

## Estructuración bibliográfica acerca de Multiview learning para clasificación de imágenes

### Bibliographic structure about Multiview learning for image classification

Nancy Yaneth Gélvez-García<sup>1</sup>, Kevin C. Díaz-M.<sup>2</sup>, Carlos Enrique Montenegro-Marín<sup>3</sup>, Paulo Alonso Gaona-García<sup>4</sup>

**Resumen:** En el presente artículo se muestra una revisión bibliográfica de la literatura académica relacionada con “Clasificación de imágenes con Multiview learning” junto con un análisis de la información presente en cada una de las fuentes bibliográficas revisadas, con la finalidad de proponer una base conceptual, teórica y estadística para trabajos de investigación que desarrollen o contengan esta temática. De igual manera se presenta brevemente la forma en la que se aborda el MVL en los diferentes escenarios de aplicación tanto académicos como prácticos.

**Palabras clave:** Análisis bibliométrico, Clasificación de imágenes, Estado del arte, Multiview Learning.

**Abstract:** This article shows a bibliographic review of the academic literature related to "Image classification with Multiview learning" together with an analysis of the information present in each

---

<sup>1</sup> MSc. en Ciencias en sistemas de Información y comunicación, Universidad Distrital Francisco José de Caldas, Colombia, Bogotá D.C., Colombia. [nygelvezg@udistrital.edu.co](mailto:nygelvezg@udistrital.edu.co) <https://orcid.org/0000-0003-3334-6959>

<sup>2</sup> Ingeniero de Sistemas, Universidad Distrital Francisco José de Caldas, Colombia. [kcdiazm@correo.udistrital.edu.co](mailto:kcdiazm@correo.udistrital.edu.co) <https://orcid.org/0000-0003-3966-6275>

<sup>3</sup> PhD. en informática, Universidad Pontificia de Salamanca, España, Salamanca. [cemontenegrom@udistrital.edu.co](mailto:cemontenegrom@udistrital.edu.co) <https://orcid.org/0000-0002-3608-7158>

<sup>4</sup> PhD. en Ingeniería de la Información y del Conocimiento, Universidad de Alcalá, España, Alcalá., Colombia. [pagaonag@udistrital.edu.co](mailto:pagaonag@udistrital.edu.co) <https://orcid.org/0000-0002-8758-1412>

of the reviewed bibliographic sources, to propose a conceptual basis, theoretical and statistical for research works that develop or contain this theme. In the same way, the way in which the MVL is approached in the different application scenarios, both academic and practical, is briefly presented.

**Keywords:** Bibliometric analysis, Image classification, State of the art, Multiview Learning.

## **1. Introducción**

Cuando se realiza una investigación académica en cualquier rama del conocimiento, es común que se comience con la revisión bibliográfica de la temática, más conocida como investigación del estado del arte; esto con la finalidad de establecer un conocimiento base, para averiguar que se ha venido trabajando en esa área del conocimiento por otros investigadores, comprender las diversas vertientes y tendencias que se están desarrollando, tener una concepción de cuáles son los retos en el campo de estudio y sobre todo inspirar nuevas ideas que se enfoquen en mejorar o llevar más allá el conocimiento que se está trabajando.

Este documento condensa una investigación enfocada en Multiview Learning (Aprendizaje de múltiples vistas) en primer lugar de forma general y en segundo lugar aplicado análisis de imágenes. Dentro de este campo se encuentran múltiples proyectos, problemáticas, y mejoras que ya se han trabajado, por tal motivo es importante realizar esta revisión previa para contextualizar y establecer los límites que se tendrán dentro de la investigación.

Para esta tarea se establece una metodología, una definición y delimitación del problema (Multiview learning para la clasificación de imágenes), la cual describe de manera clara y precisa la manera en que se trabajará con las diversas fuentes bibliográficas y su contenido para así lograr

guiar al investigador en la tarea de establecer los fundamentos teóricos básicos y necesarios para su proyecto.

## **2. Procedimiento para el análisis bibliográfico de clasificación de imágenes con multiview learning.**

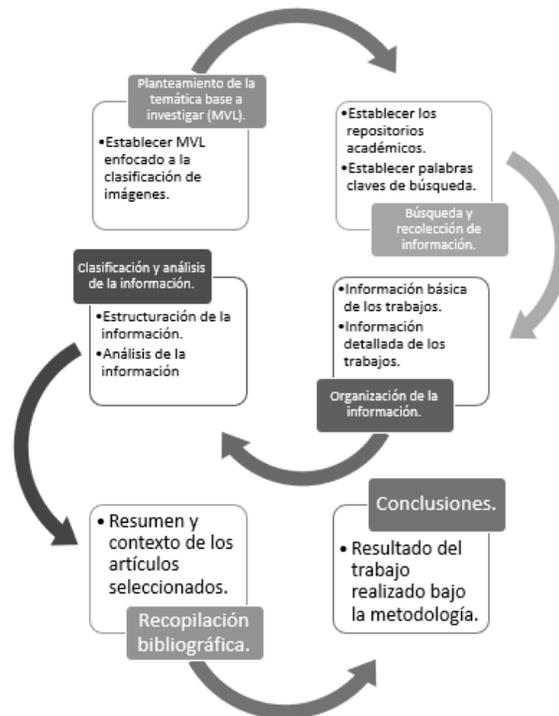
### **2.1. definición de la metodología de revisión bibliográfica enfoca al multiview learning.**

