp	Fn	Se (%)	Sp (%)	Acc (%)	ICC	E_{RMS}	E_{rel} (-/+)
<i>All-MLR_r</i>	37	11	86	10	89.58	77.08	85.42	0.80	19.83	(2.70/0.43)
<i>PCA-MLR_r</i>	29	19	90	6	93.75	60.42	82.64	0.87	12.79	(3.65/0.47)
<i>All-GRNN_r</i>	32	16	92	4	95.83	66.67	86.11	0.89	11.94	(4.33/0.34)
<i>PCA-GRNN_r</i>	32	16	90	6	93.75	66.67	84.72	0.89	12.29	(3.35/0.37)
<i>All-MLP_r</i>	39	9	86	10	89.58	81.25	86.81	0.91	11.10	(1.98/0.34)
<i>PCA-MLP_r</i>	40	8	83	13	86.46	83.33	85.42	0.90	11.70	(2.88/0.35)
<i>All-RBF_r</i>	35	13	84	12	87.50	72.92	82.64	0.89	12.64	(4.03/0.39)
<i>PCA-RBF_r</i>	36	12	85	11	88.54	75.00	84.03	0.89	13.00	(3.17/0.40)
<i>All-BY_r</i>	34	14	86	10	89.58	70.83	83.33	0.88	12.23	(3.00/0.37)
<i>PCA-BY_r</i>	41	7	85	11	88.54	85.42	87.50	0.90	11.38	(1.61/0.36)
<i>All-MARS_r</i>	41	7	85	11	88.54	85.42	87.50	0.91	11.90	(1.64/0.33)
<i>PCA-MARS_r</i>	34	14	86	10	89.58	70.83	83.33	0.72	22.72	(2.52/0.39)
<i>All-SVM_r</i>	37	11	86	10	89.58	77.08	85.42	0.88	13.72	(2.80/0.39)
<i>PCA-SVM_r</i>	41	7	84	12	87.50	85.42	86.81	0.86	13.23	(2.10/0.35)

Table 17. Results achieved on the test set by the 14 regression algorithms implemented in the study. Tn : true negatives; Fp : false positives; Tp : true positives; Fn : false negatives; Se : sensitivity; Sp : specificity; Acc : accuracy; ICC : intra-class correlation coefficient; E_{RMS} : root mean square error; E_{rel} (-/+): relative error per subject (SAHS-negative group/SAHS-positive group).

represent the 95% limits of agreement, which reflect the nature of the differences.

Figure 37 depicts Bland-Altman and regression plots for these four algorithms.

There were no substantial differences between the four plots. Bland-Altman representation shows that the average difference (bias) between both scores of AHI is close to zero. The smallest bias among these algorithms was achieved by *PCA-MLP_r*, while *All-MLP_r* provided the highest. Although bias was negative for all of them, its reduced magnitude indicates that there was not any marked trend reflecting overestimation or underestimation effects. The standard deviation of the difference for the four algorithms is around 11 as reflected by the limits of agreement. It should be noted that the plots show a trend in the variability of the difference. The scatter gets larger as the average gets higher, i.e. the error becomes larger for higher values of AHI. On the other hand, regression plots show the deviation of the estimated values from their corresponding target, which is given by the identity function (dotted line). It can be observed that the four algorithms tend to provide small error in the predictions. However, the distance from the target is significant for some subjects with high AHI.

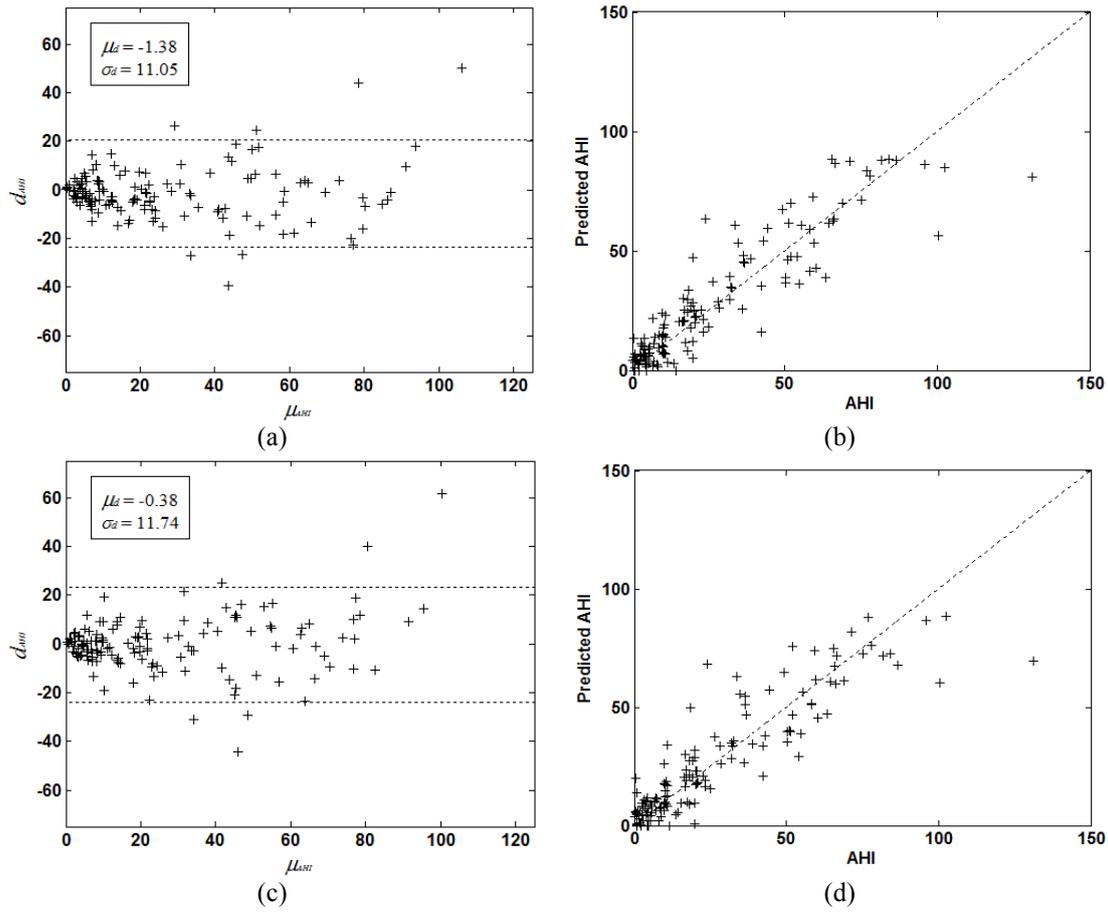


Figure 37. Bland-Altman and regression plots for the $All-MLP_r$ ((a) and (b)), the $PCA-MLP_r$ ((c) and (d)), the $PCA-BY_r$ ((e) and (f)) and the $All-MARS_r$ regression algorithms.

In addition to ICC , error measures such as E_{RMS} and E_{rel} are useful indicators of approximation accuracy. E_{rel} is higher for SAHS-negative subjects as they have smaller AHI and peak values of this measure can be found for subjects with AHI close to 0 h^{-1} . The mean E_{rel} of $All-MLP_r$, $PCA-MLP_r$, $PCA-BY_r$ and $All-MARS_r$ algorithms on the SAHS-negative group in the test set was 1.98, 2.88, 1.61 and 1.64, respectively. For SAHS-positive subjects, the E_{rel} obtained by these algorithms was 0.34, 0.35, 0.36 and 0.33. $PCA-BY_r$ was the most accurate algorithm for predicting AHI in SAHS-negative subjects, while $All-MARS_r$ achieved the smallest E_{rel} in the SAHS-positive group. Therefore, both algorithms can be pointed out as the best regression models among those implemented in the study. Indeed, if the estimated AHI is used to classify subjects as SAHS-negative or SAHS-positive, both algorithms achieved the highest correct classification rate in the test set, with an accuracy of 87.50% (88.54% sensitivity and 85.42% specificity). Both misclassified 7 SAHS-negative and 11 SAHS-positive subjects. Twelve of them (4 SAHS-negative and 8 SAHS-positive subjects) were

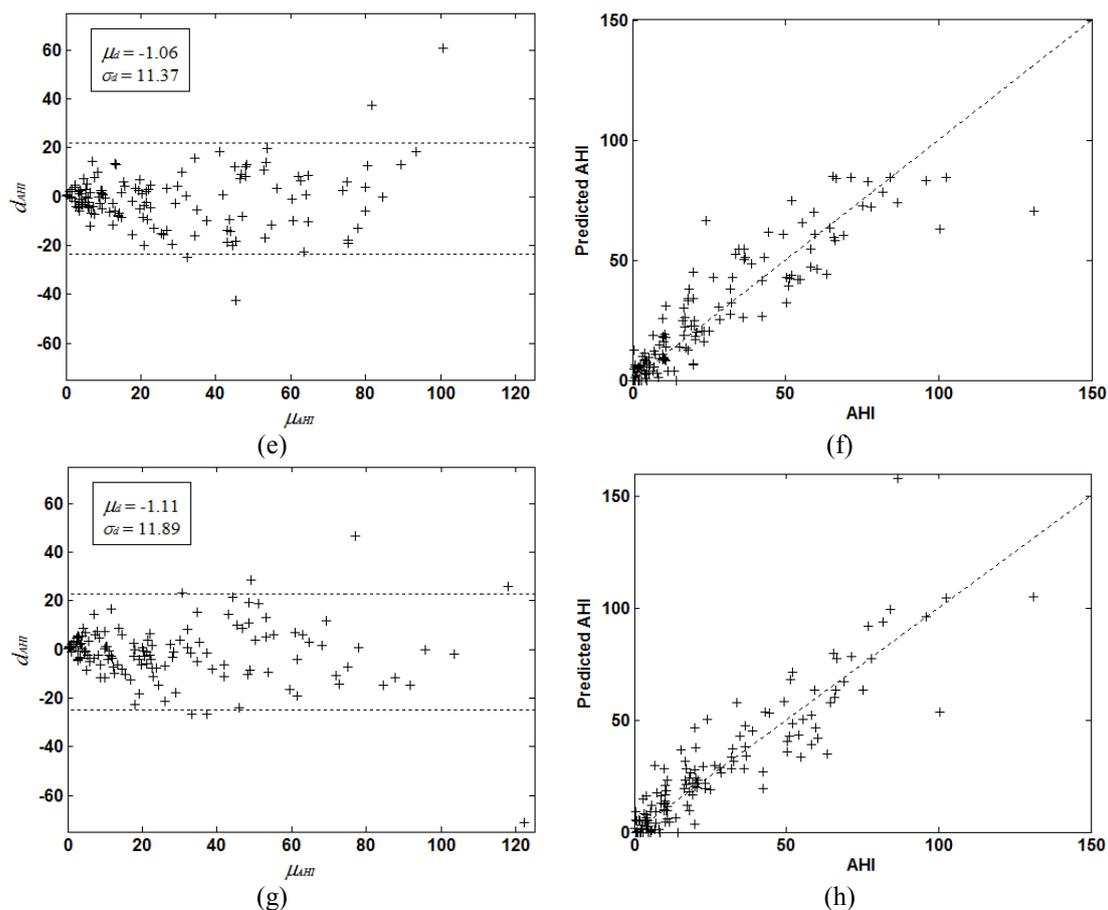


Figure 37 (Cont.). Bland-Altman and regression plots for the *All-MLP_r*, ((a) and (b)), the *PCA-MLP_r*, ((c) and (d)), the *PCA-BY_r*, ((e) and (f)) and the *All-MARS_r*, regression algorithms.

common to both algorithms. The *PCA-BY_r* algorithm provided an incorrect diagnosis for 13 subjects with mild SAHS (4 SAHS-negative subjects with $\text{AHI} \geq 5 \text{ h}^{-1}$ and 9 SAHS-negative subjects with $\text{AHI} \leq 15 \text{ h}^{-1}$). The *All-MARS_r* algorithm gave a wrong decision for 14 subjects with mild SAHS (5 SAHS-negative subjects with $\text{AHI} \geq 5 \text{ h}^{-1}$ and 9 SAHS-negative subjects with $\text{AHI} \leq 15 \text{ h}^{-1}$). Some of them had an AHI close to the threshold of 10 h^{-1} . Therefore, a small deviation in the estimated AHI led to an incorrect diagnosis. Table 18 summarises the actual and estimated AHI for the subjects misclassified by *PCA-BY_r* and *All-MARS_r* algorithms.

Finally, the influence of PCA preprocessing can be analysed. It had different effects on the evaluated methods. Dimensionality reduction represented a benefit for models based on MLR (*ICC* increased from 0.80 to 0.87) and BY networks, whereas the performance of other methods such as GRNN, MLP and RBF was not clearly influenced by PCA. In addition, it adversely influenced regression algorithms based on

<i>PCA-BY_r</i>			<i>All-MARS_r</i>		
Subject	AHI	y	Subject	AHI	y
17	7	12.0	19	8.7	16.0
19	8.7	14.9	29	7.6	17.4
24	3.3	10.1	30	6.6	29.5
29	7.6	10.6	33	5.6	12.2
30	6.6	18.5	37	8.7	13.0
42	4.1	11.3	42	4.1	16.0
48	0.6	12.6	46	2.9	14.5
51	10.2	9.0	54	14.3	0.0
54	14.3	0.0	55	10.6	9.6
55	10.6	8.3	77	10.8	6.1
58	19.9	6.5	105	19.8	3.3
77	10.8	8.2	125	10.0	9.6
105	19.8	6.9	128	10.3	4.4
113	10.2	9.5	129	11.0	9.8
128	10.3	9.2	130	18.2	9.4
132	11.5	3.8	132	11.5	4.3
141	10.0	8.5	141	10.0	9.9
143	13.5	3.7	143	13.5	6.4

Table 18. Data of the subjects in the test set misdiagnosed by *PCA-BY_r* and *All-MARS_r* algorithms. *PCA-BY_r*: regression algorithm based on BY networks and components retained from PCA; *All-MARS_r*: regression algorithm based on MARS and all the features from oximetry data; AHI: apnoea-hypopnoea index.

SVM and MARS. It was specifically remarkable for the latter since *ICC* decreased from 0.91 to 0.72 when PCA preprocessing was used.

8.8. Conventional analysis of oximetry signals

The diagnostic utility of conventional oximetry indices was evaluated using the database of SaO₂ signals available for this study. The following indices were computed off-line (Table 19 summarises the mean value of these indices for signals in the training set):

- The minimum value of saturation during sleep (minSaO₂). It represents the lowest value of saturation observed during the complete night.
- The fraction of time with a saturation level lower than 90% (CT90). It accounts for the time during which the SaO₂ value is smaller than 90%. The total duration of the signal is used as a normalisation factor.

TRAINING SET			
	All	SAHS-positive	SAHS-negative
minSaO ₂	82.80 ± 8.74	80.25 ± 9.37	87.90 ± 3.95
CT90	9.13 ± 19.08	10.09 ± 17.36	7.22 ± 22.30
ODI3	12.16 ± 19.45	17.62 ± 21.88	1.24 ± 1.14
ODI4	8.52 ± 16.95	12.52 ± 19.60	0.51 ± 0.62
SIT90	8450.43 ± 26310.47	10977.77 ± 31019.56	3395.77 ± 11240.84
Δ index	0.10 ± 0.12	0.12 ± 0.14	0.04 ± 0.01

Table 19. Mean values of the conventional oximetry indices computed from oximetry recordings in the training set.

- The oxygen desaturation index at 3% (ODI3) and 4% (ODI4). They reflect the ratio of desaturation events greater than 3% or 4% regarding the baseline level during the night, respectively. The definition of desaturation given by Magalang *et al.* (2003) was used to compute these indices: a decrease in SaO₂ greater than or equal to the set amount (3% or 4%) from baseline for at least 10 s and at a rate > 0.1%/s, returning within 60 s to a 1% interval of the initial value. Baseline was set as the mean saturation in the previous minute (Chiner *et al.* 1999). The mean level in the first 3 min of recording was used to initialise its value.
- The saturation impairment time using a threshold of 90% (SIT90). It was computed as the area of the SaO₂ signal under the level of 90% saturation (Chesson *et al.* 1993).
- Δ index. It was computed as the sum of the absolute difference between two successive points divided by the number of intervals (number of points less one) (Pepin *et al.* 1991). The Δ index was estimated as the average of the values computed for signal epochs with a length of 512 samples

The utility of each oximetry index in the proposed classification problem was estimated from ROC analysis of training data. The threshold that provided the highest classification accuracy on the training set was selected as optimum. The value of

the optimum threshold for each of the indices is shown in Table 20. Signals in the test set were classified using this optimum threshold to evaluate their diagnostic accuracy. Classification results provided by conventional oximetry indices are summarised in Table 21.

On the other hand, a linear regression model with the form of Eq. (205) was implemented for each index to analyse their utility for estimating the AHI. The *ICC* between real and estimated values of the AHI was used to measure the quality of the approximation. In addition, the diagnostic utility of the estimated AHI was evaluated. Table 22 provides the coefficients of the linear model built from each index. Their results on the test set are summarised in Table 23.

The highest classification performance was achieved by the Δ index, with an accuracy of 88.19% (86.46% sensitivity and 91.67% specificity) and an *AUC* of 0.960. It provided an incorrect decision for 4 SAHS-negative and 13 SAHS-positive subjects. On the other hand, the highest level of agreement between real and predicted AHI was

Index	Threshold
minSaO ₂	88.90
CT90	0.07
ODI3	2.72
ODI4	0.72
SIT90	11.90
Δ index	0.05

Table 20. Optimum decision threshold determined from the training set for each of the oximetry indices.

Index	<i>Tn</i>	<i>Fp</i>	<i>Tp</i>	<i>Fn</i>	<i>Se</i> (%)	<i>Sp</i> (%)	<i>Acc</i> (%)	<i>AUC</i>	κ
minSaO ₂	25	23	89	7	92.71	52.08	79.17	0.860	0.49
CT90	29	19	87	9	90.63	60.42	80.56	0.881	0.54
ODI3	42	6	83	13	86.46	87.50	86.81	0.954	0.71
ODI4	38	10	87	9	90.63	79.17	86.81	0.952	0.70
SIT90	25	23	89	7	92.71	52.08	79.17	0.873	0.49
Δ index	44	4	83	13	86.46	91.67	88.19	0.960	0.75

Table 21. Classification results individually achieved by each of the oximetry indices. *Tn*: true negatives; *Fp*: false positives; *Tp*: true positives; *Fn*: false negatives; *Se*: sensitivity; *Sp*: specificity; *Acc*: accuracy; *AUC*: area under the ROC curve; κ : Cohen's kappa coefficient.

Index	w_0	w_1
minSaO ₂	174.4800	-1.8083
CT90	21.2800	0.3799
ODI3	10.9540	1.1345
ODI4	14.3600	1.2198
SIT90	21.2360	0.0004
Δ index	9.3361	159.5600

Table 22. Coefficients of the linear regression models developed for each of the oximetry indices.

Index	Tn	Fp	Tp	Fn	Se (%)	Sp (%)	Acc (%)	ICC	E_{RMS}	E_{rel} (-/+)
minSaO ₂	14	34	92	4	95.83	29.17	73.61	0.66	21.73	(5.91/0.61)
CT90	0	48	96	0	100.00	0.00	66.67	0.31	23.17	(11.80/0.49)
ODI3	0	48	96	0	100.00	0.00	66.67	0.91	10.78	(5.92/0.25)
ODI4	0	48	96	0	100.00	0.00	66.67	0.84	14.14	(7.77/0.31)
SIT90	0	48	96	0	100.00	0.00	66.67	0.47	26.94	(11.76/0.52)
Δ index	0	48	96	0	100.00	0.00	66.67	0.77	16.79	(7.41/0.33)

Table 23. Regression results for each of the oximetry indices. Tn : true negatives; Fp : false positives; Tp : true positives; Fn : false negatives; Se : sensitivity; Sp : specificity; Acc : accuracy; ICC : intra-class correlation coefficient; E_{RMS} : root mean square error; E_{rel} (-/+): relative error per subject (SAHS-negative group/SAHS-positive group).

achieved by the linear regression algorithm built from ODI3, which provided an ICC of 0.91. However, the estimated AHI provided an accuracy of 66.67% (100.00% sensitivity and 0.00% specificity) for SAHS detection, i.e. the AHI predicted for all the SAHS-negative subjects in the test set was higher than 10 h^{-1} . The regression algorithm based on ODI3 overestimates the AHI for small values of this variable. However, it shows high accuracy in the estimation of large AHI values. A similar behaviour was observed for ODI4. This effect is reflected by the values of E_{rel} in both populations. Bland-Altman and regression plots for ODI3 are depicted in Figure 38. They show the described overestimation effect.

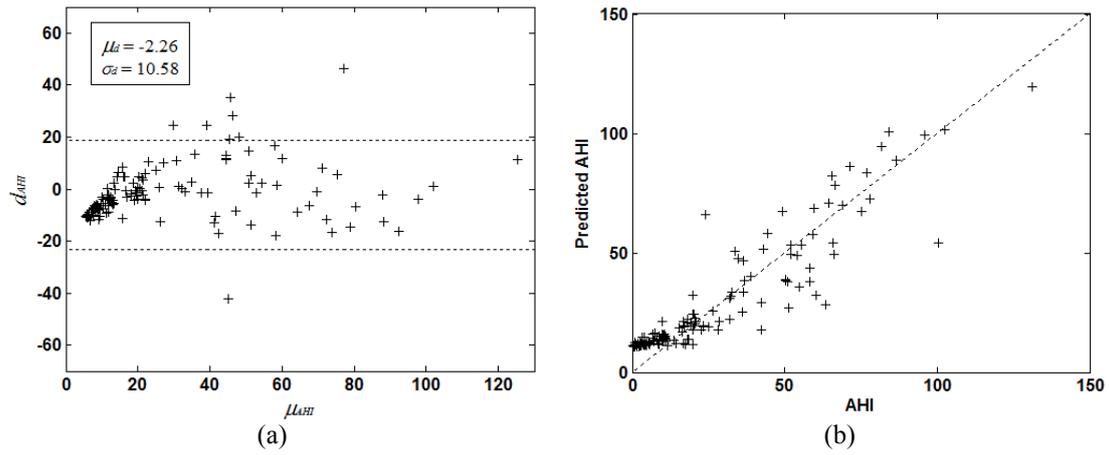


Figure 38. Bland-Altman (a) and regression (b) plots for the linear model built from ODI3.