Para construir la metodología a trabajar se establecen unos pasos a seguir los cuales serán evaluados, esto se hace teniendo en cuenta que el multiview learning (de ahora en adelante MVL) es una rama del machine learning que tiene un especial enfoque por mejorar el rendimiento y sus resultados al trabajar múltiples vistas, “El aprendizaje de múltiples vistas también se conoce como fusión de datos o integración de datos de múltiples conjuntos de características.” [1] y que a su vez es un concepto relativamente nuevo que se ha ido explorando durante la última década con diferentes enfoques, algoritmos y estrategias para el entrenamiento que dependen de los problemas a resolver.

- Planteamiento de la temática base a investigar (MVL)
- Búsqueda general del tema y recolección de múltiples elementos a revisar
- Organización de información relevante como:
  - Título
  - Abstract
  - Palabras Clave
  - Revista
  - Fecha
  - Enlace de la web

- Técnicas o metodologías usadas
- Temática que aborda
- Clasificación y análisis del contenido en variables como:
  - Fechas de publicación
  - Frecuencia de técnicas o metodologías que se usan para el multiview learning
  - Proporción de las temáticas que se ven trabajadas con multiview learning
- Recopilación bibliográfica
- Conclusiones

En la Figura 1 se presenta de forma gráfica el planteamiento de la metodología.



**Figura 1.** Diagrama de la metodología de revisión bibliográfica planteada.

**Fuente:** Elaboración propia

Estos pasos se establecen basados en una previa revisión de información bibliográfica del tema, en donde se destacan aspectos como la técnica y la temática que aborda cada artículo. Un aspecto importante es el año de publicación del trabajo investigativo, ya que, es recomendable tener información que haya sido publicada recientemente, pero a su vez no se deben descartar posibles artículos que a pesar de la fecha de publicación pueden ser relevantes por su contenido único.

## **2.2. Planteamiento de la temática.**

Para el proceso de investigación se estableció la temática de MVL enfocada al manejo de imágenes, esto con la finalidad de dar soporte a desarrollos en donde se necesite la clasificación de imágenes.

La temática delimita el espectro del trabajo, ya que, MVL también puede ser utilizado para trabajar con conjuntos de datos planos, bases de datos, entre otros conjuntos de datos que hacen que la búsqueda sea más extensa y menos enfocada al objetivo.

Al ser un espectro de trabajo enfocado a las imágenes se ha de tener en cuenta los tipos de imágenes, características particulares, temáticas en las cuales se enfocan, aplicaciones en las que se usan etc., siendo así un amplio espectro sobre el cual MVL se usa para afrontar los escenarios particulares que se presentan.

## **2.3. Búsqueda y recolección de información.**

Para la búsqueda de información se utilizaron varios repositorios de información académica, entre los que se destacan:

- Science Direct
- Springer

- Oxford Academic
- IEEE Explore Digital Library
- Google Academic

Esto con la finalidad de obtener la información más relevante, se manejaron palabras claves como:

- Multiview learning
- Image Classification
- Multiview Classification
- Deep learning classification image

De estas bases de datos se sacaron una cantidad de 50 artículos los cuales a primera vista cumplen con la idea básica de que pertenezcan a la temática a trabajar.

## **2.4. Organización de la información**

Para el control de la información se establecieron los parámetros iniciales como lo son:

- Índice del artículo
- Nombre del artículo
- Abstract (Resumen)
- Enlace de consulta

Esta primera revisión se hace con el fin de filtrar los artículos con una lectura y contextualización de la temática trabajada en cada proyecto. Una vez se tengan seleccionados los trabajos que entrarán en la clasificación y análisis, es importante tener en cuenta otros aspectos relevantes como:

- Palabras claves: Da un contexto de los temas que se están trabajando en el artículo, en este caso se establecieron las siguientes palabras claves: Multiview, machine learning, classification, image, Deep learning, etc.
- Revista: es el medio en el cual se publica el artículo o un trabajo académico, la revista también puede ser memorias de conferencias académicas. Cada una de ellas tiene una clasificación por cuartiles los cuales escalafonan la revista dando una mejor percepción de la calidad del artículo [2]. En este caso se buscaron en fuentes de información con enfoque a Revistas académicas con cuartiles tipo Q1 y Q2, pero también se tienen en cuenta otras posibles fuentes que puedan aportar a la investigación. La información de cada revista se puede consultar en <https://www.scimagojr.com/> [3]
- Tipo de artículo: Es una clasificación dada por el contenido del artículo en el cual se establece unas etiquetas para el artículo dependiendo del enfoque que tiene, por ejemplo, un artículo que su foco principal es una revisión bibliográfica, este tendrá dicha etiqueta “Revisión Bibliográfica”; si el artículo maneja una sola técnica como SVM este tendrá la etiqueta “Algoritmos”, y de esta forma clasificar el conocimiento interno del artículo.
- Temática del artículo: Esta es una clasificación basada en que áreas del conocimiento se usa el multiview (En el caso de que sea un artículo enfocado a la aplicación)
- Técnicas, metodologías o Algoritmos: Esta es una clasificación de qué temas en específico se están usando en cada uno de los artículos, esto con la finalidad de tener un detalle de los planteamientos teóricos que se están trabajando para resolver el caso de investigación particular de cada artículo.