9. Discussion and conclusions

- 9.1. Introduction
- 9.2. Discussion
- 9.3. Conclusions
- 9.4. Original contributions
- 9.5. Future work

9.1. Introduction

This Thesis addresses the development of automatic techniques for SAHS diagnosis using information from SaO_2 recordings. Clinicians and scientists have paid special attention to the study of this condition in the last decades. The high percentage of undiagnosed patients and the severity of long term effects, including cardiovascular and cerebrovascular diseases, make SAHS a public health concern. The only reliable method for SAHS diagnosis is nocturnal PSG. It is performed in a special sleep unit for monitoring multiple physiological recordings and data under supervision of qualified personnel. PSG presents some drawbacks due to its complexity and high cost. Additionally, the demand for PSG studies is progressively growing as people and clinicians are becoming aware of SAHS, whereas the available infrastructure is insufficient to support it.

Early detection of SAHS is required in order to prevent other health complications. Thus, alternative diagnostic procedures are desirable to overcome the difficulties of PSG. Different methods have been recently proposed. The common strategy is to provide a diagnosis about SAHS from a small set of recordings or clinical data. Usually, these can be obtained from clinical observation, individual forms or using portable monitoring devices at patient's home. As a result, cost and complexity associated to PSG would be reduced. In addition, the demand for in-laboratory sleep studies could be satisfied with the available resources. This is the purpose for the diagnostic algorithms developed in this Thesis. They are based on the analysis of SaO_2 recordings using pattern recognition techniques to automatically provide a decision about SAHS. Similarly, previous researchers have studied the utility of these signals for SAHS detection. In this chapter, the effectiveness of the methods proposed in the Thesis is analysed. Moreover, the achieved results are compared with those of other diagnostic techniques previously proposed. The main conclusions of the study are presented as well as possible future work to continue this research.

9.2. Discussion

Pattern recognition techniques have been used to perform SAHS diagnosis from SaO_2 data recorded through nocturnal pulse oximetry. A model composed of three

different stages was suggested to analyse SaO₂ signals: feature extraction, feature normalisation and dimensionality reduction, and pattern analysis. Initially, information in oximetry recordings was summarised into a feature pattern. The extracted features were chosen in order to reflect the presence of apnoea events in oximetry recordings. A linear re-scaling procedure was subsequently applied to each of the features to be simultaneously processed. In addition, the use of dimensionality reduction techniques by means of PCA was assessed. Two different diagnostic models were proposed for pattern analysis: classification and regression. Classification algorithms were designed to label SaO₂ signals as SAHS-positive or SAHS-negative. Regression analysis was performed to obtain a reliable estimation of the AHI. Several pattern recognition methods were evaluated for each of these tasks. As a result, a total of 16 classification algorithms and 14 regression algorithms were implemented. Some of them achieved promising results in SAHS diagnosis.

Fourteen time-domain and frequency-domain features were proposed for characterising nocturnal oximetry data. The former ones were computed from the analysis of SaO₂ signals using conventional statistical measurements (mean, variance, skewness and kurtosis) and non-linear methods. *ApEn*, *CTM* and *LZC* were applied to assess irregularity, variability and complexity of the temporal series represented by the sequence of SaO₂ samples. Previous studies found that the evaluation of these methods on oximetry signals provided significant differences between SAHS-negative and SAHS-positive populations (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Hornero *et al.* 2007). The results obtained in the Thesis for *CTM* and *LZC* are in concordance with these preceding studies. The values of these features denote more variability and complexity of SaO₂ signals from SAHS-positive patients. However, there was not a clear trend for *ApEn*, which reflected similar irregularity in both populations. The individual diagnostic capability of *ApEn* and *LZC* was very limited. Only *CTM* provided high classification accuracy from the database available for this Thesis.

Frequency-domain features were computed from the estimated PSD of oximetry recordings. The statistical properties of the frequency component were evaluated. In addition, the band between 0.010 and 0.033 Hz was analysed since it is closely related to SAHS (Zamarrón *et al.* 2003). It was observed that signal power associated to frequency components in this band tends to be higher in SAHS-positive

patients. This result agrees with previous researches (El-Solh *et al.* 2003; Zamarrón *et al.* 2003). Indeed, the PSD profiles obtained in these studies for SaO₂ signals are similar to those depicted in Figure 15. Some of the proposed frequency-domain features (S_T , S_B and PA) showed high ability to classify SaO₂ signals. However, they provided low agreement with AHI for the regression task. The latter was also observed for features computed in the time domain.

In the following stage of the algorithms, each of the extracted features was normalised to have a zero mean and unit variance distribution. This is a common practice for implementing pattern recognition algorithms. As a result, features with a larger magnitude will have the same influence as others on the output provided by the algorithm (Bishop, 1995). This linear transformation does not have any effect on linear models such as FLD, LR or MLR. Additionally, it should be noted that it affects the mapping obtained from PCA as the covariance between input features is modified. PCA was used for dimensionality reduction in order to prevent overfitting. Moreover, using a subset of uncorrelated components was expected to remove redundant information, resulting in a better representation of SaO₂ signals for classification and regression problems. However, the overall outcome indicates that PCA preprocessing did not improve the performance of the evaluated methods. The results show that $All-FLD_c$ and $All-PNN_c$ were the only classification models using the complete feature set improved by their PCA-based counterparts. Indeed, the $PCA-FLD_c$ algorithm showed high generalisation ability. It achieved the highest diagnostic accuracy (89.58% of the subjects in the test set were correctly diagnosed) among those classifiers fed by the selected PCA components. A similar tendency can be observed for regression methods. The influence of PCA was positive for MLR and BY models. In particular, ICC increased from 0.80 for the $All-MLR_r$ algorithm to 0.87 provided by the $PCA-MLR_r$ algorithm, i.e. the linear combination of the components selected from PCA led to higher agreement with AHI than that obtained from the original features. In contrast, ICC substantially decreased for MARS models when PCA was used. Therefore, the analysis of patterns in the space of PCA components enabled to improve classification and regression results only for some pattern recognition methods. This result suggests that information in the discarded components may be useful for classification and regression tasks. It is reflected, for example, in the results obtained by MLP networks or

MARS. Despite dimensionality reduction contributes to reduce model complexity, the new components do not represent an optimum subset for the proposed problems.

The design of pattern recognition algorithms aims to find the model with the highest generalisation ability. The problem is to achieve the optimum balance between bias and variance, i.e., given a training set, the optimum model complexity for the underlying problem must be determined. Usually, pattern recognition methods present some design parameters that directly influence the degree of smoothness of the final solution. A model selection process is required to assign appropriate values for these parameters. An empirical approach was used to perform model selection. Several configurations of a pattern recognition algorithm were evaluated and that with the highest generalisation ability was selected as optimum. Ideally, the hold out method should be used to estimate generalisation ability (Bishop, 1995). However, the number of signals available for the study was limited and a leave-one-out cross-validation strategy was adopted. It is worth noting that only data in the training set was used for this purpose. Once model selection was completed, a model with the selected configuration was trained using the complete training set. The definitive algorithm was then obtained. Data in the test set was preserved to obtain an unbiased estimation for the generalisation ability of this algorithm. The main drawback of this approach is that an optimum configuration is obtained from model selection but not the optimum definitive algorithm. It is derived by training a new model with the selected configuration using the complete training set.

Results on the test set were used to compare the performance of the implemented classification and regression algorithms. The classifier with the best generalisation ability was the *All-MLP_c* algorithm. It provided 94.79% sensitivity, 87.50% specificity and 92.36% accuracy. Other algorithms such as *PCA-FLD_c*, *All-LR_c*, *All-BY_c* and *All-SVM_c* also achieved high classification performance. In particular, *PCA-FLD_c* and *All-LR_c* are of special interest due to its simplicity when compared to MLP-based models. The set of adaptive parameters defining these algorithms is unique for a given training set. In addition, FLD and LR methods are not very sensitive to overfitting due to their simple architecture. Hence, they enable to avoid the exhaustive model selection process required to design MLP networks. Although classification results suggest that a non-linear decision boundary is required to separate SAHS-negative and

SAHS-positive groups, linear boundaries as those defined by $PCA-FLD_c$ and $All-LR_c$ also provided a high percentage of correct classification. Therefore, these algorithms can be pointed out as an effective solution for the proposed classification problem. They outperformed other more complex algorithms based on RBF, SVM or BY methods. The utility of linear methods was previously evaluated in a preceding study of this Thesis focused on classification of SaO_2 signals to help in SAHS diagnosis (Marcos *et al.* 2009). LDA and LR achieved 87.61% and 86.73% classification accuracy using spectral features from oximetry data, respectively. Additionally, the combination of PCA preprocessing and LDA for classifying oximetry patterns was proposed in a different research (Marcos *et al.* 2010b). An accuracy of 93.02% was reached using this classifier with a different database.

The main disadvantage of classification algorithms is that any information about the severity of SAHS is given. Some of the implemented algorithms are capable of estimating posterior probabilities, which can be used to obtain further knowledge. Probability values around the decision threshold can be interpreted as doubtful subjects who should be analysed by a human expert, i.e. the output of the algorithm can be used to indicate one of three possibilities: SAHS-negative, SAHS-positive or rejected classification. However, this procedure involves the definition of values for rejection thresholds (Bishop, 1995). Regression analysis from SaO_2 signals enables to overcome this drawback. It addresses the problem of estimating the AHI in order to quantify the severity of SAHS. This approach provides more useful information than the classification model used for SAHS diagnosis. Several regression algorithms achieved significant agreement with the true AHI derived from PSG. Specifically, $All-MLP_r$, $PCA-MLP_r$, $PCA-BY_r$ and $All-MARS_r$ reached ICC equal or greater than 0.90. $PCA-BY_r$ and $All-MARS_r$ algorithms showed the highest diagnostic ability with an accuracy of 87.50% (88.54% sensitivity and 85.42% specificity) when the estimated AHI was used to diagnose subjects. Both algorithms were more accurate than the others as reflected by their E_{rel} on SAHS-negative and SAHS-positive groups. On the other hand, the results obtained from regression analysis suggest that the functional relationship between the AHI and the features extracted from SaO_2 recordings is non-linear. The two linear algorithms implemented in the study ($All-MLR_r$ and $PCA-MLR_r$) exhibited lower generalisation ability than non-linear models. Despite the $PCA-MLR_r$ algorithm

achieved an ICC of 0.87, E_{RMS} and E_{rel} values reveal low reliability for the estimated AHI.

The performance achieved by the implemented pattern recognition algorithms was compared with that of the input features extracted from oximetry data. Tables 5 and 7 summarise the individual performance of each feature in classification and regression problems, respectively. CTM and S_B were the features with the highest discrimination capability. Diagnostic accuracy reached by CTM was 90.28% (89.58% sensitivity and 91.67% specificity) while S_B provided an accuracy of 91.67% (91.67% sensitivity and 91.67% specificity). Thus, the improvement derived from using pattern recognition techniques for classification was not relevant since similar performance was achieved with CTM or S_B alone. Conversely, algorithms based on pattern recognition techniques clearly outperformed each of the input features for the regression problem. The linear model built from CTM showed the highest agreement with AHI. It provided $ICC = 0.77$, which is far from the ICC values around 0.90 obtained through the combination of features using pattern recognition techniques.

Similarly, the algorithms proposed in the Thesis were compared with conventional oximetry indices: minSaO_2 , CT90, ODI3, ODI4, SIT90 and Δ index were considered for comparison. Their utility in classification and regression problems was evaluated. The results corresponding to each of these indices are presented in Tables 21 and 23. The Δ index achieved the highest classification accuracy. A correct decision was provided for 88.19% of subjects in the test set (86.46% sensitivity and 91.67% specificity). An AUC of 0.960 was reached using this index alone. ODI3 and ODI4 also achieved significant results in the classification problem with accuracy of 86.81% on the test set. Methods based on signal processing techniques improved the diagnostic ability of conventional oximetry indices. Only the spectral feature S_B increased classification accuracy up to 91.67%. Even higher accuracy (92.36%) was reached by processing time-domain and frequency-domain features from SaO_2 recordings using MLP networks ($All-MLP_c$ algorithm). For the regression problem, ODI3 was the most useful index to estimate the AHI. The linear model built from this feature achieved $ICC = 0.91$, which equals the highest ICC provided by the algorithms based on multivariate pattern analysis. However, it was noticed that the regression model using ODI3 overestimated AHI for small target values while it provided reliable estimations for high

AHI. This effect can be observed in Bland-Altman and regression plots depicted in Figure 38. It is also reflected by E_{rel} , which was high for SAHS-negative subjects (5.92) and small for SAHS-positive subjects (0.25). As a result, the diagnostic ability of the regression model based on ODI3 was very poor. All the controls were misclassified since it did not provide reliable estimations of the AHI for mild-SAHS subjects. Therefore, the use of pattern recognition methods for regression analysis is justified since they significantly outperformed conventional oximetry indices. $PCA-BY_r$ and $All-MARS_r$ were the most accurate regression algorithms. In addition to high ICC , these algorithms achieved small error for all the range of AHI values. Neither overestimation nor underestimation effects were observed.

The high prevalence of SAHS together with the difficulties related to PSG have motivated the search for new valid alternatives to help in SAHS diagnosis. As in the present study, most of the effort has been focused on the analysis of oximetry signals. A wide variety of methods have been evaluated, ranging from simple visual inspection to the use of signal processing techniques. Conventional oximetry indices have been commonly used for automated analysis of SaO_2 data. Their diagnostic utility was assessed in preceding studies. Table 24 summarises the results obtained by other researchers with these indices or other similar measures from oximetry recordings. There were notable differences among studies, showing a high dependence on the signals used to evaluate them. Sensitivity ranged from 25 to 100% while specificity from 40 to 100, which reflects some uncertainty about the reliability of these indices for SAHS diagnosis. In this Thesis, some of the oximetry indices provided acceptable diagnostic results. Similar findings were reported in previous studies involving a significant number of subjects. For instance, Gurubhagavatula *et al.* (2001) estimated the diagnostic capability of ODI3 in 86% sensitivity and 91% specificity using 359 subjects and a cut-off point of 5 h^{-1} to define SAHS. Nigro *et al.* (2009) evaluated a portable device to compute ODI3, which provided a sensitivity of 88% and a specificity of 94%. The study by Vázquez *et al.* (2000) is particularly interesting. It proposes to use the respiratory disturbance index (computed from the number of desaturations greater than 4%) to diagnose SAHS. This index enables to evaluate the degree of severity of SAHS in a patient. Using a threshold of $AHI = 10 \text{ h}^{-1}$ to define SAHS, it reached a sensitivity of 97% and a specificity of 80%. In addition, the correlation between the respiratory disturbance index and AHI was estimated at 0.97. The mean of the

Study	Subjects	Threshold (h ⁻¹)	Features/Methods	Se (%)	Sp (%)	Acc (%)
Chiner <i>et al.</i> 1999	275	10	ODI4	71	93	75
Epstein and Dorlac 1998	100	10	Change in SaO ₂ > 4% and < 90%	74	89	-
			Repeated short-duration fluctuations	96	55	-
Golpe <i>et al.</i> 1999	116	10	ODI4	32	97	-
			CT90	84	97	-
			Resaturation index > 3%	29	97	-
Gurubhagavatula <i>et al.</i> 2001	359	5	ODI3	86	91	-
			ODI4	85	90	-
Gyulay <i>et al.</i> 1993	98	15	ODI3	51	90	-
			ODI4	40	98	-
			CT90	93	51	-
Hussain and Fleethan 2003	30	15	ODI3	25	88	63
			ODI4	33	88	67
Lévy <i>et al.</i> 1996	301	15	Δ index	98	46	80
Magalang <i>et al.</i> 2003	224	15	ODI4	94	44	-
			Δ index	91	59	-
Nigro <i>et al.</i> 2009	154	10	ODI3	88	94	-
Olson <i>et al.</i> 1999	793	15	CT90	75	46	-
			Δ index	89	40	-
Pepin <i>et al.</i> 1991	26	5	Δ index	75	86	-
Ryan <i>et al.</i> 1995	69	15	ODI4	31	100	-
Sano <i>et al.</i> 1998	40	15	ODI3	74	83	-
Series <i>et al.</i> 1993	240	10	Decrease of > 4% in SaO ₂ leading to a value lower than 90%	98	48	71
Series <i>et al.</i> 2005	50	15	Fall in SaO ₂ > 2% followed by a rise in SaO ₂	85	93	-
Teramoto <i>et al.</i> 2002	75	15	ODI4	92	92	-
Vázquez <i>et al.</i> 2000	245	15	Respiratory disturbance index (number of drops in SaO ₂ greater than 4%)	98	88	-
		10		97	80	-
Williams <i>et al.</i> 1991	36	10	ODI4	75	100	-
This Thesis	144	10	All-MLP _c	95	88	92
			PCA-BY _r	89	85	88
			All-MARS _r	89	85	88

Table 24. Summary of the results reported in some previous studies based on the use of conventional oximetry indices to assist in SAHS diagnosis.

differences between AHI and its estimation obtained through Bland-Altman analysis was 2.18 h^{-1} , which indicates a slight underestimation tendency.

Other methods based on statistical data analysis and signal processing have been also applied to SaO₂ recordings with diagnostic purposes. Magalang *et al.* (2003) developed a predictor model for AHI based on MARS and conventional oximetry indices. It achieved high correlation (0.84) with a sensitivity of 90% and a specificity of 70% in SAHS detection. This study represents an interesting approach based on the combination of several oximetry indices using pattern recognition methods. Zamarrón *et al.* (1999) analysed the properties of SaO₂ signals in the frequency domain. They found

that the spectrum of recordings corresponding to SAHS-positive subjects tends to have a peak between 0.010 and 0.033 Hz. This result was used to define some of the frequency-domain features proposed in the present study. A similar behaviour was subsequently reported for HRV signals (Zamarrón *et al.* 2003). The dynamical properties of SaO₂ signals were also evaluated with non-linear methods. It was found that recordings corresponding to SAHS-positive subjects presented higher irregularity, variability and complexity using *ApEn*, *CTM* and *LZC*, respectively (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Hornero *et al.* 2007). As indicated before, the results reported by these studies motivated the use of these non-linear methods to characterise oximetry data in this Thesis. The analysis of signal variability using Poincaré plot-based methods such as *CTM* was also proposed in other studies. Morillo *et al.* (2009) achieved 90.9% sensitivity and 84.0% specificity using this kind of analysis. Schmittendorf *et al.* (2009) reported 100% sensitivity and 98.6% specificity by means of a modified version of *CTM*. These results confirm the effect of SAHS on SaO₂ variability as well as the diagnostic utility of *CTM*. Indeed, it was the second feature with the highest classification accuracy and the best predictor of the AHI among all the oximetry features proposed in the Thesis. Finally, techniques for multivariate pattern analysis were applied to simultaneously process spectral and non-linear features from SaO₂ data for SAHS diagnosis. Some results related to this Thesis, as well as the methodology used for the design of pattern recognition algorithms, can be found in Marcos *et al.* (2008a; 2008b; 2009; 2010b). A more extended feature set including conventional statistical analysis in time and frequency domains was used in Marcos *et al.* (2010a) and Álvarez *et al.* (2010).

Other biomedical signals and data have been considered for the development of novel diagnostic methods. In addition to oximetry signals, clinical and demographic data as well as ECG or airflow recordings have been the most commonly used. Even the combination of several signals has been suggested. Usually, a set of different quantitative measures is used to provide a decision. Thus, some kind of pattern analysis technique is required. Conventional methods such as discriminant analysis, LR or MLR are widely extended among the scientific community. Some examples of their use for SAHS detection can be found. Recently, LDA (Marcos *et al.* 2010b) and LR (Álvarez *et al.* 2010) were proposed to classify SaO₂ signals from subjects suspected of suffering from SAHS. The diagnostic accuracy reached in these studies was 93.02% and

89.7%, respectively. Friedman *et al.* (2010) reported an accuracy of 82.5% using a LDA-based classifier to process clinical features and questionnaire responses. Discriminant analysis was also used in some relevant studies focused on ECG analysis (de Chazal *et al.* 2000; de Chazal *et al.* 2003; de Chazal *et al.* 2004; Shouldice *et al.* 2004; Redmond and Heneghan 2006; Heneghan *et al.* 2008b; Méndez *et al.* 2010). The main advantages of these conventional methods are their extensive theoretic background and the absence of any design (model selection) process. However, they are expected to perform worse than other complex techniques such as neural networks in classification and regression problems, for which the optimum solution is markedly non-linear. In general, no prior information about the type of solution is available and the designer must choose the method used for pattern analysis. If the amount of data is large enough, the use of more capable methods is encouraged since they are expected to perform at least as well as conventional techniques in linear problems, while represent a more appropriate choice if non-linear modelling abilities are required.

Despite the design of complex pattern recognition methods is a difficult task, their use in novel medical diagnosis applications has increased in the last decades. Some examples related to SAHS detection can be found in the literature. Kirby *et al.* (1999) developed a regression model to estimate the AHI from clinical features using GRNN. Sensitivity and specificity were 98.9% and 80%, respectively. However, results about the quality of the estimations were not provided. Another regression algorithm based on neural networks was developed by El-Solh *et al.* (1999). Clinical and anthropomorphic features were used together with MLP networks. A feature selection stage was included to determine the optimum feature set. The MLP-based regression algorithm achieved a correlation of 0.86 with the AHI derived from PSG. It outperformed linear approximation using MLR, as observed in the results of this Thesis. The algorithm based on MARS modelling proposed by Magalang *et al.* (2003) is another example of the use of non-linear pattern recognition methods to assist in SAHS detection.

The cited studies are based on regression techniques to approximate the AHI. As suggested above, the regression approach represents a more interesting tool to model SAHS diagnosis since it provides information about the degree of severity. Additionally, the estimated AHI can be extended to different cut-off points to define

SAHS (there is no consensus about this issue and the threshold used in the literature varies from 5 h⁻¹ to 15 h⁻¹). It represents a handicap for studies focused on the classification of subjects where any information about the severity of SAHS is provided. However, complex pattern recognition techniques have been also used to model SAHS diagnosis as a classification problem. As previously indicated, Marcos *et al.* (2008a, 2008b, 2010a) used MLP and RBF networks to process oximetry features. Khandoker *et al.* (2009a, 2009b) developed a classification algorithm from ECG recordings using wavelet analysis and SVM models. A high accuracy of 92.85% was reported using a reduced test set with 42 subjects (Khandoker *et al.* 2009a).

Table 25 summarises some of the latest studies focused on SAHS diagnosis

Study	Subjects	Threshold (h ⁻¹)	Signals and data	Features/Methods	Se (%)	Sp (%)	Acc (%)
Álvarez <i>et al.</i> 2006	187	10	SaO ₂	CTM	90	83	87
Álvarez <i>et al.</i> 2010	148	10	SaO ₂	Classification of time-domain and frequency-domain features using LR	92	85	90
Caballero <i>et al.</i> 1998	29	10	Physical examination	KNN classifier	64	87	-
Caseiro <i>et al.</i> 2010	41	5	Airflow	Time-frequency analysis	81	95	-
Del Campo <i>et al.</i> 2006	187	10	SaO ₂	ApEn	88	83	-
de Chazal <i>et al.</i> 2003	30	15	HRV and EDR derived from ECG	Quadratic discriminant using time-domain and frequency-domain features from HRV and EDR	-	-	100
de Chazal <i>et al.</i> 2009	183	15	SaO ₂ and ECG	LDA for processing time-domain and frequency-domain features	95	83	-
El-Solh <i>et al.</i> 1999	269	10	Clinical and anthropomorphic features	Regression analysis using MLP networks	95	65	-
Fiz <i>et al.</i> 2010	37	15	Snoring	LR classifier from time and spectral features	80	90	-
Friedman <i>et al.</i> 2010	223	5	Questionnaire and clinical features	LDA	81	83	83
Heneghan <i>et al.</i> 2008a	59	10	SaO ₂ and ECG	LDA for processing temporal features from both signals	82	91	-
Heneghan <i>et al.</i> 2008b	92	15	ECG	LDA for classifying epochs and deriving the AHI value	92	69	-
Hilton <i>et al.</i> 1999	40	15	ECG	Spectral analysis of HRV	100	89	-
Hoffstein and Szalai 1993	594	10	Clinical features and physical examination	MLR analysis	60	63	-
Hornero <i>et al.</i> 2007	187	10	SaO ₂	ApEn	82	87	-
Khandoker <i>et al.</i> 2009a	30	10	HRV and EDR derived from ECG	SVM classifiers from wavelet features of HRV and EDR	-	-	100
Khandoker <i>et al.</i> 2009b	125	10	HRV and EDR derived from ECG	SVM classifiers from wavelet features of HRV and EDR	92	94	93

Table 25. Summary of some of the recent studies focused on SAHS diagnosis using signal processing techniques.

Study	Subjects	Threshold (h ⁻¹)	Signals and data	Features/Methods	Se (%)	Sp (%)	Acc (%)
Kirby <i>et al.</i> 1999	405	10	Clinical and demographic features	Regression analysis using GRNN	99	80	91
Magalang <i>et al.</i> 2003	224	15	SaO ₂	MARS model from conventional oximetry indices	90	70	-
Marcos <i>et al.</i> 2008a	187	10	SaO ₂	RBF classifiers and non-linear features	89	81	86
Marcos <i>et al.</i> 2008b	187	10	SaO ₂	MLP classifiers and non-linear features	90	79	86
Marcos <i>et al.</i> 2009	187	10	SaO ₂	LDA and spectral features	91	83	88
Marcos <i>et al.</i> 2010a	187	10	SaO ₂	BY classifiers using time-domain and frequency-domain features	88	82	86
Marcos <i>et al.</i> 2010b	214	10	SaO ₂	LDA and PCA from spectral and non-linear features	97	79	93
Méndez <i>et al.</i> 2009	25	10	HRV and area of the QRS complex series derived from ECG	KNN and MLP classifiers from time-domain and frequency-domain features of the series for identifying apnoeic epochs	-	-	100
Méndez <i>et al.</i> 2010	25	10	ECG and HRV	Classification of signal epochs with LDA and QDA to provide a final diagnosis	-	-	100
Morillo <i>et al.</i> 2009	117	15	SaO ₂	Poincaré plots	91	84	-
Nakano <i>et al.</i> 2004b	383	15	Tracheal sound	Spectral analysis	79	95	-
Ng <i>et al.</i> 2008	40	10	Snoring	Spectral analysis	88	82	-
Roche <i>et al.</i> 1999	91	10	HRV derived from ECG	LR classifier from time-domain features of HRV	90	98	-
Roche <i>et al.</i> 2003	147	10	HRV derived from ECG	Decision tree classifier from HRV wavelet features	76	64	-
Roche <i>et al.</i> 2002	210	15	SaO ₂ and clinical score	LR MLR	- -	- -	53 62
Schmittendorf <i>et al.</i> 2009	142	-	SaO ₂	Modified version of <i>CTM</i>	100	99	-
Zamarrón <i>et al.</i> 1999	233	10	SaO ₂	Spectral analysis	94	65	-
Zamarrón <i>et al.</i> 2003	300	10	SaO ₂ and HRV	Spectral analysis	94	82	89
Zamarrón <i>et al.</i> 2006	187	10	HRV	<i>ApEn</i>	71	79	-
This Thesis	144	10	SaO ₂	<i>All-MLP_c</i>	95	88	92
				<i>PCA-BY_r</i>	89	85	88
				<i>All-MARS_r</i>	89	85	88

Table 25 (Cont.). Summary of some of the recent studies focused on SAHS diagnosis using signal processing techniques.

using advanced signal processing techniques. Some of them reported a perfect diagnostic performance with an accuracy of 100%. However, an unbiased comparison among different methods cannot be carried out since different databases are usually analysed. Public databases are needed for a fair assessment. For instance, most of the studies based on ECG analysis used a free signal database available through the Physionet website (Goldberger *et al.* 2000). Indeed, a competition using this database of 75 ECG signals was promoted (Penzel *et al.* 2002). However, mild-SAHS subjects were removed from the analysis. The lack of a globally accepted definition of SAHS

represents an additional handicap. Diagnostic results are computed by using a specific threshold for AHI which may not correspond with that used by other researchers. In this Thesis, the validity of the proposed methods was analysed by comparing their performance with that of conventional oximetry indices from the same database. In addition, pattern recognition algorithms were compared with each of the features extracted from SaO₂ data to evaluate the benefit derived from multivariate analysis.

Despite the results achieved in the Thesis were satisfactory, some limitations can be pointed out. For instance, the size of the database could be considered as a constraint. Highly complex (flexible) pattern recognition methods were used and they could be affected by overfitting. These methods enable to model complex input-output mappings. However, it is required to control model complexity to prevent this phenomenon. An exhaustive model selection search enables to avoid model configurations with poor generalisation ability. Additionally, regularisation techniques have shown effective results for MLP networks. It has been checked that weight decay regularisation improved the capability of MLP algorithms used for classification and regression problems. Furthermore, it was checked that once the optimum weight decay parameter was found, increasing the network size was not accompanied by a decrease of generalisation ability. BY models also include regularisation during network training by means of the hyperparameter α_k . For these networks, it was observed that large architectures for which the number of weights was even greater than the number of training samples performed well. The analysis for the magnitude of the weights showed that they tended to be smaller for larger networks due to weight decay regularisation. Thus, the size of the weights represents a relevant factor for evaluating network complexity. As suggested by Bartlett (1998), it is even more important than the number of adaptive parameters. Bartlett stated that if the learning algorithm is able to find a network with small weights by minimising the error on the training set, then the generalisation depends on the size of the weights. This may explain the generalisation performance of MLP and BY networks when the number of training samples is even smaller than the number of weights. However, in the case of BY models, using excessively large network architectures may cause the Gaussian approximation to be invalid. This effect was described by MacKay (1992a) in previous experiments.

Nevertheless, increasing the size of the training set would be desirable as a better description of the statistical properties of the problem would be given. As a result, the number of functions that can fit training data could be reduced (i.e. variance is reduced), leading to more accurate algorithms. Similarly, using more test examples would increase the reliability of the final diagnostic results. These represent an estimation of the generalisation ability of the algorithms since a finite test set is used. Moreover, another property of the database to be improved is the proportion of SAHS-positive and SAHS-negative subjects. Most of the implemented classifiers provided higher sensitivity than specificity. It may be due to the unbalanced number of subjects from both classes in the training set, which encourages algorithms to minimise the number of misclassifications in the group of SAHS-positive subjects to reduce the overall error. The presence of other respiratory disorders different to SAHS may be another possible cause of false positive misclassifications. For instance, chronic obstructive pulmonary disease (COPD) can affect the dynamical behaviour of oximetry recordings similarly to SAHS, leading algorithms to produce incorrect decisions for this group of patients.

In summary, SAHS diagnosis has been modelled using pattern recognition techniques for classification and regression. The methods proposed in the Thesis enable automated analysis of oximetry recordings. They represent an efficient novel tool to help in SAHS detection since high generalisation ability was achieved in classification and regression tasks. The analysis of time-domain and frequency-domain features from SaO₂ data using pattern recognition algorithms outperformed the diagnostic ability of conventional oximetry indices. In particular, regression algorithms provide information about the degree of severity of SAHS since they are capable of accurately estimating the AHI. It represents an added value when compared with classification algorithms for SAHS diagnosis.

9.3. Conclusions

A set of 14 time-domain and frequency-domain features was used to characterise the dynamical properties of oximetry signals. They were used to implement 16 classification algorithms and 14 regression algorithms for SAHS detection. Their diagnostic ability was evaluated on an independent test set with 144 subjects. Several

algorithms achieved high performance in the SAHS diagnosis problem. The *All-MLP_c* classifier provided 92.36% accuracy and 0.958 *AUC*. It was the highest classification accuracy among all the algorithms evaluated in the study. In addition, it was checked that using one of the spectral features from SaO₂ data alone (the signal power contained in the band between 0.010 and 0.033 Hz) enabled to reach a classification accuracy of 91.67% and an *AUC* of 0.968. The regression algorithms with the highest diagnostic accuracy were *PCA-BY_r* and *All-MARS_r*, with a correct diagnosis rate of 87.50% and *ICC* higher than 0.90. The *All-MLP_r* and *PCA-MLP_r* algorithms also achieved significant results in the regression problem with an *ICC* of 0.91 and 0.90, respectively. **The analysis of SaO₂ recordings using these pattern recognition algorithms enable to achieve high diagnostic accuracy for SAHS detection. They outperformed other traditional methods for interpreting these signals such as visual inspection and conventional oximetry indices. Therefore, they could be considered as an automated method to assist clinicians in SAHS screening, contributing to reduce the number of required PSG studies.**

In addition to this global conclusion, the results of the study lead to other observations:

1. Signal processing techniques enable to extract additional information from oximetry data that can be used to detect SAHS. In this Thesis, SAHS diagnosis has been modelled using pattern recognition techniques. These methods allow to simultaneously use several features for the diagnosis task. They have the capability of adapting themselves to the data even in the case of non-linear relationships, which represents their main attribute. Table 26 summarises the results of the most accurate algorithms from this Thesis. Despite they achieved high diagnostic results, these algorithms should be validated in a larger signal database in order to be applied in clinical practice.
2. The size of the database of SaO₂ recordings was enough to perform the present study. The achieved results serve as a reference for future work developed from the same database. They enabled to extract relevant conclusions about the evaluated algorithms. However, a higher number of

Classification				
Algorithm	<i>Se</i> (%)	<i>Sp</i> (%)	<i>Acc</i> (%)	<i>AUC</i>
<i>All-MLP_c</i>	94.79	87.50	92.36	0.958
Regression				
Algorithm	<i>Se</i> (%)	<i>Sp</i> (%)	<i>Acc</i> (%)	<i>ICC</i>
<i>PCA-BY_r</i>	88.54	85.42	87.50	0.90
<i>All-MARS_r</i>	88.54	85.42	87.50	0.91

Table 26. Classification and regression algorithms with the highest diagnostic accuracy.

signals would be required for future projects involving complex methods such as neural networks. They usually have a large number of adaptive parameters that are adjusted according to the statistical properties of the problem. Thus, an important amount of data is required to provide a detailed description.

3. The results suggest that the optimum solution for classification (decision boundary) and regression (functional relationship) problems is non-linear. An exhaustive study of the most relevant pattern recognition methods has been performed. There exist other methods for classification and regression that were not included in the Thesis. However, a significant number of techniques were evaluated. They include linear methods widely used in different applications as well as more complex techniques that are capable of non-linear modelling. The diagnostic utility of 16 classification algorithms and 14 regression algorithms has been assessed. Several algorithms from each group achieved significant performance. In addition, the results show that complex non-linear algorithms outperformed linear methods. Despite a remarkable difference was not observed, it suggests that optimum solutions for classification and regression problems are non-linear. However, a more detailed description of the statistical properties of both problems is required to check this observation.
4. A thorough design and comparison of several pattern recognition algorithms is required to achieve high performance for the underlying problem. For a given training set, one or more smoothing parameters must be usually

adjusted to adapt the complexity of the algorithm to the problem. The influence of each smoothing parameter on generalisation performance must be analysed in the model selection phase. Those methods with a higher degree of flexibility require more smoothing parameters to be optimised, resulting in a more complicated design process. Conversely, these methods have the potential to solve more complex problems. Thus, there is a trade-off between the design cost and the capability of a pattern recognition method. There is no prior information about the most appropriate method for a given problem and the designer should evaluate some of them to identify the optimum. This approach was followed in this study.

5. *CTM* was the non-linear feature with the highest diagnostic capability. Previous studies reported statistically significant differences between SAHS-negative and SAHS-positive subjects by evaluating non-linear methods such as *ApEn*, *CTM* and *LZC* on SaO₂ signals (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Hornero *et al.* 2007). However, the results of the Thesis reveal that only *CTM*, which evaluates signal variability, provided high diagnostic accuracy. Signal complexity and irregularity measured by *LZC* and *ApEn*, respectively, achieved poor diagnostic results. In particular, *ApEn* did not reflect any diagnostic capability, reflecting a similar degree of irregularity for oximetry recordings corresponding to SAHS-negative and SAHS-positive subjects.
6. Spectral analysis of SaO₂ signals reveals significant differences between SAHS-negative and SAHS-positive subjects. Zamarrón *et al.* (2003) found that SaO₂ signals of patients suffering from SAHS tend to have more power in the band between 0.010 and 0.033 Hz due to periods with repeated apnoeas. A similar behaviour was observed in the study by El-Solh *et al.* (2003). The results obtained in the Thesis for spectral analysis of SaO₂ signals are consistent with these previous studies. The features extracted from the PSD (*S_T*, *S_B* and *PA*) achieved high diagnostic accuracy. Therefore, it can be stated that spectral analysis represents a more efficient technique to evaluate the dynamical properties of SaO₂ data regarding non-linear methods such as *ApEn*, *CTM* and *LZC*.

7. Pattern recognition algorithms improved the ability of conventional oximetry indices to classify subjects as SAHS-negative or SAHS-positive. The MLP classifier provided a correct decision for 92.36% of subjects in the test set using all the extracted features, which is the highest diagnostic accuracy among all the evaluated algorithms. It outperformed conventional oximetry indices for the proposed classification problem. Therefore, the proposed classification method represents a new valid technique for automated analysis of SaO₂ signals.
8. The most efficient approach for the classification problem is spectral analysis of SaO₂ signals. Despite the *All-MLP_c* classifier provided the highest diagnostic accuracy on the test set (92.36%), similar performance can be achieved by using S_B (signal power in the band between 0.010 and 0.033 Hz) computed from the PSD. It provided a classification accuracy of 91.67% using the optimum threshold determined from ROC analysis of training data, which is slightly lower than that of the MLP classifier. Therefore, multivariate analysis does not add a great advantage for the classification problem. A single feature from spectral analysis of SaO₂ recordings is enough to improve conventional measures and achieve high diagnostic performance. A substantially less complex methodology is required to compute S_B from SaO₂ data in comparison with the complete MLP algorithm. Thus, this spectral feature can be pointed out as the optimum solution for the proposed classification problem.
9. A great benefit was derived from the use of pattern recognition methods for the proposed regression problem. A linear regression model was built for each input feature to evaluate its individual correlation with the reference AHI. The model based on *CTM* provided the highest *ICC* with a value of 0.77. Several pattern recognition algorithms achieved *ICC* greater than 0.90, reflecting a remarkable improvement. Specifically, *All-MARS_r* and *PCA-BY_r* algorithms showed the highest performance, with a diagnostic accuracy of 87.50% from the estimated AHI. The proposed models for regression analysis were compared with conventional oximetry indices. A linear

regression model was built for each of them. An *ICC* of 0.91 was obtained by using ODI3 alone. However, it was checked that this model overestimated the AHI for control and mild-SAHS subjects, showing no ability to perform SAHS diagnosis. Therefore, the optimum solution for the regression problem is given by the suggested pattern recognition algorithms for SaO₂ analysis.

10. Dimensionality reduction using PCA did not improve generalisation performance for most of the evaluated methods. PCA was expected to prevent overfitting by reducing the dimensionality of the input pattern. However, the results suggest that relevant information is lost due to discarded components. PCA is an unsupervised method and it does not take into account information in the target variable (Bishop, 1995; Zhang, 2000). Therefore, the selected components (ordered depending on the variance explained by each of them) might not provide the optimum representation for the problem under study.
11. The regression approach provides more useful information about SAHS than that based on pattern classification. SAHS diagnosis from SaO₂ recordings was addressed from two different schemes: classification and regression. Despite the highest diagnostic accuracy is provided by one of the classification algorithms, the regression approach is preferred. Regression algorithms provide more useful information since they can indicate the severity of SAHS. Additionally, they are insensitive to the definition used for a positive diagnosis since algorithms are focused on the estimation of the AHI. The performance of regression algorithms was measured by means of the *ICC* between the estimated and reference AHI. It measures the reliability of the predicted outcomes instead of evaluating the linear relation between both variables, as measured by the Pearson correlation coefficient. Additional information such as that provided by E_{RMS} , E_{rel} , Bland-Altman and regression plots is required to obtain a complete assessment of regression algorithms.

12. Most of the misdiagnosed subjects correspond to mild-SAHS patients. The *All-MLP_c* algorithm misclassified 11 subjects and 7 of them had $5 \text{ h}^{-1} \leq \text{AHI} \leq 15 \text{ h}^{-1}$. The AHI estimated by the *All-MARS_r* algorithm led to an incorrect decision for 18 subjects in the test set, with 14 mild-SAHS cases. It can be expected that misclassifications will be produced for this group of subjects, which indicates a coherent behaviour of the algorithms. It must be taken into account that these algorithms were developed to analyse SaO₂ signals recorded through portable monitoring. The Portable Monitoring Task Force of the American Academy of Sleep Medicine indicates that portable devices may be used as an alternative to PSG for the diagnosis of SAHS in patients with a high pretest probability of moderate to severe SAHS (Collop *et al.* 2007). Thus, the proposed algorithms could be pointed out as a valid tool for portable SAHS diagnosis, reducing the demand for in-laboratory PSG studies.
13. The global analysis of classification and regression results reveals the high capability of MLP networks to model pattern recognition problems. These algorithms are the most popular neural network models and have been successfully applied to a wide variety of problems (Zhang, 2000). In this study, MLP networks outperformed other models such as RBF and SVM, which also have the ability of capturing non-linearities in the data. Weight decay regularisation plays an important role to implement MLP-based algorithms. It has shown to be an effective means for controlling model complexity and has been essential for properly optimisation of MLP algorithms.

9.4. Original contributions

The main contribution of the Thesis is the development of pattern recognition algorithms for SaO₂ analysis in order to provide an automated diagnosis for SAHS. Although some of the features used to characterise oximetry signals were previously evaluated in other studies, this is the first research where they are combined using pattern recognition methods to help in SAHS diagnosis. A database with 240 SaO₂ signals corresponding to subjects suspected of suffering from SAHS was

available. The study reports significant diagnostic results compared with those previously obtained from visual and automated analysis of oximetry recordings.

It is worth noting that some signal processing techniques evaluated in the Thesis were not previously used together with oximetry data. It is objectively proved by the publications related to the Thesis, which are detailed in Appendix B of this document. The original contributions of this Thesis as well as the advantages of the proposed methods are given in the following points:

1. The preprocessing algorithm used to remove artefacts from SaO₂ signals was developed in this study. This step is required since these signals usually present drops to zero due to movements during sleep. In addition, the first minutes of the signals may correspond to the awake state and should not be considered in the analysis. The preprocessing algorithm can be applied to recordings acquired with different equipment to that used in the Thesis.
2. Statistical measurements in time and frequency domains represent a novel set of features for characterising oximetry data. Previously, only the mean value of SaO₂ has been considered for the development of diagnostic rules (Herer *et al.* 1999). However, no other researchers used variance, skewness or kurtosis to characterise the distribution of SaO₂ values. On the other hand, spectral properties of SaO₂ signals have been analysed in previous studies (Zamarrón *et al.* 2003). However, statistical analysis of the frequency component was proposed in this Thesis. Indeed, the study by Marcos *et al.* (2010a) was the first one including statistics in time and frequency domains from SaO₂ recordings for SAHS detection. They have shown to provide relevant information for classification and regression, as shown from ARD results obtained for BY models.
3. This is the first study in which time-domain and frequency-domain features from oximetry data were combined to assist in SAHS diagnosis. It has been shown that the combination of both groups of features can improve the description of SaO₂ recordings. Multivariate pattern analysis was proposed for SAHS detection in other preceding studies using features from SaO₂ data

(Roche *et al.* 2002; Magalang *et al.* 2003). Nevertheless, only conventional oximetry indices were used for this purpose. Typically, they were combined with clinical score and/or demographic data. As suggested in this Thesis, a previous work focused on magnetoencephalogram analysis also proposed the combination of non-linear measurements computed in the time-domain with spectral information for identifying Alzheimer's disease (Hornero *et al.* 2008).