## 2.5. Clasificación y análisis de la información.

Para esta clasificación y posterior análisis se estableció una tabla de información en la cual se condensan aspectos útiles para la consulta rápida como para establecer el nivel de importancia que se le puede asignar a un artículo dentro del conjunto investigado.

Esta tabla cuenta con los siguientes campos:

<b>Campo</b>	<b>Descripción del campo</b>
<b>Id Artículo</b>	Identificador del artículo
<b>Título Original</b>	Título en el idioma original
<b>Título español</b>	Título en español
<b>Abstract Original</b>	Abstract en el idioma original
<b>Abstract español</b>	Abstract en español
<b>Revista</b>	Nombre de la revista del artículo
<b>Calificación</b>	Calificación científica que tiene la revista al momento de realizar el artículo (Q1, Q2...)
<b>Categoría Revista</b>	Enfoque en conocimiento que tiene la revista
<b>Fecha Artículo</b>	Fecha de publicación del artículo
<b>URL</b>	Dirección URL donde es tomado el artículo
<b>Valoración Investigación</b>	Calificación de 1 a 3 (1 más importante, 3 menos importante) conforme a que tanto puede aportar el artículo para la investigación
<b>Tipo de artículo</b>	Categorización del tipo de artículo

**Tabla 1.** Descripción de los campos de la tabla principal de la información

**Fuente:** Elaboración propia

También se crearon 3 tablas auxiliares (Tabla II, Tabla III y Tabla IV) en las cuales se relacionan información importante que se puede extraer de cada artículo y que brinda valor para el análisis, estas tablas están estructuradas de la siguiente forma:

<b>Campo</b>	<b>Descripción del campo</b>
<b>Id Artículo</b>	Identificador del artículo
<b>PALABRAS CLAVE</b>	Palabra clave que está contenida en el artículo

**Tabla 2.** Descripción de los campos de la tabla de Palabras Clave

**Fuente:** Elaboración propia

<b>Campo</b>	<b>Descripción del campo</b>
<b>Id Artículo</b>	Identificador del artículo
<b>TEMÁTICA</b>	Temática que es abordada en el artículo

**Tabla 3.** Descripción de los campos de la tabla de temática

**Fuente:** Elaboración propia

<b>Tabla IV</b>	
<b>Descripción de los campos de la tabla de Técnicas, metodologías o Algoritmos</b>	
<b>Campo</b>	<b>Descripción del campo</b>
<b>Id Artículo</b>	Identificador del artículo
<b>PALABRAS CLAVE</b>	Es la técnica, la metodología o el algoritmo que aparece en uso en el respectivo artículo.

**Tabla 4.** Descripción de los campos de la tabla de Técnicas, metodologías o Algoritmos

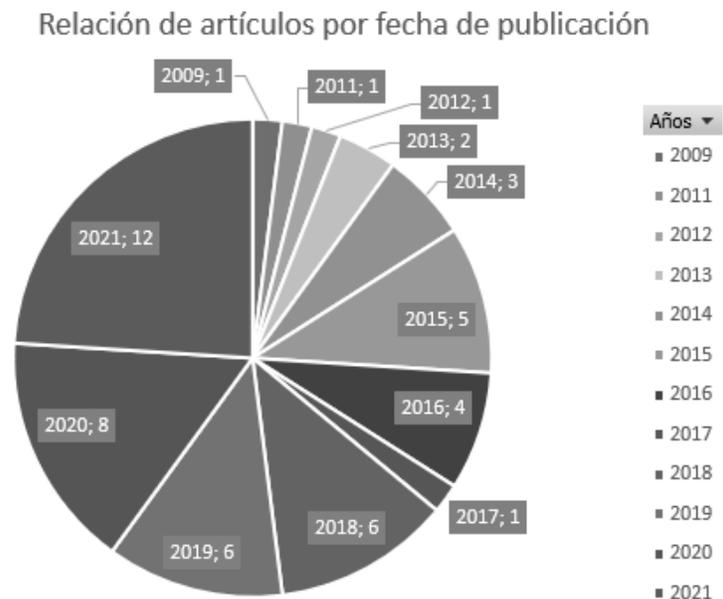
**Fuente:** Elaboración propia

Con esta estructura se procede a diligenciar la información conforme cada uno de los artículos que se tuvieron en cuenta, tanto por su título, por el abstract, por las palabras y por el posible aporte que pueda tener al tema de MVL.