4. The Thesis proposes a comparison between classification and regression approaches to assist in SAHS diagnosis. It represents an interesting viewpoint since the main advantages and disadvantages of both approaches can be compared under the same conditions. Preceding works analysing both schemes for SAHS diagnosis were not published. Several studies focused on the development of pattern classifiers have been cited. Some recent examples are those by de Chazal *et al.* (2009), Fiedman *et al.* (2010) or Khandoker *et al.* (2009b). Similarly, other researchers focused on the estimation of the AHI as proposed in the regression approach (El-Solh *et al.* 1999; Vázquez *et al.* 2000; Magalang *et al.* 2003). This comparative analysis yield one of the conclusions extracted from the Thesis, which reflects that the regression approach is more useful for clinical practice.
5. To our knowledge, RBF and BY networks were not previously used for SAHS detection. In addition to common techniques such as KNN, FLD, LR or MLR, other complex methods were evaluated by other researchers. For instance, MLP, GRNN and MARS were used for AHI regression (El-Solh *et al.* 1999; Kirby *et al.* 1999; Magalang *et al.* 2003). Similarly, PNN or SVM classifiers were developed for identifying patients with SAHS (El-Solh *et al.* 2003; Khandoker *et al.* 2009a; Khandoker *et al.* 2009b). On the other hand, RBF and BY networks were not previously used to assist in SAHS diagnosis. Preceding results of this Thesis were obtained for these methods (Marcos *et al.* 2008a; Marcos *et al.* 2010a).
6. The first SVM-based algorithm for AHI estimation was proposed in this Thesis. Previously, other diagnostic algorithms based on SVM models were

developed (Khandoker *et al.* 2009a; Khandoker *et al.* 2009b). However, a classification approach was considered. The number of studies using these algorithms for solving regression problems is substantially smaller. The satisfactory results achieved by SVM-based regression algorithms in the Thesis reflect their utility in this kind of tasks.

7. The Thesis explores the performance of different pattern recognition methods as a strategy to find the optimum solution for a given problem. As previously described, some of them were applied to SAHS diagnosis by other researchers. Nevertheless, this is the first study in which an exhaustive comparison of several algorithms is carried out. It is a required strategy since the designer does not usually have prior knowledge about the method that best fits the underlying problem.
8. The analysis of regression results using the *ICC* represents an added value for the study. As explained before, it is a robust statistical measure to quantify the reliability for the AHI estimations (Bravo and Potvin 1991). Previous studies focused on AHI regression used the Pearson correlation coefficient to evaluate the quality of the results. However, the problem with this measure is that it only quantifies the linear relationship between both variables and does not take detect systematic errors (Weir, 2005).

9.5. Future work

In this Thesis, SaO₂ recordings were analysed by using pattern recognition methods to help in SAHS diagnosis. The initial database was randomly divided into a training set and a test set to perform the study. Two approaches were proposed to model the diagnosis task: classification and regression. The former aims to provide a correct diagnostic decision considering SAHS-negative and SAHS-positive as possible choices. The objective of the latter approach is to develop accurate methods to estimate the AHI from oximetry data. Several algorithms were evaluated for both schemes. The analysis of the results shows the validity of the methods implemented in the study. Despite the large number of algorithms proposed for both problems, there are some points that can be addressed in future work.

Increasing the number of available SaO₂ signals must be considered. The developed algorithms should be validated in a larger database to confirm the results shown in the Thesis. In addition, it should be noted that oximetry signals analysed in the study were acquired during PSG tests performed in a special sleep unit. Thus, the proposed methods should be evaluated on SaO₂ signals recorded at patient's home using a portable oximeter. Indeed, this is a required practice in order to avoid PSG studies. Portable monitoring devices enable to record data in the usual patient's environment, providing a more reliable analysis of oximetry during sleep. Furthermore, it enables to develop telemedicine-based applications for SAHS diagnosis. Specifically, it would be of special interest transmitting SaO₂ signals from the portable device to a host including the software with the algorithms for automated analysis.

The proposed algorithms achieved high accuracy in classification and regression problems. A high percentage of misclassifications corresponds to mild-SAHS subjects. Future efforts should be focused on providing accurate diagnosis for this group of subjects. Due to the number of algorithms evaluated in the Thesis, a possible way to increase classification and regression accuracy is the use of ensemble methods to combine them. For instance, voting or averaging schemes can be used for combining classification and regression models, respectively (Haykin, 1999). Other ensemble methods to be considered are bagging and boosting, which combine multiple algorithms of a given method (Zhou *et al.* 2002). The bagging algorithm employs bootstrap sampling to generate many training sets from the original training set (Breiman, 1996). Then, a model is trained from each of those training sets. Predictions are combined via simple averaging for regression and majority voting for classification. The boosting algorithms used for classification and regression are AdaBoost and AdaBoost.R2, respectively (Freund, 1995). Both algorithms sequentially generate a series of trained models, where the training instances that are wrongly predicted by the previous models will play more important role in the training of later models (Freund, 1995). Predictions are combined via weighted averaging for regression and weighted voting for classification.

Additional features can be considered for classification and regression problems. Other measures can be obtained from SaO₂ analysis in time and frequency

domains. For instance, other entropy measures could be evaluated since *ApEn* did not provide valuable information for classification or regression when used alone. Another possibility is to use conventional oximetry indices as input features together with those proposed in the study. Some of these indices, such as Δ index and ODI3, have shown an acceptable level of diagnostic accuracy.

As explained in the Thesis, increasing the number of input features may have negative effects for some of the pattern recognition methods. Therefore, reducing the number of inputs may be required to prevent overfitting and poor generalisation performance. PCA was evaluated for dimensionality reduction. However, it did not yield satisfactory results due to its unsupervised nature. Feature selection methods can be considered as a previous stage to the pattern recognition method (Bishop, 1995). The aim is to find the subset of features that leads to the highest generalisation performance. This approach can be useful if there are inputs which carry little useful information or if there is strong correlation between sets of inputs (i.e. the same information is repeated in several variables).

Finally, oximetry recordings could be used together with other physiological data in order to improve diagnostic accuracy. Other biomedical signals related to SAHS can be considered. For instance, some studies suggest combining SaO₂ signals with clinical data (Roche *et al.* 2002), HRV recorded during pulse oximetry (Zamarrón *et al.* 2003) or ECG recordings (Heneghan *et al.* 2008a; de Chazal *et al.* 2009). However, the latter may increase the complexity of the data acquisition process. One of the advantages of SaO₂ signals is the simplicity of pulse oximetry, which only requires to use a finger or ear probe. Thus, it does not disturb patient's sleep and portable equipment could be easily manipulated by patients. New alternative diagnostic methods should consider the possibility of performing portable sleep monitoring at patient's home. Therefore, the effect of adding more signals on the acquisition process must be analysed, taking into account that non-expert users will have to manipulate monitoring devices.

Appendix A. Glossary of terms

<i>Acc</i>	Accuracy
AHI	Apnoea-hypopnoea syndrome
ANOVA	Analysis of variance
<i>ApEn</i>	Approximate entropy
ARD	Automatic relevance determination
<i>AUC</i>	Area under the ROC curve
BMI	Body mass index
BY	Bayesian
CODI	Cardiac oximetry disturbance index
COPD	Chronic obstructive pulmonary disease
CPAP	Continuous positive airway pressure
<i>CTM</i>	Central tendency measure
CT90	Cumulative time spent below 90% of saturation
ECG	Electrocardiogram
EDR	ECG-derived respiratory signal
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
FLD	Fisher's linear discriminant
<i>Fn</i>	False negative
<i>Fp</i>	False positive

GPE	Generalised prediction error
GRNN	Generalised regression neural network
HRV	Heart rate variability
<i>ICC</i>	Intraclass correlation coefficient
IRLS	Iterated re-weighted least squares
LDA	Linear discriminant analysis
<i>LZC</i>	Lempel-Ziv complexity
KNN	<i>K</i> -nearest neighbour
LR	Logistic regression
MAP	Maximum a posteriori
MARS	Multivariate adaptive regression splines
MLP	Multilayer perceptron
MLR	Multiple linear regression
ODI3	Oxygen desaturation index over 3%
ODI4	Oxygen desaturation index over 4%
PCA	Principal component analysis
PNN	Probabilistic neural network
PSD	Power spectral density
PSG	Polysomnography
QDA	Quadratic discriminant analysis
RBF	Radial basis function
RDI	Respiratory disturbance index
ROC	Receiver operating characteristic
SAHS	Sleep apnoea-hypopnoea syndrome
<i>Se</i>	Sensitivity

SaO ₂	Blood oxygen saturation
SCG	Scaled conjugate gradient
SIT	Saturation impairment time
<i>Sp</i>	Specificity
SVM	Support vector machine
<i>Tn</i>	True negative
<i>Tp</i>	True positive

Appendix B. Publications derived from the Doctoral Thesis

- B.1. Publications in peer-reviewed journals included in the Journal Citation Reports
- B.2. Publications in international conferences
- B.3. Publications in national conferences

B.1. Publications in peer-reviewed journals included in the Journal Citation Reports

- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., López, M. and Zamarrón, C. (2008a) Radial basis function classifiers to help in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. *Medical & Biological Engineering & Computing*, 46, pp. 323–332.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M. (2008b) Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. *Computer Methods and Programs in Biomedicine*, 72, pp. 79–89.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and Zamarrón, C. (2009) Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry. *Medical Engineering & Physics*, 31, pp. 971–978.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Nabney, I.T., Del Campo, F. and Zamarrón, C. (2010a) The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome. *Physiological Measurement*, 31, pp. 375–394.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and Aboy, M. (2010b) Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis. *Medical & Biological Engineering & Computing*, 48, pp. 895–902.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V. and Del Campo, F. (2010) Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. *IEEE Transactions on Biomedical Engineering*, 57, pp. 2816–2824.

- ▣ Del Campo, F., Hornero, R., Zamarrón, C., Álvarez, D. and Marcos, J.V. (2010) Variability of pulse signal frequency obtained using nocturnal pulse oximetry in patients with sleep apnoea/hypoapnoea syndrome. *Archivos de Bronconeumología*, 46, 3, pp. 116–121.

B.2. Publications in international conferences

- ▣ Álvarez, D., Gutiérrez, G.C., Marcos, J.V., Del Campo, F. and Hornero, R. (2010) Spectral analysis of single-channel airflow and oxygen saturation recordings in obstructive sleep apnea detection. In: *Proceedings of the 32nd Annual International Conference of the IEEE EMBS, Buenos Aires, September 2010*. Buenos Aires: IEEE, pp. 847–850.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and Zamarrón, C. (2009) Classification of nonlinear features from oximetry signals by means of support vector machines to help in the diagnosis of obstructive sleep apnea. In: *Proceedings of the 6th International Workshop on Biosignal Interpretation, New Haven, June 2009*. New Haven: IEEE, pp. 132–135.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F. and Zamarrón, C. (2009) Spectral entropy and sample entropy analyses of blood oxygen saturation from nocturnal oximetry in obstructive sleep apnea diagnosis. In: *Proceedings of the 6th International Workshop on Biosignal Interpretation, New Haven, June 2009*. New Haven: IEEE, pp. 152–155.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and Zamarrón, C. (2009) A classification algorithm based on spectral features from nocturnal oximetry and support vector machines to assist in the diagnosis of obstructive sleep apnea. In: *Proceedings of the 31st Annual International Conference of the IEEE EMBS, Minneapolis, September 2009*. Minneapolis: IEEE, pp. 5547–5550.

- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F. and López, M. (2009) Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis. In: *Proceedings of the 31st Annual International Conference of the IEEE EMBS, Minneapolis, September 2009*. Minneapolis: IEEE, pp. 400–403.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M. (2008) Single layer network classifiers to assist in the detection of obstructive sleep apnea syndrome from oximetry data. In: *Proceedings of the 30th Annual International Conference of the IEEE EMBS, Vancouver, August 2008*. Vancouver: IEEE, pp. 1651–1654.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F., Zamarrón, C. and López, M. (2008) Applying time, frequency and nonlinear features from nocturnal oximetry to OSA diagnosis. In: *Proceedings of the 30th Annual International Conference of the IEEE EMBS, Vancouver, August 2008*. Vancouver: IEEE, pp. 3872–3875.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and López, M. (2007) Applying neural network classifiers in the diagnosis of the obstructive sleep apnea syndrome from nocturnal pulse oximetric recordings. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, August 2007*. Lyon: IEEE, pp. 5174–5177.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F. and López, M. (2007) Obstructive sleep apnea detection using clustering classification of nonlinear features from nocturnal oximetry. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, August 2007*. Lyon: IEEE, pp. 1937–1940.

B.3. Publications in national conferences

- ▣ Álvarez, D., Hornero, R., Gutiérrez, G.C., Marcos, J.V. and Del Campo, F. (2010) Análisis de las variaciones en la saturación de oxígeno y flujo aéreo

- en la ayuda al diagnóstico de la apnea del sueño. In: *Actas del XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2010, Madrid, November 2010*. Madrid: CASEIB, pp. 105.
- ▣ Arroyo, C.A., Del Campo, F., Hornero, R., De Frutos, J., Ruiz, T., Álvarez, D., Marcos, J.V., Sánchez, A. and González, M. (2010) Análisis no lineal de la actividad electroencefalográfica en pacientes con trastornos respiratorios del sueño. In: *Libro de actas del XXIX Congreso de la Sociedad Castellano-Leonesa y Cántabra de Patología Respiratoria (SOCALPAR), Santander, May 2010*. Santander: SOCALPAR.
- ▣ Marcos, J.V. Hornero, R., Álvarez, D., Del Campo, F. and Zamarrón, C. (2009) Aplicación de redes neuronales bayesianas para la ayuda en el diagnóstico de la apnea del sueño a partir de la oximetría. In: *Actas del XXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2009, Cádiz, November 2009*. Cádiz: CASEIB, pp. 537-540.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F. and López, M. (2009) Análisis espectral y no lineal de las señales de oximetría y electroencefalograma para la detección del síndrome de la apnea obstructiva del sueño. In: *Actas del XXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2009, Cádiz, November 2009*. Cádiz: CASEIB, pp. 533–536.
- ▣ Del Campo, F., Zamarrón, C., Hornero, R., Álvarez, D. and Marcos, J.V. (2009) Estudio de la sincronización entre la frecuencia cardiaca y la saturación de oxihemoglobina en pacientes con trastornos respiratorios del sueño. In: *Actas del XLII Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR), Santander, June 2009*. Santander: SEPAR, pp. 235.
- ▣ Del Campo, F., Hornero, R., Zamarrón, C., Álvarez, D. and Marcos, J.V. (2009) Utilidad de una red neuronal en el diagnóstico del síndrome de apnea hipopnea del sueño. In: *Libro de actas del XXVIII Congreso de la Sociedad*

Castellano Leonesa y Cántabra de Patología Respiratoria (SOCALPAR), Zamora, May 2009. Zamora: SOCALPAR.

- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M. (2008) Combinación de una red perceptrón multicapa y una red de funciones de base radial para la ayuda en la detección de la apnea obstructiva del sueño a partir de la oximetría nocturna. In: *Libro de Actas del XXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2008, Valladolid, October 2008*. Valladolid: CASEIB, pp. 261–264.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F., Zamarrón, C. and López, M. (2008) Empleo de características espectrales y no lineales en la detección del síndrome de la apnea obstructiva del sueño mediante análisis discriminante lineal. In: *Libro de Actas del XXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2008, Valladolid, October 2008*. Valladolid: CASEIB, pp. 277–280.
- ▣ Del Campo, F., De Frutos, J., Hornero, R., Álvarez, D. and Marcos, J.V. (2008) Aplicación de una red neuronal a las señales de saturación de oxígeno y tiempo de tránsito de pulso como ayuda diagnóstica en el síndrome de apnea del sueño. In: *Actas del XLI Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR), Santa Cruz de Tenerife, June 2008*. Santa Cruz de Tenerife: SEPAR, pp. 177–178.
- ▣ Del Campo, F., Zamarrón, C., Hornero, R., Marcos, J.V. and Álvarez, D. (2007) Utilidad de una red neuronal artificial aplicada a la señal de saturación de oxihemoglobina en el diagnóstico del síndrome de apnea hipopnea del sueño. In: *Actas del XL Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR), Barcelona, June 2007*. Barcelona: SEPAR, pp. 198–199.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M. (2007) Evaluación de un clasificador perceptrón multicapa para

la ayuda en el diagnóstico de la apnea obstructiva del sueño a partir de parámetros espectrales de la oximetría nocturna. In: *Libro de Actas del XXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2007, Cartagena, November 2007*. Cartagena: CASEIB, pp. 164–167.

- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F., Zamarrón, C. and López, M. (2007) Clasificación de características no lineales extraídas de los registros de saturación de oxígeno en sangre para la ayuda en el diagnóstico de la apnea obstructiva del sueño. In: *Libro de Actas del XXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2007, Cartagena, November 2007*. Cartagena: CASEIB, pp. 511–514.
- ▣ Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M. (2006) Análisis de la saturación de oxígeno en sangre mediante redes neuronales para la ayuda en el diagnóstico de la apnea del sueño. In: *Libro de Actas del XXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2006, Pamplona, November 2006*. Pamplona: CASEIB, pp. 387–390.
- ▣ Álvarez, D., Hornero, R., Marcos, J.V., Del Campo, F., Zamarrón, C. and López, M. (2006) Análisis de la entropía de los registros de saturación de oxígeno en sangre en múltiples escalas para la ayuda en el diagnóstico del síndrome de la apnea obstructiva del sueño. In: *Libro de Actas del XXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB) 2006, Pamplona, November 2006*. Pamplona: CASEIB, pp. 427–430.

Appendix C. Síntesis de la Tesis Doctoral en castellano

- C.1. Introducción
- C.2. El síndrome de la apnea obstructiva del sueño
- C.3. Estado de la técnica
- C.4. Sujetos y señales
- C.5. Extracción de características, normalización y reducción de dimensionalidad
- C.6. Técnicas de reconocimiento de patrones para clasificación
- C.7. Técnicas de reconocimiento de patrones para regresión
- C.8. Resultados
- C.9. Discusión y conclusiones
 - C.9.1. Conclusiones
 - C.9.2. Contribuciones originales
 - C.9.3. Líneas futuras

C.1. Introducción

El estudio presentado en la Tesis Doctoral pertenece al ámbito de la Ingeniería Biomédica. Esta disciplina emplea métodos y principios de la ingeniería para abordar problemas propios de la medicina o la biología (Nebeker, 2002). Concretamente, la Tesis se centra en el procesado de señales biomédicas con fines diagnósticos, haciendo especial hincapié en la utilización de técnicas de reconocimiento de patrones.

La Tesis propone un nuevo método para el diagnóstico del síndrome de la apnea-hipopnea del sueño (SAHS). Actualmente, sólo es posible obtener un diagnóstico fiable para esta patología mediante polisomnografía nocturna (PSG), que consiste en el estudio detallado del sueño del paciente durante una noche completa. Del mismo se deriva el valor del índice de apnea-hipopnea (AHI) del paciente, que indica el número de eventos apneicos observados por hora de sueño (Qureshi and Ballard 2003). Este parámetro se emplea como criterio para establecer el diagnóstico definitivo. Generalmente, un AHI igual o mayor que 10 h^{-1} conlleva un diagnóstico positivo de SAHS (Bloch, 2003).

Sin embargo, la PSG presenta varios inconvenientes. Este tipo de estudios debe realizarse en una unidad del sueño especializada y bajo la supervisión de personal técnico. Este entorno resulta extraño para el paciente, lo que puede alterar su patrón de sueño habitual. Además, el número de unidades de sueño disponibles en los centros hospitalarios es insuficiente para cubrir la creciente demanda de estudios de PSG (Flemons *et al.* 2004). Por otro lado, la PSG presenta una complejidad elevada, ya que durante la prueba se registran múltiples parámetros fisiológicos. Finalmente, debe tenerse en cuenta el alto coste económico asociado a la PSG, pues se precisa de equipamiento y personal específico para su realización (Bennet and Kinnear 1999).

En la Tesis se propone emplear únicamente la señal de saturación de oxígeno en sangre (SaO_2) registrada mediante pulsioximetría para obtener un diagnóstico final del SAHS. Esta señal forma parte de los registros monitorizados durante la PSG y constituye una pieza fundamental para la interpretación de la prueba. La pulsioximetría permite registrar la SaO_2 de forma no invasiva. Los costes asociados

a esta técnica son muy inferiores a los de la PSG y, además, podría realizarse en el hogar del paciente (Bloch, 2003; Netzer *et al.* 2003). Por tanto, la obtención de un diagnóstico a partir de la pulsioximetría nocturna eliminaría los inconvenientes derivados de la PSG. Tradicionalmente, la señal de SaO₂ ha sido utilizada para el diagnóstico del SAHS mediante la inspección visual y la utilización de índices oximétricos clásicos. Sin embargo, estos procedimientos no proporcionan la fiabilidad necesaria para el diagnóstico médico (Netzer *et al.* 2003). Por tanto, en la Tesis se plantea la siguiente hipótesis: la capacidad diagnóstica de los registros de SaO₂ podría incrementarse mediante la utilización de técnicas de procesamiento de señal avanzadas.

El objetivo de la Tesis es el desarrollo de métodos automáticos basados en técnicas de procesamiento de señal para la ayuda en el diagnóstico del SAHS a partir del registro de SaO₂. El comportamiento dinámico de las señales de SaO₂ en relación con la enfermedad del SAHS hace pensar que el análisis de estos registros, tanto en el dominio del tiempo como en el de la frecuencia, puede arrojar parámetros de gran utilidad para el diagnóstico. Además, sería posible combinar múltiples parámetros entre sí empleando técnicas de reconocimiento de patrones, que permitirían procesar una mayor cantidad de información simultáneamente.

Las técnicas de reconocimiento de patrones permiten llevar a cabo operaciones de clasificación y regresión (Bishop, 1995). En la Tesis se propone modelar el problema de diagnóstico del SAHS utilizando ambos procedimientos. El problema de clasificación consistiría en asignar una señal de SaO₂ a uno de dos posibles grupos: SAHS negativo o SAHS positivo. Por otro lado, el problema de regresión consistiría en obtener una estimación del AHI a partir de la información contenida en la señal de SaO₂. La metodología a aplicar en ambos casos es similar. Ésta se compone de tres etapas: 1) extracción de características, 2) normalización y reducción de la dimensionalidad, y 3) reconocimiento de patrones. En la última etapa se diseñarán e implementarán los algoritmos de reconocimiento de patrones para llevar a cabo la clasificación o regresión a partir de datos multivariante (patrones). Entre los primeros, se emplearon la técnica de *K*-vecinos próximos (KNN), el discriminante lineal de Fisher (FLD), la regresión logística (LR), las redes neuronales probabilísticas (PNN), las redes neuronales perceptrón multicapa (MLP), las redes de funciones de base radial (RBF), las redes Bayesianas (BY) y las máquinas de vectores soporte (SVM). Por otro lado, las

técnicas de regresión evaluadas fueron la regresión lineal múltiple (MLR), las redes de regresión generalizada (GRNN), las redes MLP, las redes RBF, las redes BY, los “*splines*” multivariante de regresión adaptativa (MARS) y los modelos SVM.

C.2. El síndrome de la apnea-hipopnea del sueño

El síndrome de la apnea-hipopnea del sueño (SAHS) es un trastorno derivado de la oclusión intermitente y repetitiva de la vía aérea superior durante el sueño (Qureshi and Ballard 2003). Los episodios de apnea originan alteraciones en el intercambio gaseoso intrapulmonar, así como la fragmentación del sueño, impidiendo un correcto descanso. Si no es tratado convenientemente, puede llegar a tener consecuencias severas sobre la calidad de vida del enfermo, causando desde trastornos neuropsiquiátricos como depresión, irritabilidad y deterioro intelectual, hasta complicaciones cardíacas y cerebrovasculares. Además, el SAHS es responsable indirecto de accidentes laborales o de tráfico debido a la somnolencia diurna excesiva (Van Houwelingen *et al.* 1999). La prevalencia del SAHS se estima entre el 1 y el 5% de los adultos (Young *et al.* 2002).

Actualmente, el método diagnóstico de referencia para el SAHS es la PSG, que debe realizarse en una unidad del sueño especializada (McNicholas, 2008). Una PSG completa registra típicamente el electroencefalograma (EEG), los movimientos oculares o electrooculogramas (EOG), el tono muscular mediante un electromiograma (EMG) mentoniano, el flujo aéreo buco-nasal mediante un termistor o una sonda de presión, el esfuerzo respiratorio mediante un cinturón torácico-abdominal, el electrocardiograma (ECG), el movimiento de las extremidades, ronquidos y somniloquias, y la SaO₂ mediante pulsioximetría digital (Qureshi and Ballard 2008; McNicholas, 2008). A pesar de su efectividad diagnóstica, la PSG es un método complejo y caro. Requiere la estancia del paciente en un entorno diferente al habitual, lo que puede afectar a las características habituales del sueño, de forma que el estudio no sea representativo de su desarrollo normal (Bennet and Kinnear 1999).

Los inconvenientes descritos han originado la búsqueda de alternativas diagnósticas más sencillas. Debido a su sencillez y relativo bajo coste, la oximetría nocturna, que proporciona una lectura permanente de la SaO₂, es una técnica de especial

interés para el estudio del sueño. La monitorización de la SaO₂ mediante la pulsioximetría es un procedimiento no invasivo. Generalmente, basta con colocar una pequeña sonda en el dedo del paciente. El principio de funcionamiento de los sensores de saturación se basa en las propiedades ópticas de las moléculas de hemoglobina (Jubran, 2004). Además, el equipo electrónico posee unas dimensiones reducidas, por lo que el estudio podría realizarse en el hogar del paciente. Estas características hacen de la pulsioximetría una línea de investigación de interés para el desarrollo de métodos de diagnóstico del SAHS alternativos a la PSG.

Una vez que el diagnóstico de SAHS ha sido confirmado, es preciso llevar a cabo un tratamiento efectivo de la enfermedad con el objetivo de prevenir posibles complicaciones. En primer lugar, es necesario que el paciente adopte hábitos saludables, evitando la obesidad, la falta de sueño y el consumo de alcohol o tabaco (Qureshi and Ballard 2003). La presión positiva en las vías respiratorias (CPAP) es el tratamiento más frecuente para el SAHS en adultos (Caples *et al.* 2005). Otras medidas posibles son la cirugía o la utilización de dispositivos cuyo cometido es corregir la posición de la mandíbula, favoreciendo la respiración (Van Houwelingen *et al.* 1999).

C.3. Estado de la técnica

Las técnicas de procesamiento de señales e inteligencia artificial han sido ampliamente utilizadas para el desarrollo de sistemas clínicos de decisión (Lisboa, 2002). Concretamente, pueden encontrarse ejemplos relativos a su aplicación en la medicina del sueño y como elementos de ayuda al diagnóstico del SAHS (Penzel and Conradt 2000). Esta patología influye sobre diferentes sistemas fisiológicos, de forma que las apneas producidas durante la noche quedan reflejadas en la actividad cerebral, cardíaca o respiratoria. El movimiento de las extremidades o el ronquido también pueden reflejar estos eventos. Además, ciertas variables clínicas y demográficas como el índice de masa corporal (BMI), la edad, la somnolencia diurna o la longitud de la circunferencia del cuello pueden ser indicativas del SAHS. Por tanto, existen diferentes fuentes de información que pueden ser usadas para el desarrollo de métodos automáticos de diagnóstico, tal y como se propone en la Tesis.

Concretamente, otros investigadores han estudiado exhaustivamente la utilidad de la señal de SaO_2 como herramienta de diagnóstico del SAHS. Los índices oximétricos clásicos como el índice de desaturación por encima del 3% (ODI3) o el 4% (ODI4), el tiempo acumulado con un nivel de saturación inferior al 90% (CT90), el índice Δ y el índice de severidad de las desaturaciones en referencia al 90% (SIT90) representan el método tradicional para analizar automáticamente los registros de SaO_2 . Sin embargo, existe gran discrepancia entre los estudios publicados hasta la fecha en cuanto a la sensibilidad y especificidad diagnóstica de estos índices.

Otras señales comúnmente empleadas para llevar a cabo el diagnóstico del SAHS son el ECG (de Chazal *et al.* 2003; Heneghan *et al.* 2008b) y el flujo respiratorio (Caseiro *et al.* 2010). Además, existen varios estudios que proponen diferentes reglas de diagnóstico basadas en parámetros clínicos (Friedman *et al.* 2010) o en la combinación de éstos con índices oximétricos o medidas obtenidas a partir del ECG (Roche *et al.* 2002).

Tal y como se propone en la Tesis, se han utilizado técnicas de reconocimiento de patrones para el análisis simultáneo de varios parámetros, representando cada uno de ellos una variable relacionada con el SAHS. Generalmente, el análisis discriminante y la regresión logística han sido empleados para este fin. Además, algunos trabajos emplean técnicas más avanzadas como son las redes MLP (El-Solh *et al.* 1999) y SVM (Khandoker *et al.* 2009a; Khandoker *et al.* 2009b) con variables clínicas y parámetros obtenidos del ECG, respectivamente.

C.4. Sujetos y señales

Un total de 240 sujetos sospechosos de padecer SAHS participaron en el estudio. El análisis del sueño se efectuó desde la medianoche hasta las 8:00 AM en el Hospital Río Hortega de Valladolid, España. La pulsioximetría nocturna fue realizada simultáneamente con la PSG. Las señales de SaO_2 fueron registradas a una frecuencia de 1 Hz. Los registros obtenidos mediante PSG fueron analizados por los expertos médicos para determinar el diagnóstico de cada sujeto, de forma que un AHI superior a 10 h^{-1} fue el umbral considerado para dictaminar un diagnóstico positivo de la enfermedad.

Las señales de SaO_2 fueron almacenadas en un equipo informático. Se aplicó un algoritmo de preprocesado sobre éstas para eliminar los artefactos originados durante la noche. El conjunto inicial de señales fue dividido en un grupo de entrenamiento (96 registros) y otro de test (144 registros) para el desarrollo y evaluación de los algoritmos de clasificación y regresión.

C.5. Extracción de características, normalización y reducción de dimensionalidad

Las dos primeras etapas de la metodología propuesta en la Tesis comprenden la extracción de características y las operaciones de normalización y reducción de dimensionalidad.

En la etapa de extracción de características se pretende cuantificar las propiedades de los registros de SaO_2 relacionadas con la presencia de SAHS y su grado de severidad. El objetivo es resumir la información de estos registros en un grupo reducido de variables, facilitando así su análisis. Las señales de SaO_2 de enfermos de SAHS suelen presentar un comportamiento inestable debido a las apneas producidas durante la noche. Éstas quedan reflejadas mediante profundas desaturaciones y la posterior restauración del valor de saturación. Por el contrario, el registro de oximetría tiende a permanecer constante, en torno al 96%, en sujetos de control (Netzer *et al.* 2001). Estas diferencias en el comportamiento dinámico de la señal de SaO_2 pueden utilizarse en la detección del SAHS y la estimación del AHI. Con el fin de reflejar las propiedades dinámicas de la señal de oximetría, se propone utilizar un total de 14 características calculadas mediante el análisis de los registros en el dominio del tiempo y la frecuencia.

Las características extraídas a través del análisis temporal de la oximetría fueron las siguientes:

- Análisis estadístico de la señal de SaO_2 en el dominio del tiempo. Se obtuvo una estimación para la media, varianza, asimetría y curtosis (Jobson, 1991b) de la variable que representa el valor de saturación de oxígeno.

- Análisis no lineal de la señal de SaO_2 en el dominio del tiempo. Se estimaron la entropía aproximada ($ApEn$), medida de la tendencia central (CTM) y complejidad de Lempel-Ziv (LZC). Estos parámetros representan una medida de la irregularidad (Pincus, 2001), la variabilidad (Cohen *et al.* 1996) y la complejidad de la señal (Lempel and Ziv 1976), respectivamente.

Además, se calcularon las siguientes características a partir de la densidad espectral de potencia (PSD) de la señal de oximetría:

- Análisis estadístico de la PSD del registro de SaO_2 . Se calcularon la media, la varianza, la asimetría y la curtosis (Jobson, 1991a) de la variable que representa la componente frecuencial de la señal.
- Análisis de la PSD del registro de SaO_2 en la banda comprendida entre 0.010 y 0.033 Hz (Zamarrón *et al.* 2003). En estudios anteriores se demostró que el espectro de señales procedentes de sujetos enfermos tiende a presentar mayor potencia en esta banda como consecuencia de la repetición de las apneas y su duración (Zamarrón *et al.* 2003). Por tanto, se calcularon la potencia total de la señal (S_T), la potencia en la banda de interés (S_B) y el valor de pico de la PSD en dicha banda frecuencial (PA).

Las características extraídas fueron normalizadas de forma que la distribución de cada una de ellas tuviera media nula y desviación típica igual a la unidad. Esta operación permite igualar el rango dinámico de cada característica, propiedad que es recomendable para implementar eficientemente los algoritmos de reconocimiento de patrones (Bishop, 1995). Por último, se evaluó el impacto producido por la reducción de dimensionalidad del patrón de entrada en los resultados de clasificación y regresión. La técnica empleada para llevar a cabo dicha operación fue el análisis de componentes principales (PCA).

C.6. Técnicas de reconocimiento de patrones para clasificación

En total se evaluaron 8 técnicas diferentes para la clasificación de patrones: KNN, FLD, LR, redes PNN, redes MLP, redes RBF, redes BY y SVM. Éstas incluyen procedimientos simples y comúnmente empleados, como pueden ser KNN, FLD, LR o las redes PNN, junto con otros métodos más complejos, como son las redes MLP, RBF, BY y SVM. FLD y LR únicamente son capaces de definir fronteras de decisión lineal en el espacio de entrada. En cambio, el resto de técnicas permiten resolver problemas de clasificación en los que la frontera de decisión entre ambas categorías es no lineal. En principio, las redes MLP, RBF y BY así como los algoritmos basados en SVM poseen un mayor potencial para modelar las propiedades estadísticas del problema.

C.7. Técnicas de reconocimiento de patrones para regresión

Se evaluaron 7 técnicas diferentes para modelar el problema de regresión propuesto en la Tesis: MLR, redes GRNN, redes MLP, redes RBF, redes BY, MARS y SVM. Al igual que sucede con la clasificación, el objetivo fue utilizar técnicas con diferente grado de complejidad para analizar su comportamiento ante el problema propuesto. En este caso, MLR representa el método más común para realizar tareas de regresión a partir de patrones multivariante. La solución asumida por MLR propone aproximar la variable objetivo mediante la combinación lineal de las variables de entrada. Las otras técnicas propuestas poseen la capacidad de definir dependencias no lineales entre el conjunto de características de entrada y la variable de salida.

C.8. Resultados

Dado que en las tareas de clasificación y regresión se persiguen objetivos diferentes, se utilizaron sendas medidas del rendimiento para evaluar ambos tipos de algoritmos. La precisión, que indica la tasa de acierto en el diagnóstico, fue utilizada para comparar la capacidad de los algoritmos de clasificación. En el caso de los algoritmos de regresión, se utilizó el coeficiente de correlación intraclass (*ICC*), que cuantifica la fiabilidad de la estimación (Shrout and Fleiss 1979).

Antes de diseñar los diferentes algoritmos de clasificación y regresión, se evaluó el rendimiento individual de cada una de las características extraídas en la resolución de los dos problemas propuestos. El parámetro espectral S_B proporcionó el valor más alto de precisión en la clasificación de señales: 91.67% en el conjunto de test. En cuanto a la aproximación del AHI, el *CTM* logró el *ICC* más alto con un valor de 0.77.

PCA permitió reducir el número de características a 5. El criterio para la selección de componentes consistió en establecer un umbral del 90% sobre la varianza total (Nabney, 2002).

Posteriormente, se implementaron los algoritmos de clasificación y regresión utilizando los métodos de reconocimiento de patrones propuestos. En total, se construyeron 16 algoritmos de clasificación, dos por cada una de las 8 técnicas seleccionadas para modelar el problema. En la primera versión del algoritmo, se tomaron como variables de entrada las 14 características extraídas de la SaO_2 una vez normalizadas. En la segunda, se aplicó PCA sobre las características normalizadas. Posteriormente, las componentes seleccionadas se utilizaron como variables de entrada. El objetivo fue analizar el efecto de la reducción de dimensionalidad mediante PCA.

Por otro lado, se implementaron 14 algoritmos de regresión para la estimación del AHI. Tal y como se propuso para el problema de clasificación, se utilizaron dos patrones diferentes por cada técnica: el conjunto total de características normalizadas y el conjunto de componentes obtenidas mediante PCA.

El algoritmo de clasificación que logró la mayor precisión diagnóstica fue aquél constituido por una red MLP y el conjunto total de 14 características normalizadas, con una tasa de acierto de 92.36% en el conjunto de test. En el problema de regresión, varios algoritmos alcanzaron un *ICC* superior a 0.9. Finalmente, la red BY con PCA previo y la utilización de un modelo MARS con el conjunto total de 14 características fueron los procedimientos más destacados.

Por último, se evaluó la capacidad diagnóstica de los índices oximétricos convencionales sobre el conjunto de señales de SaO_2 analizado en la Tesis. El objetivo

de estudio fue comparar la efectividad de los algoritmos propuestos frente a los parámetros utilizados tradicionalmente para el procesado automático de la oximetría. El índice Δ proporcionó los mejores resultados para el problema de clasificación, con una precisión de 88.19%. El valor de *ICC* más alto fue logrado mediante el ODI3, con un índice de correlación de 0.91. Sin embargo, se observó que el error en la estimación utilizando este índice era notable en el caso de sujetos con AHI inferior a 10 h^{-1} .

C.9. Discusión y conclusiones

El esquema propuesto para el procesado de señales de SaO_2 se compone de tres etapas: extracción de características, normalización y reducción de dimensionalidad, y reconocimiento de patrones. En la primera etapa, se ha definido un conjunto de características para cuantificar diferentes propiedades de la señal de SaO_2 en el dominio del tiempo y de la frecuencia. Como características temporales, se han propuesto los cuatro primeros momentos estadísticos y métodos no lineales como son *ApEn*, *CTM* y *LZC*. Estos últimos fueron aplicados previamente en otros estudios para diferenciar señales de SaO_2 correspondientes a enfermos de SAHS y sujetos de control, proporcionando valores de precisión destacables (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Hornero *et al.* 2007). Los resultados obtenidos en la Tesis para *CTM* y *LZC* concuerdan con los publicados en los estudios anteriores, ya que las señales analizadas tienden a ser más variables y complejas en el caso de pacientes con SAHS. En cambio, no se ha apreciado ninguna tendencia en el valor de *ApEn* (irregularidad), ya que es similar para sujetos enfermos y sanos. Por otro lado, el análisis en frecuencia de las señales proporcionó los resultados esperados, ya que pudo observarse como los registros asociados a un mayor valor de AHI tienden a presentar más potencia en la banda entre 0.010 y 0.033 Hz, tal y como se adelantaba en estudios anteriores (Zamarrón *et al.* 2003).

En la siguiente etapa, se normalizaron las características extraídas para homogeneizar su rango dinámico. Asimismo, se evaluó la utilidad de reducir la dimensionalidad del patrón de características mediante PCA. Los resultados reflejan que esta técnica no proporciona ninguna ventaja para la clasificación o la regresión, ya que se pierde información relevante en las componentes descartadas. Por tanto, PCA no representa un procedimiento óptimo para la reducción de dimensionalidad.

En la tercera etapa se emplearon técnicas de reconocimiento de patrones tanto para la clasificación de señales como para la estimación del AHI. El trabajo presentado en esta Tesis muestra la necesidad de llevar a cabo un diseño detallado de estas técnicas para optimizar su rendimiento. Este proceso de diseño consiste en encontrar el equilibrio óptimo entre sesgo y varianza para el conjunto de entrenamiento dado. El sesgo y la varianza varían con la complejidad del modelo, que viene dada por el valor de diferentes parámetros de diseño. Por tanto, es necesario comparar el rendimiento de diferentes configuraciones (diferentes valores de los parámetros de diseño) para identificar la más adecuada para el problema propuesto. En esta Tesis se ha utilizado un procedimiento de validación cruzada dejando uno fuera a partir de los datos en el conjunto de entrenamiento para llevar a cabo la selección del modelo. Este procedimiento permite reservar los datos en el conjunto de test para la validación de los algoritmos finalmente seleccionados. Sin embargo, mediante la validación cruzada se comparan únicamente configuraciones. Así, el algoritmo de clasificación o regresión que finalmente es implementado y validado se obtiene a partir del entrenamiento de un nuevo modelo con la configuración seleccionada previamente.

Los resultados logrados en esta Tesis muestran la validez de la metodología propuesta. En la clasificación, se llegó a alcanzar una precisión de 92.36% mediante una red MLP. Los algoritmos de regresión con mejores prestaciones, basados en MARS y redes BY, proporcionaron un *ICC* superior a 0.90. Estos algoritmos mejoran los resultados de clasificación y regresión obtenidos con cada característica de entrada por separado. No obstante, el análisis espectral de la SaO₂ permite alcanzar una precisión de 91.67% en la clasificación de señales de oximetría. Por tanto, la utilización de técnicas de reconocimiento de patrones no supone una ventaja considerable para la clasificación, ya que con un único parámetro de entrada es posible alcanzar niveles de precisión similares. En cambio, la utilización de técnicas de reconocimiento de patrones para el problema de regresión permite mejorar sustancialmente los resultados de estimación de cada parámetro por separado. Por otro lado, los algoritmos de reconocimiento de patrones superan notablemente el rendimiento de los índices oximétricos clásicos. Éstos fueron evaluados individualmente para las tareas de clasificación y regresión utilizando la base de datos de esta Tesis.

Finalmente, ha de puntualizarse que la comparación de los resultados de esta Tesis con los alcanzados previamente por otros investigadores estaría totalmente sesgada, ya que la base de datos utilizada es diferente. Aún así, puede decirse que los resultados presentados en esta Tesis son comparables con el estado de la técnica, lo que demuestra el potencial de los métodos propuestos.

C.9.1. Conclusiones

La principal conclusión de la Tesis Doctoral es que el análisis de los registros de SaO₂ mediante los algoritmos de reconocimiento de patrones propuestos permite alcanzar una elevada precisión diagnóstica para la detección del SAHS. Estos algoritmos mejoran las prestaciones de otros métodos tradicionalmente usados para la interpretación de estos registros, como son la inspección visual y los índices oximétricos clásicos. Por tanto, podrían emplearse como un método automático de exploración para el diagnóstico del SAHS, contribuyendo así a reducir el número de PSG.