A continuación, se realiza el análisis de la información teniendo en cuenta los por los siguientes aspectos:

### 2.5.1. Por fecha de publicación

Este aspecto se tiene en cuenta para determinar qué tan recientes son los artículos encontrados, y así poder tener en cuenta los trabajos que sean más recientes, de igual manera, determinar desde qué año se ha venido trabajando con MVL en la clasificación de imágenes. A manera de resultado se obtuvo (Figura 2):



**Figura 2.** Artículos por fecha de publicación.

**Fuente:** Elaboración propia

En la Figura 2 se puede observar que ha existido un crecimiento en la investigación de MVL en cuanto al análisis de imágenes se refiere, durante los últimos 5 años, y que la temática de MVL se viene trabajando desde el año 2009.

### 2.5.2 Por fecha de publicación

En este análisis se buscó establecer qué palabras son utilizadas de forma recurrente por los autores que trabajan MVL y sus temáticas asociadas. Con la información recopilada con la estructura de la tabla II se obtuvieron los siguientes resultados:



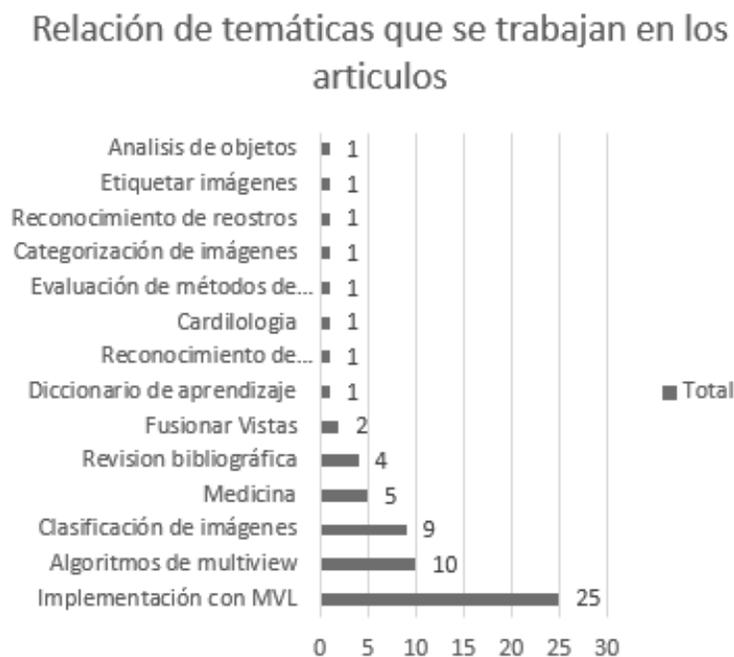
**Figura 3.** Gráfica de representación de las palabras clave (Keywords) por concurrencia en los artículos.

**Fuente:** Elaboración propia

En la Figura 3 se observa como en la mayoría de los artículos consultados se maneja el concepto de Multiview learning, seguida de conceptos relacionados como Deep learning, image classification y Convolutional neural network, los cuales son los temas más relacionados a la hora de trabajar el MVL enfocado al tratamiento de imágenes. NOTA: Solo se presentan en la gráfica los conceptos que aparecen más de una vez.

### 2.5.3. Por técnicas y/o metodologías que usan el multiview

En este análisis se buscó establecer qué temáticas son trabajadas en los artículos, teniendo en cuenta que un artículo puede tener más de una temática. Un ejemplo de ello es utilizar el MVL enfocado a reconocer violencia en contenido multimedia. Con la información recopilada con la estructura de la tabla III se logró llegar a los siguientes resultados



**Figura 4.** Gráfica de relación de temáticas por concurrencia en los artículos.

**Fuente:** Elaboración propia

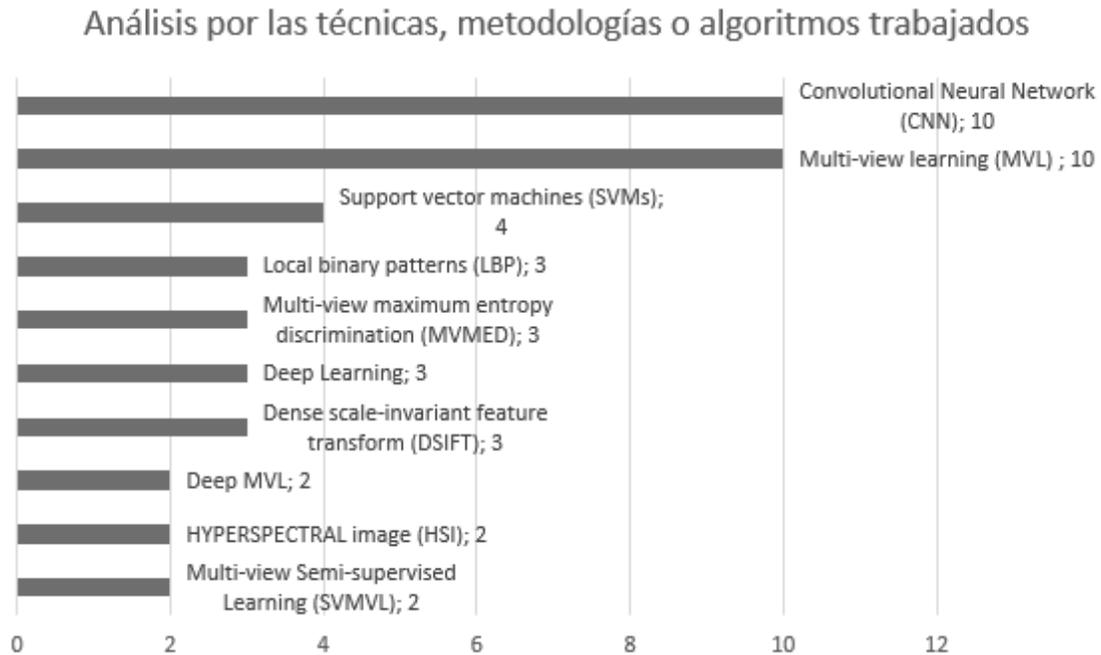
Como se evidencia en la Figura 4, la mayoría de los artículos se enfocan en la implementación del MVL para el manejo de un aspecto de mejora en la clasificación o en la optimización de algún algoritmo, seguido de algoritmos que se enfocan específicamente el manejo de MVL y posteriormente en la clasificación de imágenes, cabe resaltar que aparecen temas como la medicina, el análisis de objetos, reconocimiento de violencia o fusión de vistas.

#### ***2.5.4. Por técnicas, metodologías o algoritmos que trabajan con el multiview***

Por último, un tema importante es la utilización de una técnica, metodología o algoritmo en conjunto con otros algoritmos o un algoritmo como parte del MVL, esto es algo usual en varios articulo o en su defecto se evidencia en los artículos como que hay múltiples técnicas que se usan para abordar diversos problemas que a pesar de que tengan una misma temática como lo son las imágenes su propósito es diferente, Entre las que se destacan el MVL como base fundamental de la mayoría de los artículos en conjunto con las CNN y las SVMs, también es importante recalcar que entre todas las técnicas se encuentran algoritmos de bajo nivel como DSIFT, SIFT, LBP y HOG para extraer características de las imágenes, y también aparecen variaciones o extensiones de MVL como Deep MVL, MVL semi supervisado, MVL CNN, etc.

Y para finalizar este análisis se evidencia que hay una gran variedad de temas que se ven relacionados con el MVL lo cual muestra una gran aceptación de la idea de correlacionar, fusionar y establecer coincidencias entre múltiples vistas ya sea de imágenes o etiquetas de las imágenes para así lograr entrenar sistemas que sean óptimos mediante la utilización de algoritmos y funciones matemáticas que cada uno de los artículos aborda de la forma más minuciosa y rigurosa posible.

NOTA: Solo se presentan en la gráfica los conceptos que aparecen más de una vez.



**Figura 5.** Gráfica de análisis de técnicas, metodologías o algoritmos por concurrencia en los artículos.

**Fuente:** Elaboración propia

Con la información de la Figura 5 se denota que hay temas que son recurrentes en el trabajo con MVL, un ejemplo de ello son los algoritmos de bajo nivel que se usan como base para la creación de las vistas y por ende forman parte fundamental de algunos procesos de MVL para el análisis de datos.

### **3. Artículos relevantes sobre MVL enfocado a imágenes**

Al momento de realizar la lectura de cada uno de los artículos se buscó rescatar la información más relevante de cada uno de ellos con la idea de poder hacer un breve resumen de su contenido y hacer referencia a dichos trabajos.

El primero de ellos realiza una clasificación de los algoritmos que se usan para el MVL en categorías, estableciendo 3 categorías por forma de entrenamiento de los algoritmos, estas son:

- Algoritmos de estilo de co-entrenamiento
- Algoritmos de estilo de co-regularización
- Algoritmos de estilo de consistencia de márgenes

El trabajo titulado “Multi-view learning overview: Recent progress and new challenges and new challenges.” [1] , se presentan algunos algoritmos para cada una de las categorías y su respectivo funcionamiento siendo un punto de referencia clave al ser un artículo de revisión bibliográfica enfocado al MVL.

Un artículo interesante el cual sirve como referencia para el manejo de imágenes es el artículo “Multiview discriminative marginal metric learning for makeup face verification” [4] ya que toma como referencia algunos algoritmos de bajo nivel como los son HOG, SIFT, TPLBP, LBP Y FBD , con el fin de determinar diferencias entre dos imágenes (una imagen con maquillaje y otra sin el maquillaje, esto en la misma persona), posteriormente aplican el principio básico de multiview discriminative marginal metric learning (MDMML) para aplicar el modelo de clasificación.