Además de esta conclusión global, los resultados del estudio conducen a las siguientes observaciones:

1. Las técnicas de procesamiento de señal permiten obtener información adicional de los registros de oximetría que puede ser usada en el diagnóstico del SAHS. En esta Tesis se ha propuesto modelar el problema de diagnóstico del SAHS mediante técnicas de reconocimiento de patrones. Estos algoritmos posibilitan el análisis simultáneo de varias variables. Su principal ventaja es su capacidad para capturar las propiedades estadísticas de los datos de entrada, incluso en el caso de problemas de naturaleza no lineal. Los resultados logrados por los mejores algoritmos de clasificación y regresión implementados en la Tesis se recogen en la Tabla 26. A pesar de su elevada capacidad diagnóstica, sería necesario un nuevo proceso de validación de los mismos mediante una base de datos mayor para que fueran empleados en la práctica clínica.
2. La base de datos de señales de oximetría fue suficiente para llevar a cabo el estudio propuesto. Los resultados alcanzados en este trabajo constituyen una

referencia para futuros trabajos que empleen la misma base de datos. Además, estos resultados reflejan conclusiones importantes del estudio realizado. Aún así, hubiera sido deseable disponer de una base de datos más amplia. La utilización de métodos de reconocimiento de patrones de elevada complejidad requiere una descripción detallada de las propiedades estadísticas del problema a modelar. Para ello, es preciso contar con un conjunto de entrenamiento de gran tamaño.

3. Los resultados indican que la solución óptima tanto para el problema de clasificación (frontera de decisión en el espacio de entrada) como para la regresión (relación funcional entre el AHI y las características de entrada) es de naturaleza no lineal. Se ha realizado un estudio detallado de las técnicas de reconocimiento de patrones más relevantes, incluyendo métodos lineales comúnmente empleados y técnicas complejas capaces de modelar problemas de tipo no lineal. En total, se han implementado 16 algoritmos de clasificación y 14 algoritmos de regresión, algunos de los cuales han mostrado un rendimiento destacable. Los resultados reflejan que los métodos no lineales ofrecieron mejor rendimiento tanto para la clasificación como para la regresión. Por tanto, puede decirse que la solución óptima para ambos problemas es de naturaleza no lineal. Sin embargo, sería necesario comprobar este resultado mediante la utilización de una base de datos más amplia.
4. Es necesario diseñar y comparar varios algoritmos de reconocimiento de patrones para lograr optimizar los resultados ante un problema concreto. Dado el conjunto de entrenamiento, es necesario ajustar uno o más parámetros de diseño para adaptar el nivel de complejidad de un algoritmo al problema propuesto. Para ello, se debe analizar la influencia de cada uno de estos parámetros de diseño en el rendimiento del algoritmo durante la fase de selección del modelo. Los métodos con un mayor potencial (flexibilidad) para modelar un problema poseen un mayor número de parámetros de diseño. Por tanto, existe un compromiso entre la dificultad de la etapa de diseño y la capacidad del método. En principio, no es posible saber qué método es el más adecuado para resolver un determinado

problema, por lo que es necesario realizar una comparación entre el mayor número de técnicas posible para optimizar la solución. Ésta ha sido precisamente la estrategia adoptada en la Tesis Doctoral.

5. El *CTM* fue la característica no lineal con mayor capacidad diagnóstica. Previamente, otros estudios hallaron diferencias significativas entre poblaciones de SAHS negativo y positivo para las variables *ApEn*, *CTM* y *LZC* calculadas a partir de la SaO_2 (Álvarez *et al.* 2006; Del Campo *et al.* 2006; Hornero *et al.* 2007). Sin embargo, los resultados de la Tesis muestran que únicamente el *CTM*, que cuantifica la variabilidad de la señal, posee una capacidad diagnóstica elevada por sí solo. Por el contrario, la irregularidad y la complejidad de la señal, medidas a través de la *ApEn* y el *LZC*, respectivamente, reflejaron una capacidad diagnóstica muy inferior. Este hecho es especialmente destacable en el caso de la *ApEn*, ya que los valores obtenidos indican un nivel de irregularidad similar en señales procedentes de ambas poblaciones.
6. El análisis espectral de los registros de oximetría refleja diferencias significativas entre los sujetos SAHS negativo y SAHS positivo. Zamarrón *et al.* (2003) observaron cómo las señales correspondientes a sujetos enfermos poseen mayor potencia en la banda entre 0.010 y 0.033 Hz como consecuencia de los periodos con repetidas apneas. Este resultado fue corroborado por El-Solh *et al.* (2003). Los resultados de la Tesis son consistentes con estos estudios. Así, las tres características obtenidas directamente de la PSD (S_T , S_B y PA) proporcionaron una elevada precisión diagnóstica. Por tanto, puede decirse que el análisis espectral es una herramienta más eficaz que el análisis no lineal para evaluar el comportamiento dinámico de la oximetría nocturna.
7. Los algoritmos de reconocimiento de patrones superaron a los índices oximétricos clásicos en la clasificación de señales de SaO_2 como SAHS negativo o positivo. El clasificador MLP con el conjunto total de características logró una precisión del 92.36%, lo que supone el valor más alto entre todos los algoritmos de clasificación implementados. Este

algoritmo mejoró las prestaciones de los índices oximétricos clásicos en el problema de clasificación propuesto. Por tanto, puede decirse que el método desarrollado en la Tesis es un procedimiento válido para la clasificación automática de señales de SaO₂.

8. La estrategia más eficiente para el problema de clasificación es el análisis espectral de las señales de SaO₂. A pesar de que la precisión más alta fue obtenida mediante el algoritmo de clasificación basado en redes MLP mencionado anteriormente (92.36%), puede obtenerse una precisión diagnóstica similar mediante el parámetro S_B (potencia de la señal en la banda entre 0.010 y 0.033 Hz) calculado a partir de la PSD de la señal. Concretamente, este parámetro permite obtener una precisión del 91.67%. En este caso, la clasificación se realiza mediante la comparación del valor de S_B con el umbral óptimo determinado mediante el análisis de la curva de característica de receptor (ROC) sobre los datos del conjunto de entrenamiento. Por tanto, el análisis multivariante no representa una ventaja sustancial para el problema de clasificación, ya que una única característica de entrada es suficiente para lograr una precisión similar. Además, la complejidad del método en este caso es notablemente inferior, por lo que puede decirse que el análisis espectral es la solución óptima para el problema de clasificación propuesto.
9. La utilización de técnicas de reconocimiento de patrones representa una ventaja para el problema de regresión propuesto. La evaluación del rendimiento de cada característica de entrada en esta tarea se realizó a partir de un modelo de regresión lineal. Así, el modelo construido a partir del *CTM* proporcionó el valor más alto de *ICC* (0.77). Varios algoritmos de reconocimiento de patrones lograron un *ICC* superior a 0.90, lo que supone una mejora considerable respecto a las características de entrada. Éstos fueron los basados en MARS y redes BY para el procesado del conjunto total de características y del patrón derivado de PCA, respectivamente. La precisión diagnóstica de ambos algoritmos fue del 87.50%, obtenida a partir del valor de AHI estimado. Asimismo, estos algoritmos fueron comparados con los índices oximétricos clásicos. En este caso, el índice con mejor

rendimiento fue el ODI3 ($ICC = 0.91$). Sin embargo, se comprobó que la aproximación al AHI obtenida a partir de este parámetro tendía a sobreestimar el verdadero AHI en el caso de sujetos de control o con SAHS leve, anulando la capacidad diagnóstica del modelo. Por tanto, la solución óptima para el problema de regresión son los algoritmos de reconocimiento de patrones mencionados anteriormente.

10. La utilización de PCA no contribuyó a mejorar el rendimiento de la mayoría de los métodos evaluados. Se propuso PCA como una medida para reducir el sobreaprendizaje gracias a la reducción de la dimensionalidad. Sin embargo, los resultados reflejan que las componentes descartadas poseen información relevante para el problema. PCA es un método no supervisado, por lo que no se tiene en cuenta el valor de la variable objetivo para la selección de componentes, que son ordenadas en función del porcentaje de varianza total explicado por cada una de ellas. Por tanto, el subconjunto de componentes seleccionado puede no ser el óptimo para llevar a cabo las tareas de clasificación y regresión propuestas.
11. El modelo de regresión para el diagnóstico del SAHS proporciona información más útil que el de clasificación. Como se ha comentado, se han utilizado dos modelos para el diagnóstico del SAHS: clasificación y regresión. La mayor precisión diagnóstica ha sido lograda por uno de los algoritmos de clasificación implementados en la Tesis. Sin embargo, el modelo de regresión es preferible ya que permite identificar el grado de severidad de la enfermedad. Además, es independiente de la definición utilizada para el SAHS (umbral para el valor de AHI) ya que los algoritmos no son entrenados para proporcionar una decisión de diagnóstico, sino el valor del AHI. Los algoritmos de regresión han sido evaluados mediante el ICC obtenido entre el AHI verdadero y el aproximado, proporcionando una medida de la fiabilidad en la estimación. Este parámetro refleja con mayor fidelidad la calidad del estimador que el coeficiente de correlación de Pearson empleado comúnmente, ya que éste sólo indica si existe una relación lineal entre las dos variables. Además, se han empleado medidas de error y gráficos de Bland-Altman para evaluar estos algoritmos.

12. La mayoría de los sujetos diagnosticados incorrectamente por los algoritmos sufren SAHS leve. El algoritmo de clasificación con la precisión más alta (basado en redes MLP) falló en el diagnóstico de 11 sujetos. De ellos, 7 estaban afectados por SAHS leve ($5 \text{ h}^{-1} \leq \text{AHI} \leq 10 \text{ h}^{-1}$). Por otro lado, el algoritmo de regresión más destacado diagnosticó erróneamente a 18 sujetos, 14 de los cuales tenían SAHS leve. Estos resultados reflejan un comportamiento coherente de los algoritmos, ya que cabe esperar que los errores se produzcan para sujetos de estas características. Hay que considerar que los algoritmos propuestos fueron desarrollados para ser aplicados en señales registradas mediante equipos portátiles. En este sentido, la Academia Americana de la Medicina del Sueño indica que estos equipos deberían emplearse como una alternativa a la PSG únicamente en pacientes con un alto riesgo de padecer SAHS. Por tanto, las técnicas implementadas en la Tesis pueden considerarse como una herramienta válida para el análisis de pacientes potenciales de SAHS mediante dispositivos portátiles. Esta práctica contribuiría a reducir la demanda de estudios de PSG.
13. Los resultados de clasificación y regresión muestran el alto potencial de las redes MLP para la resolución de problemas de reconocimiento de patrones. Estos algoritmos son los modelos de red neuronal más populares y han sido aplicados a diversos problemas. En la Tesis, los algoritmos de redes MLP proporcionaron mejores resultados que otros basados en redes RBF o SVM, a pesar de que éstos también tienen la capacidad de modelar problemas de naturaleza no lineal. Un aspecto importante para el diseño y la optimización de las redes MLP es la regularización mediante “*weight decay*” (Bishop, 1995), que permite controlar la complejidad de la red de acuerdo al conjunto de entrenamiento disponible.

C.9.2. Contribuciones originales

La principal contribución de la Tesis es el desarrollo de algoritmos de reconocimiento de patrones para el diagnóstico automático del SAHS a partir de la

oximetría nocturna. Algunas de las características de entrada de los algoritmos propuestos han sido empleadas en otros estudios previos. Sin embargo, éste es el primer trabajo en el que se propone combinar estas características mediante técnicas de reconocimiento de patrones para la ayuda en el diagnóstico del SAHS. La base de datos utilizada para la investigación se compuso de 240 señales de SaO₂ correspondientes a sujetos sospechosos de padecer la enfermedad. En comparación con los trabajos publicados hasta la fecha, los resultados alcanzados sobre la misma son destacables y muestran la validez de la metodología propuesta.

Algunos de los métodos de procesado de señal empleados no han sido utilizados anteriormente para el análisis de la oximetría nocturna. Una muestra objetiva de ello son las publicaciones derivadas de la Tesis que se detallan en el Apéndice B del documento. Las contribuciones originales de la Tesis así como las principales ventajas de los métodos propuestos en la misma se detallan a continuación:

1. En este estudio se ha desarrollado el método de preprocesado para la eliminación de artefactos en señales de SaO₂. La etapa de preprocesado es fundamental, ya que las señales suelen presentar descensos bruscos en su perfil como consecuencia de los movimientos del paciente. Además, es conveniente descartar los primeros minutos del registro, pues el paciente podría estar despierto durante los mismos. El algoritmo de preprocesado desarrollado en la Tesis no sólo es válido para las señales utilizadas en la misma, sino que podría aplicarse sobre señales de SaO₂ adquiridas con un equipo diferente.
2. Los parámetros estadísticos representan un nuevo conjunto de medidas para la caracterización de los registros de SaO₂. Anteriormente, sólo el valor medio de SaO₂ había sido utilizado para el desarrollo de métodos de diagnóstico (Herer *et al.* 1999). La varianza, la asimetría o la curtosis no se han empleado anteriormente para caracterizar la distribución de muestras de SaO₂. Por otro lado, otros estudios propusieron con anterioridad el análisis espectral de la señal de SaO₂, aunque las propiedades estadísticas de la componente frecuencial no fueron analizadas. De hecho, el trabajo de Marcos *et al.* (2010a) fue el primero en proponer el análisis estadístico

clásico de la señal tanto en el dominio del tiempo como en el de la frecuencia. Estas características proporcionan información relevante para la clasificación y regresión, como se deduce de los resultados obtenidos para las redes BY.

3. Esta Tesis es el primer estudio en el que se combinan características temporales y frecuenciales para llevar a cabo el diagnóstico del SAHS. Los resultados indican que el análisis simultáneo de ambos grupos de características permite obtener una descripción más detallada de las señales de SaO₂. Previamente, el análisis de patrones multivariante se utilizó como estrategia para la ayuda en la detección del SAHS utilizando parámetros de la oximetría nocturna (Roche *et al.* 2002; Magalang *et al.* 2003). Únicamente se emplearon índices oximétricos clásicos. Típicamente, éstos fueron combinados con variables clínicas y/o demográficas para obtener el diagnóstico final. Tal y como se propone en la Tesis, un estudio anterior propone combinar parámetros no lineales calculados en el dominio del tiempo con medidas espectrales, ambos obtenidos de la señal de magnetoencefalograma, para la ayuda en el diagnóstico de la enfermedad de Alzheimer (Hornero *et al.* 2008).
4. En la Tesis se comparan dos modelos diferentes para el problema de diagnóstico del SAHS: clasificación y regresión. Esta propuesta es de gran interés ya que permite apreciar las ventajas y desventajas de ambos modelos bajo las mismas circunstancias de aplicación. No se han encontrado trabajos previos en los que se proponga esta comparación en el contexto del diagnóstico del SAHS. En cambio, existen varios estudios que se centran bien en el modelo de clasificación (de Chazal *et al.* 2009; Fiedman *et al.* 2010; Khandoker *et al.* 2009b) o en el de regresión (El-Solh *et al.* 1999; Vázquez *et al.* 2000; Magalang *et al.* 2003).
5. Hasta donde conocemos, las redes RBF y BY no han sido empleadas previamente para la ayuda al diagnóstico del SAHS. Además de los métodos convencionales como son KNN, FLD, LR o MLR, otros investigadores han aplicado técnicas más complejas para el reconocimiento de patrones. Así, se

han utilizado redes MLP y GRNN para la aproximación del AHI (El-Solh *et al.* 1999; Kirby *et al.* 1999; Magalang *et al.* 2003). De la misma forma, se han implementado algoritmos de clasificación basados en redes PNN y SVM para la identificación de pacientes con SAHS (El-Solh *et al.* 2003; Khandoker *et al.* 2009a; Khandoker *et al.* 2009b). Sin embargo, no se han encontrado ejemplos de aplicación de las redes RBF y BY, para las que fueron publicados algunos de los resultados preliminares de esta Tesis (Marcos *et al.* 2008a; Marcos *et al.* 2010a).

6. La Tesis expuesta supone el primer estudio en el que se utilizan algoritmos SVM para la estimación del AHI. Anteriormente, se han desarrollado algoritmos de diagnóstico del SAHS basados en clasificadores SVM (Khandoker *et al.* 2009a; Khandoker *et al.* 2009b). Precisamente, el número de aplicaciones que proponen SVM para llevar a cabo tareas de regresión es muy inferior en comparación con su utilización como clasificador. Los resultados de la Tesis sirven como ejemplo del buen funcionamiento de los modelos SVM en problemas de regresión.
7. La Tesis analiza el rendimiento de diferentes técnicas de reconocimiento de patrones para hallar la solución óptima al problema propuesto. Como se indicó anteriormente, algunas de estas técnicas habían sido utilizadas previamente para la ayuda en el diagnóstico del SAHS. Sin embargo, este es el primer estudio en el que se realiza una comparación entre diferentes técnicas de reconocimiento de patrones. Se trata de una estrategia necesaria, ya que el diseñador no conoce previamente cuál es el método más adecuado para el problema bajo estudio.
8. El análisis de los resultados de estimación del AHI mediante el *ICC* representa un elemento diferenciador del estudio presentado. Se trata de un parámetro robusto para evaluar la fiabilidad de las estimaciones (Bravo and Potvin 1991). En otros estudios centrados en la aproximación del AHI se utilizó el coeficiente de correlación de Pearson para evaluar la precisión del estimador. Sin embargo, este parámetro sólo indica la relación lineal entre

dos variables y no detecta la posible existencia de un error sistemático en la estimación.

C.9.3. Líneas futuras

La Tesis presenta los resultados alcanzados mediante un número desatado de algoritmos de reconocimiento de patrones: 16 algoritmos para la clasificación de señales de SaO₂ y 14 algoritmos para la estimación del AHI. Sin embargo, existen diferentes aspectos que habrán de ser tratados en futuras investigaciones.

Los algoritmos propuestos deberían ser validados en una base de datos más amplia para confirmar los resultados logrados. Por tanto, sería preciso aumentar el número de señales en la base de datos actual. Además, las señales analizadas en la Tesis fueron registradas durante el estudio de PSG realizado en una unidad del sueño especializada. Dado que el objetivo es eliminar los inconvenientes derivados de la PSG, la finalidad de los algoritmos implementados debería ser el análisis de señales de SaO₂ adquiridas mediante dispositivos portátiles. Así, estos algoritmos deberán ser evaluados sobre señales recogidas en el hogar del paciente mediante equipos portátiles de pulsioximetría. De hecho, en próximos proyectos podría pensarse en una solución propia de la telemedicina consistente en transmitir los registros de SaO₂ desde el equipo de adquisición a un servidor central. Éste contendría la lógica necesaria para llevar a cabo el procesado de las señales y proporcionar un diagnóstico de forma automática.

Otro aspecto a tratar tras la Tesis sería la combinación de los algoritmos desarrollados. El objetivo sería mejorar los resultados logrados hasta el momento por cada uno de ellos individualmente. Además, podría considerarse la inclusión de nuevas características de entrada. En este sentido, los índices oximétricos clásicos podrían ser empleados junto con los parámetros temporales y frecuenciales propuestos en la Tesis. Sin embargo, la utilización de un patrón de entrada de grandes dimensiones podría afectar negativamente al rendimiento de ciertos modelos, como son, por ejemplo, las redes MLP o RBF. Sería necesario mejorar la etapa de reducción de dimensionalidad utilizando nuevos métodos para la selección de características.

Finalmente, podría pensarse en combinar la información extraída de los registros de oximetría con otras señales biomédicas relacionadas con el SAHS. Por ejemplo, el ECG o el flujo aéreo pueden proporcionar información complementaria de interés para el diagnóstico. Sin embargo, la inclusión de nuevas señales conllevaría un incremento sustancial de la complejidad del procedimiento de adquisición de datos.

Bibliography

- Abbot, R.D. and Carroll, R.J. (1984) Interpreting multiple logistic regression coefficients in prospective observational studies. *American Journal of Epidemiology*, 119, pp. 830–836.
- Abdullah, H., Maddage, N.C., Cosic, I. and Cvetkovic, D. (2010) Cross-correlation of EEG frequency bands and heart rate variability for sleep apnoea classification. *Medical & Biological Engineering & Computing*, 48, pp. 1261–1269.
- Acir, N., Öztura, I., Kuntalp, M., Baklan, B. and Güzelis, C. (2005) Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Transactions on Biomedical Engineering*, 52, pp. 30–40.
- Álvarez, D., Hornero, R., Abásolo, D., Del Campo, F. and Zamarrón, C. (2006) Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. *Physiological Measurement*, 27, 399–412.
- Álvarez, D., Hornero, R., García, M., Del Campo, F. and Zamarrón, C. (2007) Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure. *Artificial Intelligence in Medicine*, 41, pp. 13–24.
- Álvarez, D., Hornero, R., Marcos, J.V. and Del Campo, F. (2010) Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. *IEEE Transactions on Biomedical Engineering*, 57, pp. 2816–2824.

- Álvarez-Estévez, D. and Moret-Bonillo, V. (2009) Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome. *Expert Systems with Applications*, 36, pp. 7778–7785.
- Al-Abed, M., Manry, M., Burk, J.R., Lucas, E.A. and Behbehani, K. (2007) A method to detect obstructive sleep apnea using neural network classification of time-frequency plots of the heart rate variability. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, August 2007*. Lyon: IEEE, pp. 6101–6104.
- Al-Ani, T., Haman, Y., Fodil, R., Lofaso, F. and Isabey, D. (2004) Using hidden Markov models for sleep disordered breathing identification. *Simulation Modelling Practice and Theory*, 12, pp. 117–128.
- Amick, D.J. and Walberg, H.J. (1975) *Introductory multivariate analysis*. Berkley: McCutchan Publishing Corp..
- Andersson, L. and Brattström, V. (1991) Cephalometric analysis of permanently snoring patients with and without obstructive sleep apnea syndrome. *International Journal of Oral and Maxillofacial Surgery*, 20, pp. 159–162.
- Ayas, N.T., Pittman, S., MacDonald, M. and White, D.P. (2003) Assessment of a wrist-worn device in the detection of obstructive sleep apnea. *Sleep Medicine*, 4, pp. 435–442.
- Baltzan, M.A., Verschelden, P., Al-Jahdali, H., Olha, A.E. and Kimoff, R.J. (2000) Accuracy of oximetry with thermistor (OxiFlow) for diagnosis of obstructive sleep apnea and hypopnea. *Sleep*, 23, pp. 61–69.
- Bartlett, P.L. (1998) The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Transactions on Information Theory*, 44, pp. 525–536.