Al realizar este estudio se encontraron algunos algoritmos interesantes y relevantes, los cuales son: algoritmos de bajo nivel, sobre los cuales también se soporta otro algoritmo de MVL llamado multi-view maximum entropy discriminant (MVMED+) basado en: model for learning with different numbers of views. o en español (modelo discriminante de máxima entropía de múltiples vistas (MVMED+)); este fue trabajado en el artículo “Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features” [5], en el cual se trabaja bajo el objetivo de lograr reconocer imágenes que representan violencia física,

para ello se soportan en los algoritmos de bajo nivel que luego se unirán con MVL, estos algoritmos son:

- DSIFT Feature: una variante del algoritmo SIFT que busca manejar descriptores más densos de los que se manejan en el algoritmo de SIFT, para ello realizan más iteraciones en un tamaño específico de la grilla que se usa en SIFT, para así lograr extraer características invariantes de la imagen.
- HOG Features: el Histograma orientado por gradiente es un descriptor de características usado para la detección de objetos, este se usó en conjunto con una máquina de soporte vectorial o support vector machine (SVM) para obtener el histograma de cada píxel y así normalizar las imágenes.
- LBP Features: las características de patrones binarios locales es un descriptor de características locales basados en la textura en donde se compara un píxel en conjunto con sus vecinos circundantes para así obtener un ponderado de esa textura local y establecer un valor en una escala de grises.

Con la contextualización anterior en el artículo se introduce de forma matemática y algorítmica el concepto de la discriminación por máxima entropía en donde a grandes rasgos se trata de optimizar una función objetivo en la cual se incluyen los resultados de cada uno de los algoritmos de bajo nivel trabajados previamente. Adicionalmente establecer los valores requeridos para un SMO (Sequential minimal optimization). Todo esto en conjunción logra establecer unos resultados para un conjunto de imágenes en comparación con las imágenes tratadas de forma individual o en combinaciones pares de las vistas como se muestra en la Figura 6 del artículo.

Views	Life	Scene	Physical
view1	77.82 ± 0.47	89.94 ± 0.62	88.35 ± 0.70
view2	78.81 ± 0.45	92.38 ± 0.52	91.70 ± 0.47
view3	85.99 ± 0.57	92.26 ± 0.47	88.74 ± 0.95
view1 + view2	80.29 ± 0.63	94.06 ± 0.54	92.58 ± 0.63
view1 + view3	85.56 ± 0.71	93.07 ± 0.36	90.60 ± 0.60
view2 + view3	86.07 ± 0.46	94.29 ± 0.63	92.52 ± 0.70
view1 + view2 + view3	<b>86.12 ± 0.42</b>	<b>95.10 ± 0.52</b>	<b>92.60 ± 0.41</b>

**Figura 6.** Tabla de “The accuracy of MVMED<sup>+</sup> with different number of views.

**Fuente:** [5]

Otro trabajo importante es el artículo denominado “A Survey on Multi-view Learning” [6], en el cual se presentan los dos principios de relación entre las múltiples vistas, los cuales son:

- El principio de consenso: el cual tiene como objetivo, manejar la información de las diferentes vistas y maximizar el “consenso” o la igualdad de la información entre las vistas. Esto muestra la conexión entre estas vistas bajo algunas superposiciones y leves desigualdades que puedan existir al maximizar la correlación de la información. Esto también se puede aplicar utilizando una minimización dependiendo del caso en concreto.
- El principio de complemento: Este principio se basa en que cada vista puede contener información que otra vista no puede contener mostrando diferente información sobre un mismo objeto de estudio en capas que de por sí manejan información que en esencia no es igual. Una forma de manejar este concepto es utilizar múltiples clasificadores para extraer información complementaria y que de forma paralela se puedan entrenar para hallar comportamientos que unan la información pero que a su vez den el complemento esperado.

Continuando con el artículo se plantea el proceso de creación de uso básico de las técnicas de MVL, partiendo de la manera en la que se deben crear las vistas, cómo se deberían combinar, manejando

técnicas de co-entrenamiento y estableciendo algunos algoritmos de múltiple aprendizaje que se han trabajado a través de la evolución de MVL, como lo son Multiple Kernel Learning (MKL) y el Subspace Learning-based Approaches que a su vez exploran todos los algoritmos y técnicas que cada uno de estos conllevan.

En el artículo titulado “Multi-View Image Classification With Visual, Semantic and View Consistency” [7], se trabaja la idea de consenso entre las múltiples vistas, esto implica que se busca que entre las vistas que se están manejado existen características que sean iguales o semejantes, para ello lo que realizaron fue aplicar cada una de las coherencias en vistas separadas y posteriormente utilizar clasificadores pre entrenados para combinar las coherencias, esto con un método denominado “Linear Multi-View Combination”, posteriormente comienzan a experimentar con varios datasets para ver comportamiento, y objetos que clasifica; también para realizar una comparación en desempeño conforme a otros métodos que se usaron como su línea base de comparación lo anterior se puede observar en la Figura 7.