- Baxt, W.G. (1995) Application of artificial neural networks to clinical medicine. *Lancet*, 346, pp. 1135–1138.
- Bebis, G and Georgiopoulos, M. (1994) Feed-forward neural networks. *IEEE Potentials*, 13, pp. 27–31.
- Bennet, J.A. and Kinnear, W.J.M. (1999) Sleep on the cheap: the role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome. *Thorax*, 11, pp. 958–959.
- Bishop, C.M. (1995) *Neural Networks for pattern recognition*. Oxford: Oxford University Press.
- Bloch, K.E. (2003) Getting the most out of nocturnal pulse oximetry. *Chest*, 124, pp. 1628–1630.
- Bradley, A.P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30, pp. 1145–1159.
- Bradley, P.A., Mortimore, I.L. and Douglas, N.J. (1995) Comparison of polysomnography with ResCare Autoset in the diagnosis of the sleep apnoea/hypopnoea syndrome. *Thorax*, 50, pp. 1201–1203.
- Bradley, T.D. and Floras, J.S. (2009) Obstructive sleep apnoea and its cardiovascular consequences. *Lancet*, 373, pp. 82–93.
- Bravo, G. and Potvin, L. (1991) Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *Journal of Clinical Epidemiology*, 44, pp. 381–390.
- Breiman, L. (1996) Bagging predictors. *Machine Learning*, 24, pp. 123–140.
- Broomhead, D.S. and Lowe, D. (1988) Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, pp. 321–355.

- Bruce, E.N. (2001) *Biomedical signal processing and signal modelling*. New York: John Wiley and Sons.
- Burwell, C.S., Robin, E.D., Whaley, R.D. and Bickelmann, A.G. (1956) Extreme obesity associated with alveolar hypoventilation—A pickwickian syndrome. *The American Journal of Medicine*, 21, pp. 811–818.
- Bystricky, W. and Safer, A. (2004) Identification of individual apnea events from the ECG using neural networks and a dynamic markovian state model. *Computers in Cardiology*, 31, pp. 297–300.
- Caballero, P., Álvarez-Sala, R., García-Río, F., Prados, C., Hernán, M.A., Villamar, J. and Álvarez-Sala, J.L. (1998) CT in the evaluation of the upper airway in healthy subjects and in patients with obstructive sleep apnea syndrome. *Chest*, 113, pp. 111–116.
- Cabrero-Canosa, M., Castro-Pereiro, M., Graña-Ramos, M., Hernández-Pereira, E., Moret-Bonillo, V., Martín-Egaña, M. and Vereá-Hernando, H. (2003) An intelligent system for the detection and interpretation of sleep apneas. *Expert Systems with Applications*, 24, pp. 335–349.
- Cabrero-Canosa, M., Hernández-Pereira, E. and Moret-Bonillo, V. (2004) Intelligent diagnosis of sleep apnea syndrome. *IEEE Engineering in Medicine and Biology Magazine*, 23, pp. 72–81.
- Cacoullos, T. (1966) Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18, pp. 179–189.
- Caples, S.M., Gami, AS and Somers, V.K. (2005) Obstructive sleep apnea. *Annals of Internal Medicine*, 142, pp. 187–197.

- Caseiro, P., Fonseca-Pinto, R. and Andrade, A. (2010) Screening of obstructive sleep apnea using Hilbert-Huang decomposition of oronasal airway pressure recordings. *Medical Engineering & Physics*, 32, pp. 561–568.
- Cattell, R.B. (1966) Scree test for number of factors. *Multivariate Behavioural Research*, 1, pp. 245–276.
- Chen, C.C. and Barnhart, H.X. (2008) Comparison of ICC and CCC for assessing agreement for data without and with replications. *Computational Statistics and Data Analysis*, 53, pp. 554–564.
- Chesson, A.L, Anderson, W.McD., Walls, R.C. and Bairnsfather, L.E. (1993) Assessment of hypoxemia in patients with sleep disorders using saturation impairment time. *American Review of Respiratory Disorders*, 148, pp. 1592–1598.
- Chiner, E., Signes-Costa, J., Arriero, J.M., Marco, J., Fuentes, I. and Sergado, A. (1999) Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies?. *Thorax*, 54, pp. 968–971.
- Chua, C.P., Garvey, J., Redmond, S., Heneghan, C. and McNicholas, W.T. (2007) Towards automated sleep state estimation using a Holter-oximeter. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, August 2007*. Lyon: IEEE, pp. 3998–4001.
- Cirignotta, F., Mondini, S., Gerardi, R., Mostacci, B. and Sancisi, E. (2001) Unreliability of automatic scoring of MESAM 4 in assessing patients with complicated obstructive sleep apnea syndrome. *Chest*, 119, pp. 1387–1392.
- Cohen, A. (2006) Biomedical signals: origin and dynamic characteristics; frequency-domain analysis. In Bronzino, J.D. (ed.) *The Biomedical Engineering Handbook. Medical devices and systems*. 3rd ed. Boca Raton: Taylor & Francis, pp. 1.

- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, pp. 37–46.
- Cohen, M.E., Hudson, D.L. and Deedwania, P.C. (1996) Applying continuous chaotic modelling to cardiac signals. *IEEE Engineering in Medicine and Biology Magazine*, 15, pp. 97–102.
- Collop, N.A., Anderson, W.Mc., Boehlecke, B., Claman, D., Goldberg, R., Gottlieb, D.J., Hudgel, D., Sateia, M. and Schwab, R. (2007) Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. *Journal of Clinical Sleep Medicine*, 3, pp. 737–747.
- Cooper, B.G., Veale, D., Griffiths, C.J. and Gibson, G.J. (1991) Value of nocturnal oxygen saturation as a screening test for sleep apnoea. *Thorax*, 46, pp. 586–588.
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20, pp. 273–297.
- Crino, S. and Brown, D.E. (2007) Global optimization with multivariate adaptive regression splines. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 37, pp. 333–340.
- Crocker, B.D., Olson, L.G., Saunders, N.A., Hensley, M.J., McKeon, J.L., Allen, K.M. and Gyulay, S.G. (1990) Estimation of the probability of disturbed breathing during sleep before a sleep study. *The American Review of Respiratory Disease*, 142, pp. 14–18.
- Davies, R.J., Ali, N.J. and Stradling, J.R. (1992) Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome. *Thorax*, 47, pp. 101–105.
- Deegan, P.C. and McNicholas, W.T. (1996) Predictive value of clinical features for the obstructive sleep apnoea syndrome. *European Respiratory Journal*, 9, pp. 117–124.

- Del Campo, F., Hornero, R., Zamarrón, C., Abásolo, D. and Álvarez, D. (2006) Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea. *Artificial Intelligence in Medicine*, 37, 111–118.
- de Boor, C. (1978) *A practical guide to splines*. New York: Springer.
- de Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P. and O'Malley, M. (2000) Automatic classification of sleep apnea epochs using the electrocardiogram. *Computers in Cardiology*, 27, pp. 745–748.
- de Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P. and O'Malley, M. (2003) Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Transactions on Biomedical Engineering*, 50, pp. 686–696.
- de Chazal, P., Penzel, T. and Heneghan, C. (2004) Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram, *Physiological Measurement*, 25, pp. 967–983.
- de Chazal, P., Heneghan, C. and McNicholas, W.T. (2009) Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals. *Philosophical Transactions of the Royal Society A*, 367, pp. 369–389.
- Dingli, K., Coleman, E.L., Vennelle, M., Finch, S.P., Wraith, P.K., Mackay, T.W. and Douglas, N.J. (2003) Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome. *European Respiratory Journal*, 21, pp. 253–259.
- Dixon, J.B., Schachter, L.M. and O'Brien, P.E. (2003) Predicting sleep apnea and excessive day sleepiness in the severely obese. *Chest*, 123, pp. 1134–1141.
- Di Eugenio, B. and Glass, M. (2004) The Kappa statistic: a second look. *Computational Linguistics*, 30, pp. 95–101.

- Durán, J., Esnaola, S, Rubio, R. and Iztueta, A. (2001) Obstructive sleep apnea–hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr. *American Journal of Respiratory and Critical Care Medicine*, 163, pp. 685–689.
- El-Solh, A.A., Mador, M.J., Ten-Brock, E., Shucard, D.W., Abul-Khoudoud, M. and Grant, B.J.B. (1999) Validity of neural network in sleep apnea. *Sleep*, 22, pp. 105–111.
- El-Solh, A.A., Magalang, U.J., Mador, M.J., Dmochowski, J., Veeramachaneni, S., Saberi, A., Draw, A.M., Lieber, B.B. and Grant, B.J.B. (2003) The utility of neural network in the diagnosis of Cheyne-Stokes respiration. *Journal of Medical Engineering and Technology*, 27, pp. 54–58.
- Enderle, J.D., Bronzino, J.D. and Blanchard, S.M. (2005) *Introduction to biomedical engineering*. Burlington: Elsevier Academic Press.
- Epstein, L.J. and Dorlac, G.R. (1998) Cost-effectiveness analysis of nocturnal oximetry as a method of screening for sleep apnea-hypopnea syndrome. *Chest*, 113, pp. 97–103.
- Fisher, R.A. (1925) *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, R.A. (1936) The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7, pp. 179–188.
- Fiz, J.A., Jané, R., Solá-Soler, J., Abad, J., García, M.A. and Morera, J. (2010) Continuous analysis and monitoring of snores and their relationship to the apnea-hypopnea index. *Laryngoscope*, 120, pp. 854–862.
- Flemons, W.W., Whitelaw, W.A., Brant, R. and Remmers. J.E. (1994) Likelihood ratios for a sleep apnea clinical prediction rule. *American Journal of Respiratory and Critical Care Medicine*, 150, pp. 1279–1285.

- Flemons, W.W. and McNicholas, W.T. (1997) Clinical prediction of the sleep apnea syndrome. *Sleep Medicine Reviews*, 1, pp. 19–32.
- Flemons, W.W., Douglas, N.J., Kuna, S.T., Rodenstein, D.O. and Wheatley, J. (2004) Access to diagnosis and treatment of patients with sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, vol. 169, pp. 668–672.
- Fontela-Romero, O., Guijarro-Berdiñas, B., Alonso-Betanzos, A. and Moret-Bonillo, V. (2005) A new method for sleep apnea classification using wavelets and feedforward neural networks. *Artificial Intelligence in Medicine*, 34, pp. 65–76.
- Freund, Y. (1995) Boosting a weak learning algorithm by majority. *Information and Computation*, 121, pp. 256–285.
- Friedman, J.H. and Silverman, B.W. (1989) Flexible parsimonious smoothing and additive modelling. *Technometrics*, 31, pp. 3–39.
- Friedman, J.H. (1991) Multivariate adaptive regression splines. *The Annals of Statistics*, 19, pp. 1–141.
- Friedman, M., Wilson, M.N., Pulver, T., Pandya, H., Joseph, N.J., Lin, H.C. and Chang, H.W. (2010) Screening for obstructive sleep apnea/hypopnea syndrome: subjective and objective factors. *Otolaryngology–Head and Neck Surgery*, 142, pp. 531–535.
- Gagnadoux, F., Pelletier-Fleury, N., Philippe, C., Rakatonanahary, D. and Bernard, F. (2002) Home unattended vs hospital telemonitored polysomnography in suspected obstructive sleep apnea syndrome: a randomized crossover trial. *Chest*, 121, pp. 753–758.
- Galvin, J.R., Rooholamini, S.A. and Stanford, W. (1989) Obstructive sleep apnea: diagnosis with ultrafast CT. *Radiology*, 171, pp. 775–778.

- García, M., Sánchez, C.I., Poza, J., López, M.I. and Hornero, R. (2009) Detection of hard exudates in retinal images using a radial basis function classifier. *Annals of Biomedical Engineering*, 37, pp. 1448–1463.
- Gastaut, H., Tassinari, C.A. and Duron, B. (1966) Polygraphic study of the episodic diurnal and nocturnal (hypnic and respiratory) manifestations of the Pickwick syndrome. *Brain Research*, 1, pp. 167–186
- Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K. and Stanley, H.E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101, pp. e215–e220.
- Golpe, R., Jiménez, A., Carpizo, R. and Cifrian, J.M. (1999) Utility of home oximetry as a screening test for patients with moderate to severe symptoms of obstructive sleep apnea. *Sleep*, 22, pp. 932–937.
- Golpe, R, Jiménez, A. and Carpizo, R. (2002) Home sleep studies in the assessment of sleep apnea/hypopnea syndrome. *Chest*, 122, pp. 1156–1161.
- Gottlieb, D.J., Yao, Q., Redline, S., Ali, T. and Mahowald, M.W. (2000) Does snoring predict sleepiness independently of apnea and hypopnea frequency?. *American Journal of Respiratory and Critical Care Medicine*, 162, pp. 1512–1517.
- Gugger, M. (1997) Comparison of ResMed AutoSet (version 3.03) with polysomnography in the diagnosis of the sleep apnoea/hypopnoea syndrome. *European Respiratory Journal*, 10, pp. 587–591.
- Guilleminault, C., Eldridge, F. and Dement, W.C. (1973) Insomnia with sleep apnea: a new syndrome. *Science*, 181, pp. 856–858.
- Guimaraes, G., Peter, J.H., Penzel, T. and Ultsch, A. (2001) A method for automated temporal knowledge acquisition applied to sleep-related breathing disorders. *Artificial Intelligence in Medicine*, 23, pp. 211–237.

- Gurubhagavatula, I., Maislin, G. and Pack, A.I. (2001) An algorithm to stratify sleep apnea risk in a sleep disorders clinic population. *American Journal of Respiratory and Critical Care Medicine*, 164, pp. 1904–1909.
- Gyulay, S., Olson, L.G., Hensley, M.J., King, M.T., Allen, K.M. and Saunders, N.A. (1993) A comparison of clinical assessment and home oximetry in the diagnosis of obstructive sleep apnea. *The American Review of Respiratory Disease*, 147, pp. 50–53.
- Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a receiving operating characteristic (ROC) curve. *Radiology*, 143, pp. 29–36.
- Hanning, C.D. and Alexander-Williams, J.M. (1995) Pulse oximetry: a practical review. *British Medical Journal*, 311, pp. 367–370.
- Haykin, S. (1996) Neural networks expand SP's horizons. *IEEE Signal Processing Magazine*, 13, pp. 24–49
- Haykin, S. (1999) *Neural networks: a comprehensive foundation*. Upper Saddle River: Prentice Hall International.
- Heneghan, C., Chua, C.P., Garvey, J.F., de Chazal, P., Shouldice, R., Boyle, P. and McNicholas, W.T. (2008a) A portable automated assessment tool for sleep apnea using a combined Holter-oximeter. *Sleep*, 31, pp. 1432–1439.
- Heneghan, C., de Chazal, P., Ryan, S., Chua, C.P., Doherty, L., Boyle, P., Nolan, P. and McNicholas, W.T. (2008b) Electrocardiogram recording as a screening tool for sleep disordered breathing. *Journal of Clinical Sleep Medicine*, 4, pp. 223–228.
- Herer, B., Roche, N., Carton, M., Roig, C., Poujol, V. and Huchon, G. (1999) Value of clinical, functional, and oximetric data for the prediction of obstructive sleep apnea in obese patients. *Chest*, 116, pp. 1537–1544.

- Herer, B., Fuhrman, C., Roig, C. and Housset, B. (2002) Prediction of obstructive sleep apnea by OxiFlow in overweight patients. *Sleep Medicine*, 3, pp. 417–422.
- Hilton, M.F., Bates, R.A., Godfrey, K.R., Chappell, M.J. and Cayton, R.M. (1999) Evaluation of frequency and time-frequency spectral analysis of heart rate variability as a diagnostic marker of the sleep apnoea syndrome. *Medical & Biological Engineering & Computing*, 37, pp. 760–769.
- Hoffstein, V. and Szalai, J.P. (1993) Predictive value of clinical features in diagnosing obstructive sleep apnea. *Sleep*, 16, pp. 118–122.
- Holmström, L., Koistinen, P., Laaksonen, J. and Oja, E. (1997) Neural and Statistical Classifiers—Taxonomy and Two Case Studies. *IEEE Transactions on Neural Networks*, 8, pp. 5–17.
- Hornero, R., Álvarez, D., Abásolo, D., Del Campo, F. and Zamarrón, C. (2007) Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome. *IEEE Transactions on Biomedical Engineering*, 54, pp. 107–113.
- Hornero, R., Escudero, J., Fernández, A., Poza, J. and Gómez, C. (2008) Spectral and nonlinear analyses of MEG background activity in patients with Alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 55, pp. 1658–1665.
- Hornik, K. (1991) Approximation capabilities of multilayer feedforward networks, *Neural Networks*, 4, pp. 251–257.
- Hsu, C.W. and Lin, C.J. (2002) A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13, pp. 415–425.
- Hush, D.R. and Horne, B.G. (1993) Progress in supervised neural networks. *IEEE Signal Processing Magazine*, 10, pp. 8–39.

- Hussain, S.F. and Fleetham, J.A. (2003) Overnight home oximetry: can it identify patients with obstructive sleep apnea-hypopnea who have minimal daytime sleepiness?. *Respiratory Medicine*, 97, pp. 537–540.
- Issa, F.G., Morrison, D., Hadjuk, E., Iyer, A., Feroah, T. and Remmers, J.E. (1993) Digital monitoring of sleep-disordered breathing using snoring sound and arterial oxygen saturation. *The American Review of Respiratory Disease*, 148, pp. 1023–1029.
- Jain, A.K., Duin, R.P.W. and Mao, J. (2000) Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, pp. 4–37.
- Jarvis, M.R. and Mitra, R.R. (2000) Apnea patients characterized by 0.02 Hz peak in the multitaper spectrogram of electrocardiogram signals. *Computers in Cardiology*, 27, pp. 769–772
- Jobson, J.D. (1991a) *Applied multivariate data analysis. Volume I: regression and experimental design*. New York: Springer.
- Jobson, J.D. (1991b) *Applied multivariate data analysis. Volume II: categorical and multivariate methods*. New York: Springer.
- Johns, M.W. (1991) A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep*, 14, pp. 540–545.
- Jolliffe, I.T. (2002) *Principal Component Analysis*. New York: Springer.
- Jubran, A. (2004) Pulse oximetry. *Intensive Care Medicine*, 30, pp. 2017–2020.
- Jung, R. and Kuhlo, W. (1965) Neurophysiological studies of abnormal night sleep and the Pickwickian syndrome. *Progress in Brain Research*, 18, pp. 140–159.
- Kaiser, H.F. (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, pp. 141–151.

- Khandoker, A.H., Karmakar, C.K. and Palaniswami, M. (2009a) Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings. *Computers in Biology and Medicine*, 39, pp. 88–96.
- Khandoker, A.H., Palaniswami, M. and Karmakar, C.K. (2009b) Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. *IEEE Transactions on Information Technology in Biomedicine*, 13, pp. 37–48.
- Kiely, J.L., Delahunty, C., Matthews, S. and McNicholas, W.T. (1996) Comparison of a limited computerized diagnostic system (ResCare Autoset[®]) with polysomnography in the diagnosis of obstructive sleep apnoea syndrome. *European Respiratory Journal*, 9, pp. 2360–2364.
- Kirby, S.C., Anderson, W.McD., Chesson, A.L. and George, R.B. (1992) Computer quantitation of saturation impairment time as an index of oxygenation during sleep. *Computer Methods and programs in Biomedicine*, 38, pp. 107–115.
- Kirby, S.D., Danter, W., George, C.F.P., Francovic, T., Ruby, R.P.F. and Ferguson, K.A. (1999) Neural network prediction of obstructive sleep apnea from clinical criteria. *Chest*, 116, pp. 409–415.
- Kryger, M.H. (1985) Fat, sleep, and Charles Dickens: literary and medical contributions to the understanding of sleep apnea. *Clinics in Chest Medicine*, 6, pp. 555–562.
- Kulkarni, S.R., Lugosi, G. and Venkatesh, S.S. (1998) Learning pattern classification-A survey. *IEEE Transactions on Information Theory*, 40, pp. 2178–2206.
- Kushida, C.A., Littner, M.R., Morgenthaler, T., Alessi, C.A., Bailey, D., Coleman, J., Friedman, L., Hirshkowitz, M., Kapen, S., Kramer, M., Lee-Chiong, T., Loube, D.L., Owens, J., Pancer, J.P. and Wise, M. (2005) Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*, 48, pp. 499–521.

- Lacassagne, L., Didier, A., Murriss-Espin, M., Charlet, J.P., Chollet, P., Léophonte-Domairon, M.L., Tiberge, M., Pessey, J.J. and Léophonte, P. (1997) Role of nocturnal oximetry in screening for sleep apnea syndrome in pulmonary medicine. Study of 329 patients. *Revue des Maladies Respiratoires*, 14, pp. 201–207.
- Laguna, P. and Sörnmo, L. (2009) Introduction. Editorial on Signal processing in vital rhythms and signs. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367, pp. 207–211.
- Ledley, R.S. and Lusted, L.B. (1960) The use of electronic computers in medical data processing: aids in diagnosis, current information retrieval, and medical record keeping. *IRE Transactions on Medical Electronics*, ME-7, pp. 31–47.
- Lempel, A. and Ziv, J. (1976) On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22, pp. 75–81.
- Lehman, C., Koenig, T., Jelic, V., Prichep, L., John, L.E., Wahlund, L.O., Dodge, Y. and Diers, T. (2007) Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *Journal of Neuroscience Methods*, 161, pp. 342–350.
- Lévy, P., Pépin, J.L., Deschaux-Blanc, C., Paramelle, B. and Brambilla, C. (1996) Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. *Chest*, 109, pp. 395–399.
- Lin, R., Lee, R.G., Tseng, C.L., Zhou, H.K., Chao, C.F. and Jiang, J.A. (2006) A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks. *Biomedical Engineering-Applications, Basis and Communications*, 18, pp. 138–143.
- Lisboa, P.J.G. (2002) A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks*, 15, pp. 11–39.