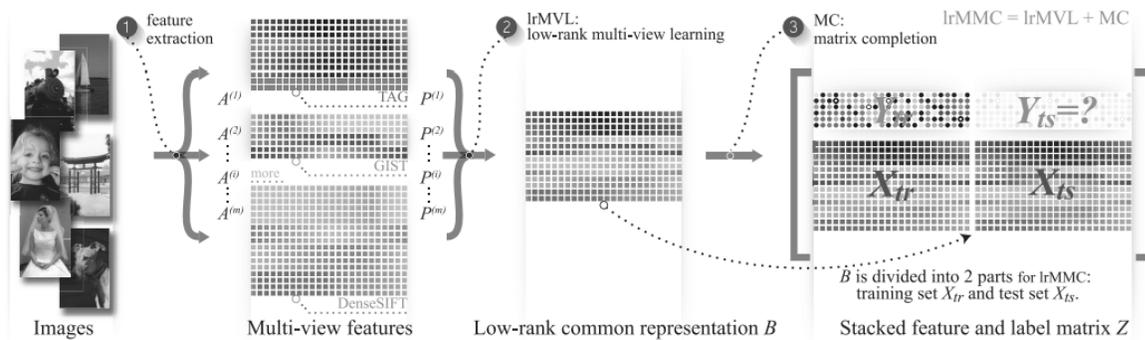
Methods	BB	Performance (%)
FC-VGG [45]	no	70.4
bilinear CNN [61]	no	84.1
LRBP[62]	no	84.2
PR-CNN [16]	yes	73.5
Triplet-A [63]	yes	80.7
STN [64]	no	84.1
BoostCNN [65]	no	86.2
VGG-D [66]	yes	82.0
RBF [73]	no	79.0
AGAL [74]	yes	85.5
VSVC(average)	no	87.5
VSVC	no	87.9

**Figura 7.** Tabla de “Performance Comparison os VSVC another Baselne Methods.”

**Fuente:** [7].

Un artículo que es importante tener en cuenta por establecer un marco de aprendizaje semi supervisado es el artículo “Multiview semi-supervised learning for image classification” [8], en el cual toman la información de imágenes que están pre etiquetadas, se establece un sistema de etiquetado de imágenes para entrenar y clasificar las imágenes de forma independiente y posteriormente con estas vistas se usa un principio de mínima entropía para establecer la categoría de la etiqueta que debería llevar las imágenes y así poder clasificar la imagen de forma certera. Algo similar se ve en el artículo de la conferencia de la IEEE ICME del 2009 llamado “Multi-view multi-label active learning for image classification” [9], el cual realiza un aprendizaje activo de múltiples vistas estableciendo una serie de incertidumbres en las vistas para posteriormente fusionarse a partir de una muestra-etiquetada que se tiene como base.

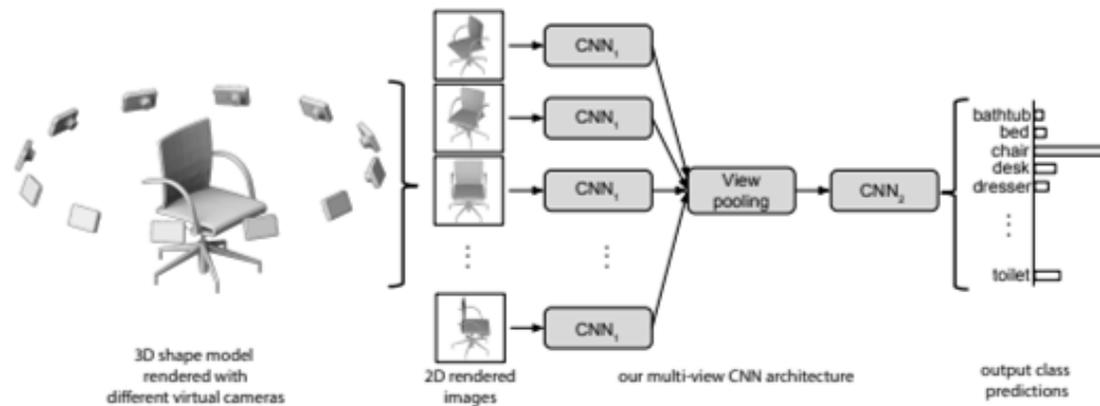
En el tema de multiview es algo recurrente buscar la mejora de los algoritmos que ya se están utilizando esto conforme a la complejidad en los tiempos de clasificación y procesamiento, un artículo que habla de ello es “Low-Rank Multi-View Learning in Matrix Completion for Multi-Label Image Classification” [10] en este artículo se basan en la complejión de matrices para realizar la clasificación de múltiples etiquetas, ya que es un método que se ha trabajado, aquí también se trabajan algoritmos de bajo nivel como lo son DSIFT, TAG, GIST para tener las múltiples vistas de características de las imágenes, para posteriormente aplicar el modelo planteado “low-rank multi-view matrix completion (lrMMC) model” y por último se utilizó la matriz de complejión para clasificar la imagen, tal como se muestra en la Figura 8.