- Liu, D., Pang, Z. and Lloyd, S.R. (2008) A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG. *IEEE Transactions on Neural Networks*, 19, pp. 308–318.
- Macey, P.M., Li, J.S.J. and Ford, R.P.K. (1998) Expert system for the detection of apnoea. *Engineering Applications of Artificial Medicine*, 11, pp. 425–438.
- Mackay, D.J.C. (1992a) A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4, pp. 448–472.
- Mackay, D.J.C. (1992b) The evidence framework applied to classification networks. *Neural Computation*, 4, pp. 720–726.
- Magalang, U.J., Dmochowski, J., Veeramachaneni, S., Draw, A., Mador, M.J., El-Solh, A. and Grant, B.J.B. (2003) Prediction of the apnea-hypopnea index from overnight pulse oximetry. *Chest*, 124, pp. 1694–1701.
- Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., López, M. and Zamarrón, C. (2008a) Radial basis function classifiers to help in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. *Medical & Biological Engineering & Computing*, 46, pp. 323–332.
- Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M. (2008b) Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. *Computer Methods and Programs in Biomedicine*, 72, pp. 79–89.
- Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and Zamarrón, C. (2009) Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry. *Medical Engineering & Physics*, 31, pp. 971–978.
- Marcos, J.V., Hornero, R., Álvarez, D., Nabney, I.T., Del Campo, F. and Zamarrón, C. (2010a) The classification of oximetry signals using Bayesian neural networks to

- assist in the detection of obstructive sleep apnoea syndrome. *Physiological Measurement*, 31, pp. 375–394.
- Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F. and Aboy, M. (2010b) Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis. *Medical & Biological Engineering & Computing*, 48, pp. 895–902.
- Masa, J.F., Barbé, F., Capote, F., Chiner, E., Díaz, J., Durán, J., López, S., Marín, J.M., Montserrat, J.M., Terán, J. and Zamarrón, C. (2004) Resources and delays in the diagnosis of sleep apnea-hypopnea syndrome. *Archivos de Bronconeumología*, 43, pp. 188–198.
- Mayer, P., Meurice, J.C., Philip-Joet, F., Cornette, A., Rakotonanahary, D., Meslier, N., Pepin, J.L., Lévy, P. and Veale, D. (1998) Simultaneous laboratory-based comparison of ResMed Autoset™ with polysomnography in the diagnosis of sleep apnoea/hypopnoea syndrome. *European Respiratory Journal*, 12, pp. 770–775.
- McNames, J.N. and Fraser, A.M. (2000) Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram. *Computers in Cardiology*, 27, pp. 749–752.
- McNicholas, W.T. (2008) Diagnosis of obstructive sleep apnea in adults. *Proceedings of the American Thoracic Society*, 5, pp. 154–160.
- Mendelson, Y. (1992) Pulse oximetry: theory and applications for noninvasive monitoring. *Clinical Chemistry*, 38, 1601–1607.
- Méndez, M.O., Bianchi, A.M., Matteucci, M., Cerutti, S. and Penzel, T. (2009) Sleep apnea screening by autoregressive models from a single ECG lead. *IEEE Transactions on Biomedical Engineering*, 56, pp. 2838–2850.

- Méndez, M.O., Corthout, J., Van Huffel, S., Matteucci, M., Penzel, T., Cerutti, S. and Bianchi, A.M. (2010) Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis. *Physiological Measurement*, 31, pp. 273–289.
- Moller, M.F. (1993) A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, 6, pp. 525–533.
- Moody, J. and Darken, C.J. (1989) Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1, pp. 281–294.
- Moody, J.E. (1992) The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems. In: Moody, J.E., Hanson, S.J. and Lippmann, R.P. (eds.) *Advances in neural information processing systems*. San Mateo: Morgan Kaufmann Publishers
- Morillo, D.S., Rojas, J.L., Crespo, L.F., León, A. and Gross, N. (2009) Poincaré analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hypopnea syndrome. *Physiological Measurement*, 30, pp. 405–420.
- Moyle, J.T. (1996) Uses and abuses of pulse oximetry. *Archives of Disease in Childhood*, 74, pp. 77–80.
- Nabney, I.T. (2002) *NETLAB: Algorithms for pattern recognition*. Berlin: Springer.
- Nabney, I.T. (2004) Efficient training of RBF networks for classification. *International Journal of Neural Systems*, 14, pp. 201–208.
- Nakano, H., Ikeda, T., Hayashi, M., Ohshima, E., Itoh, M., Nishikata, N. and Shinohara, T. (2004a) Effect of body mass index on overnight oximetry for the diagnosis of sleep apnea. *Respiratory Medicine*, 98, pp. 421–427.

- Nakano, H., Hayashi, M., Ohshima, E., Nishikata, N. and Shinohara, T. (2004b) Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome. *Sleep*, 27, pp. 951–957.
- Nazeran, H., Almas, A., Behbehani, K., Burk, J., and Lucas, E. (2001) A fuzzy inference system for detection of obstructive sleep apnea. In: *Proceedings of the 23th Annual International Conference of the IEEE EMBS, Istanbul, October 2001*. Istanbul: IEEE, pp. 1645–1648.
- Neal, R.M. (1996) *Bayesian learning for neural networks*. New York: Springer.
- Nebeker, F. (2002) Golden accomplishments in biomedical engineering. *IEEE Engineering in Medicine and Biology Magazine*, 21, pp. 17–47.
- Nettleton, D. and Muñiz, J. (2001) Processing and representation of meta-data for sleep apnea diagnosis with an artificial intelligence approach. *International Journal of Medical Informatics*, 63, pp. 77–89.
- Netzer, N., Eliasson, A.H., Netzer, C. and Kristo, D.A. (2001) Overnight pulse oximetry for sleep-disordered breathing in adults: a review. *Chest*, 120, pp. 625–633.
- Ng, A.K., Koh, T.S., Baey, E., Lee, T.H., Abeyratne, U.R. and Puvanendran, K. (2008) Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?. *Sleep Medicine*, 9, pp. 894–898.
- Nigro, C.A., Aimaretti, S., González, S. and Rhodius, E. (2009) Validation of the WristOx 3100™ oximeter for the diagnosis of sleep apnea/hypopnea syndrome. *Sleep and Breathing*, 13, 127–136.
- Olson, L.G., Ambrogetti, A. and Gyulay, G. (1999) Prediction of disordered-breathing by unattended overnight oximetry. *Journal of Sleep Research*, 8, pp. 51–55.

- Özbay, Y., Ceylan, R. and Karlik, B. (2006) A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine*, 36, pp. 376–388.
- Pang, K.P. and Terris, D.J. (2006) Screening for obstructive sleep apnea: an evidence-based analysis. *American Journal of Otolaryngology–Head and Neck Medicine and Surgery*, 27, pp. 112–118
- Pardo, M. and Sberveglieri, G. (2004) Remarks on the use of multilayer perceptrons for the analysis of chemical sensor array data. *IEEE Sensors Journal*, 4, pp. 355–363.
- Parra, O., García-Esclasans, N., Montserrat, J.M., García, L., Ruíz, J., López, J.A., Guerra, J.M. and Sopena, J.J. (1997) Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies?. *European Respiratory Journal*, 10, pp. 1720–1724.
- Parzen, E. (1962) On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33, pp. 1065–1076.
- Patangay, A., Vemuri, P. and Tewfik, A. (2007) Monitoring of obstructive sleep apnea in heart failure patients. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, August 2007*. Lyon: IEEE, pp. 1043–1046.
- Penzel, T. and Conradt, R. (2000) Computer based sleep recording and analysis. *Sleep Medicine Reviews*, 4, pp. 131–148.
- Penzel, T., McNames, J., de Chazal, P., Raymond, B., Murray, A. and Moody, G. (2002) Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. *Medical & Biological Engineering & Computing*, 40, pp. 402–407.
- Pepin, J.L., Lévy, P., Lepaulle, B., Brambilla, C. and Guilleminault, C. (1991) Does oximetry contribute to the detection of apneic events?. *Chest*, 99, pp. 1151–1157.

- Philips Respironics (2010) Alice 5 Diagnostic Sleep System, Philips Respironics [WWW] Available from: <http://alice5.respironics.com/> [Accessed 15/12/2010].
- Pincus, S.M. (2001) Assessing serial irregularity and its implications for health. *Annals of the New York Academy of Science*, 954, pp. 245–267.
- Portier, F., Portmann, A., Czernichow, P., Vascaut, L., Devin, E., Benhamou, D., Cuvelier, A. and Muir, J.F. (2000) Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome. *American Journal of Respiratory and Critical Care Medicine*, 162, pp. 814–818.
- Preul, M.C., Caramanos, Z., Collins, D.L., Villemure, J.G., Leblanc, R., Olivier, A., Pokrupa, R. and Arnold, D.L. (1996) Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Nature Medicine*, 2, pp. 323–325.
- Qureshi, A. and Ballard, R.D. (2003) Obstructive sleep apnea. *The Journal of Allergy and Clinical Immunology*, 112, pp. 643–651.
- Raymond, B., Cayton, R.M., Bates, R.A. and Chappell, M.J. (2000) Screening for obstructive sleep apnoea based on the electrocardiogram—The computers in Cardiology Challenge. *Computers in Cardiology*, 27, pp. 267–270.
- Raymond, B., Cayton, R.M. and Chappell, M.J. (2003) Combined index of heart rate variability and oximetry in screening for the sleep apnoea/hypopnoea syndrome. *Journal of Sleep Research*, 12, pp. 53–61.
- Rechtschaffen, A and Kales, A (1968) *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Los Angeles: Brain Information Services, Brain Research Institute, University of California.
- Redmond, S.J. and Heneghan, C. (2006) Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. *IEEE Transactions on Biomedical Engineering*, 53, pp. 485–496.

- Reichert, J.A., Bloch, D.A., Cundiff, E. and Votteri, B.A. (2003) Comparison of the NovaSom QSGTM, a new sleep apnea home-diagnostic system, and polysomnography. *Sleep Medicine*, 4, 213–218.
- Rencher, A.C. (1992) Interpretation of canonical discriminant functions, canonical variates, and principal components. *The American Statistician*, 46, pp. 217–225.
- Roche, F., Gaspoz, J.M., Court-Fortune, I., Minini, P., Pichot, V., Duverney, D., Costes, F., Lacour, J.R., Barthélémy, J.C. (1999) Screening of obstructive sleep apnea syndrome by heart rate variability analysis. *Circulation*, 100, pp. 1411–1415.
- Roche, F., Pichot, V., Sforza, E., Court-Fortune, I., Duverney, D., Costes, F., Garet, M. and Barthélémy, J.C. (2003) Predicting sleep apnoea syndrome from heart period: a time-frequency wavelet analysis. *European Respiratory Journal*, 22, pp. 937–942.
- Roche, N., Herer, B., Roig, C. and Huchon, G. (2002) Prospective testing of two models based on clinical and oximetric variables for prediction of obstructive sleep apnea. *Chest*, 121, pp. 747–752.
- Rodríguez, J.M., De Lucas, P., Sánchez, M.J., Izquierdo, J.L., Peraíta, R. and Cubillo, J.M. (1996) Usefulness of the visual analysis of night oximetry as a screening method in patients with suspected clinical obstructive sleep apnea syndrome. *Archivos de Bronconeumología*, 32, pp. 437–441.
- Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, pp. 386–408.
- Rumelhart, D.E., Hinton, G.E. and R. J. Williams (1986) Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. and the PDP Research Group (eds.) *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: foundations*. Cambridge, MA: MIT Press, pp. 318–362.

- Ryan, P.J., Hilton, M.F., Boldy, D.A., Evans, A., Bradbury, S., Sapiano, S., Prowse, K. and Cayton, R.M. (1995) Validation of British Thoracic Society guidelines for the diagnosis of the sleep apnoea/hypopnoea syndrome: can polysomnography be avoided?. *Thorax*, 50, pp. 972–975.
- Salisbury, J.I. and Sun, Y. (2007) Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique. *Medical Engineering & Physics*, 29, pp. 336–343.
- Sano, K., Nakano, H., Ohnishi, Y., Ishii, Y., Nakamura, T., Matuzawa, K., Maekawa, J. and Narita, N. (1998) Screening of sleep apnea/hypopnea syndrome by home pulse oximetry. *Nihon Kokyuki Gakkai Zasshi*, 36, pp. 948–952.
- Schmittendorf, E., Schultheiß, B., Goffart, M. and Böhning, N. (2009) Obstructive sleep apnea screening using nonlinear characteristics of overnight pulse oximetry recordings. In: *Proceedings of the World Congress on Medical Physics and Biomedical Engineering, Munich, September 2009*. Berlin: Springer Berlin Heidelberg, pp. 724–727.
- Schwaighofer, A., Tresp, V., Mayer, P., Krause, A., Beuthan, J., Rost, H., Metzger, G., Müller, G.A. and Scheel, A.K. (2003) Classification of rheumatoid joint inflammation based on laser imaging. *IEEE Transactions on Biomedical Engineering*, 50, pp. 375–382.
- Schwenker, F., Kestler, H.A. and Palm, G. (2001) Three learning phases for radial-basis-function networks. *Neural Networks*, 14, pp. 439–458.
- Sériès, F., Marc, I., Cormier, Y. and La Forge, J. (1993) Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome. *Annals of Internal Medicine*, 119, pp. 449–453.
- Sériès, F., Kimoff, R.J., Morrison, D., Leblanc, M.H., Smilovitch, M., Howlett, J., Logan, A.G., Floras, J.S., and Bradley, T.D. (2005) Prospective evaluation of

- nocturnal oximetry for detection of sleep-related breathing disturbances in patients with chronic heart failure. *Chest*, 127, pp. 1507–1514.
- Sezgin, N. and Tagluk, M.E. (2009) Energy based feature extraction for classification of sleep apnea syndrome. *Computers in Biology and Medicine*, 39, pp. 1043–1050.
- Sher, A.E., Schechtman, K.B. and Piccirillo, J.F (1996) The efficacy of surgical modifications of the upper airway in adults with obstructive sleep apnea syndrome. *Sleep*, 19, pp. 156-177.
- Shouldice, R., O'Brien, L., de Chazal, P., Gozal, D. and Heneghan, C. (2004) Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features. *Sleep*, 27, pp. 784–792.
- Shrout, P.E. and Fleiss, J.L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 36, pp. 420–428.
- Sivan, Y., Kornecki, A. and Schonfeld, T. (1996) Screening obstructive sleep apnoea syndrome by home videotape recording in children. *European Respiratory Journal*, 9, pp. 2127–2131.
- Smola, A.J. and Schölkopf, B. (2004) A tutorial on support vector regression. *Statistics and Computing*, 14, pp. 199–222.
- Specht, D.F. (1990) Probabilistic neural networks. *Neural Networks*, 3, pp. 109–118.
- Specht, D.F. (1991) A general regression neural network. *IEEE Transactions on Neural Networks*, 2, pp. 568–576.
- Stoohs, R. and Guilleminault, C. (1992) MESAM 4: an ambulatory device for the detection of patients at risk for obstructive sleep apnea syndrome (OSAS). *Chest*, 101, pp. 1221-1227.

- Sugi, T., Kawana, F. and Nakamura, M. (2009) Automatic EEG arousal detection for sleep apnea syndrome. *Biomedical Signal Processing and Control*, 4, pp. 329–337.
- Su, S., Baroody, F.M., Kohrman, M. and Suskind, D. (2004) A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome. *Otolaryngology–Head and Neck Surgery*, 131, pp. 844–850.
- Tagluk, M.E., Akin, M. and Sezgin, N. (2010) Classification of sleep apnea by using wavelet transform and artificial neural networks. *Expert Systems with Applications*, 37, pp. 1600–1607.
- Taktak, A.F.G., Simpson, S., Patel, S. and Meyer, G. (2000) Neural network analysis of oxygenation signals in infants during sleep. *Physiological Measurement*, 21, pp. N11–N22.
- Tassinari, C. and Lugaresi, E. (1970) Obstructive sleep apnoea-hypopnoea syndrome. *Revue de Neurologique*, 123, pp. 267–268.
- Teramoto, S., Matsuse, T. and Fukuchi, Y. (2002) Clinical significance of nocturnal oximeter monitoring for detection of sleep apnea syndrome in the elderly. *Sleep Medicine*, 3, pp. 67–71.
- Theodoridis, S. and Koutroumbas, K. (2009) *Pattern recognition*. New York: Academic Press.
- Tishler, P.V., Larkin, E.K., Schluchter, M.D. and Redline, S. (2003) Incidence of sleep-disordered breathing in an urban adult population. *Journal of the American Medical Association*, 289, pp. 2230–2237.
- Tremper, K.K. (1989) Pulse oximetry. *Chest*, 95, pp. 713–715.

- Tsai, W.H., Remmers, J.E., Brant, R., Flemons, W.W., Davies, J. and Macarthur, C. (2003) A decision rule for diagnostic testing in obstructive sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 167, 1427–1432.
- Van Houwelingen, K.G., Van Uffelen, R. and Van Vliet, A.C.M. (1999) The sleep apnoea syndromes. *European Heart Journal*, 20, pp. 858–866.
- Vapnik, V.N. (1999) An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10, pp. 988–999.
- Vapnik, V. and Chapelle, O. (2000) Bounds on error expectation for support vector machines. *Neural Computation*, 12, pp. 2013–2036.
- Várady, P., Micsik, T., Benedek, S. and Benyó, Z. (2002) A novel method for the detection of apnea and hypopnea events in respiration signals. *IEEE Transactions on Biomedical Engineering*, 49, pp. 936–942.
- Vázquez, J.C., Tsai, W.H., Flemons, W.W., Masuda, A., Brant, R., Hadjuk, E., Whitelaw, W.A. and Remmers, J.E. (2000) Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea. *Thorax*, 55, pp. 302–307.
- Viner, S., Szalai, J.P. and Hoffstein, V. (1991) Are history and physical examination a good screening test for sleep apnea?. *Annals of Internal Medicine*, 115, pp. 356–359.
- Wang, R.C., Elkins, T.P., Keech, D., Wauquier, A. and Hubbard, D. (1998) Accuracy of clinical evaluation in pediatric obstructive sleep apnea. *Otolaryngology–Head and Neck Surgery*, 118, pp. 69–73.
- Weinreich, G., Armitstead, J. and Teschler, H. (2008) Pattern recognition of obstructive sleep apnoea and Cheyne–Stokes respiration. *Physiological Measurement*, 29, pp. 869–878.

- Weir, J.P. (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19, pp. 231–240.
- Welch, P. (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15, pp. 70–73.
- Westbrook, P.R., Levendowski, D.J., Cvetinovic, M., Zavora, T., Velimirovic, V., Henninger, D. and Nicholson, D. (2005) Description and validation of the apnea risk evaluation system. *Chest*, 128, pp. 2166–2175.
- Williams, A.J., Yu, G., Santiago, S. and Stein, M. (1991) Screening for sleep apnea using pulse oximetry and a clinical score. *Chest*, 100, pp. 631–635.
- Yadollahi, A., Giannouli, E. and Moussavi, Z. (2010) Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals. *Medical & Biological Engineering & Computing*, 48, pp. 1087–1097.
- Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S. and Badr, S. (1993) The occurrence of sleep-disordered breathing among middle-aged adults. *New England Journal of Medicine*, 328, pp. 1230–1235.
- Young T., Evans L., Finn L. and Palta M (1997) Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*, 20, pp. 705–706.
- Young, T., Peppard, P.E. and Gottlieb, D.J. (2002) Epidemiology of obstructive sleep apnea: a population health perspective. *American Journal of Respiratory and Critical Care Medicine*, 165, pp. 1217–1239.
- Zamarrón, C., Romero, P.V., Rodríguez, J.R. and Gude, F. (1999) Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea. *Clinical Science*, 97, pp. 467–473.

- Zamarrón, C., Gude, F., Barcala, J., Rodríguez, J.R. and Romero, P.V. (2003) Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome. *Chest*, 123, pp 1567–1576.
- Zamarrón, C., Hornero, R., Del Campo, F., Abásolo, D. and Álvarez, D. (2006) Heart rate regularity analysis obtained from pulse oximetric recordings in the diagnosis of obstructive sleep apnea. *Sleep and Breathing*, 10, pp. 83–89.
- Zhang, G.P. (2000) Neural networks for classification: a survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 30, pp. 451–462.
- Zhou, Z.H., Wu, J. and Tang, W. (2002) Ensembling neural networks: many could be better than all. *Artificial Intelligence*, 137, pp. 239–263.
- Zigel, Y., Tarasiuk, A. and Goldshtein, A. (2008) Analysis of speech signals among obstructive sleep apnea patients. In: *Proceedings of the IEEE 25th Convention of Electrical and Electronic Engineers in Israel, Eliat, December 2008*. Eliat: IEEE, pp. 760–764.
- Zou, D., Grote, L., Peker, Y., Lindblad, U. and Hedner, J. (2006) Validation a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography. *Sleep*, 29, pp. 367–374.
- Zweig, M.H. and Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39, pp. 561–577.
- Zywietz, C.W., von Einem, V., Widiger, B. and Joseph, G. (2004) ECG analysis for sleep apnea detection. *Methods of Information in Medicine*, 43, pp. 56–59.