**Figura 8.** Muestra del proceso de multiview con IrMVL y MC.

**Fuente:** [10].

Un desafío que se presenta en la literatura consiste en el manejo de imágenes en 2D que representan objetos en tercera dimensión (3D) para lo cual se ha trabajado el MVL apoyado en técnicas de redes neuronales convolucionales, esto se plasma en el trabajo titulado “Multi-view Convolutional Neural Networks for 3D Shape Recognition” [11], en este artículo trabajan las imágenes de perspectivas de una objeto de forma separada, cada una de estas imágenes entra a una red neuronal, el resultado de cada una de ellas se agrupa y posteriormente se pasa la información a una red neuronal convolucional, para lograr identificar que estas vistas coinciden con un objeto en específico (una imagen de la figura 3D que está relacionada con múltiples imágenes 2D), esto se puede ver representado en la Figura 9.



**Figura 9.** Multi-view CNN for 3D shape recognition.

**Fuente:** [11].

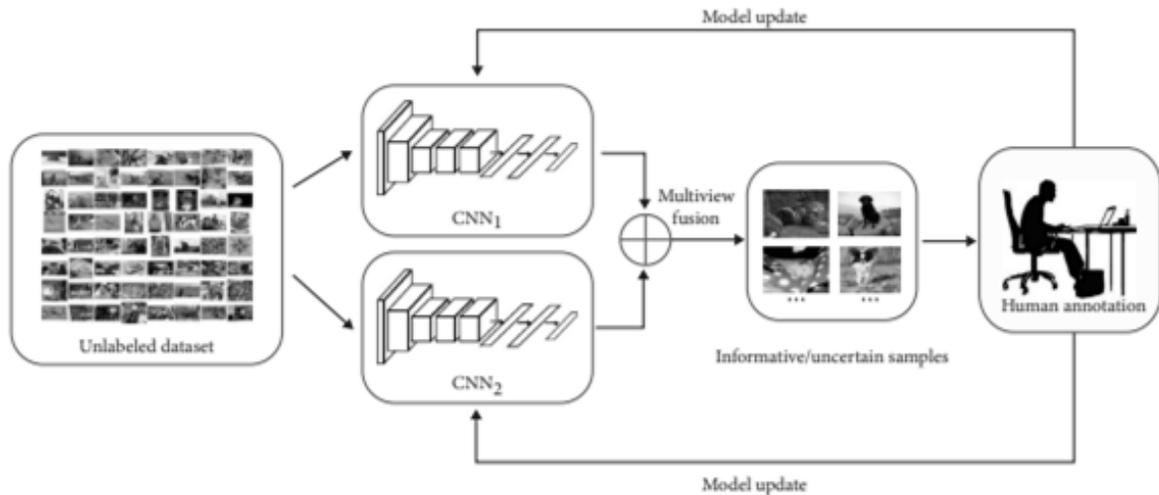
Otro gran desafío que se trabaja en el MVL costa de buscar formas de mejorar el rendimiento de las clasificaciones, esto ha generado que se integren procesos con ramas de la inteligencia artificial como lo son las redes neuronales y más específicamente las redes neuronales convolucionales, esto se muestra en el artículo “Multi-view classification with convolutional neural networks” [12], en este trabajo se es utilizan las redes neuronales convolucionales para extraer la información de los objetos (Plantas insectos y autos) para así lograr implementar 3 formas diferentes de fusionar las múltiples vistas, estas son (1) fusionar mapas de características convolucionales a diferentes profundidades de red (Early fusion); (2) fusión de representaciones latentes de cuellos de botella antes de la clasificación (Late fusion); y (3) fusión de puntuación (Score fusion). De esta forma estiman cual es la mejor forma de combinar estas múltiples vistas para los entornos de dominio planteados.

Como se ha visto a lo largo de los artículos que se han revisado se manejan una gran variedad de algoritmos enfocados a diferentes partes del proceso de MVL, una parte del multiview es la fusión, concatenación o correlación de las múltiples vistas generadas desde los métodos establecidos en

cada investigación, la forma de combinar todas estas vistas se realiza con redes neuronales, en este nuevo artículo llamado “Deep Multi-View Learning using Neuron-Wise Correlation-Maximizing Regularizers” [13], en dicho artículo se toma como base para la fusión de las vistas las redes neuronales profundas (DNN) en su forma básica, para después introducir el concepto a trabajar: Regularizadores por correlación (CorrReg) estos se aplican en las capas de fusión para regularizar los parámetros y posteriormente optimizar la validación del set de datos.

Continuando con la línea de la clasificación de imágenes utilizando redes neuronales, en el siguiente artículo titulado “A Deep Multiview Active Learning for Large-Scale Image Classification” [14], se presenta el “Deep multiview active learning (DMAL) framework” el cual busca mejorar el aprendizaje activo de múltiples vistas (Multiview active learning (MAL)) esto realizando pruebas con variaciones de cantidad y tamaño de las imágenes, estas acceden a dos redes neuronales convolucionales generando dos capas que se unen usando multiview learning para posteriormente tomar anotaciones de una persona y hacer retroalimentación activa a las redes neuronales y así complementar la información para su mejora en la clasificación.

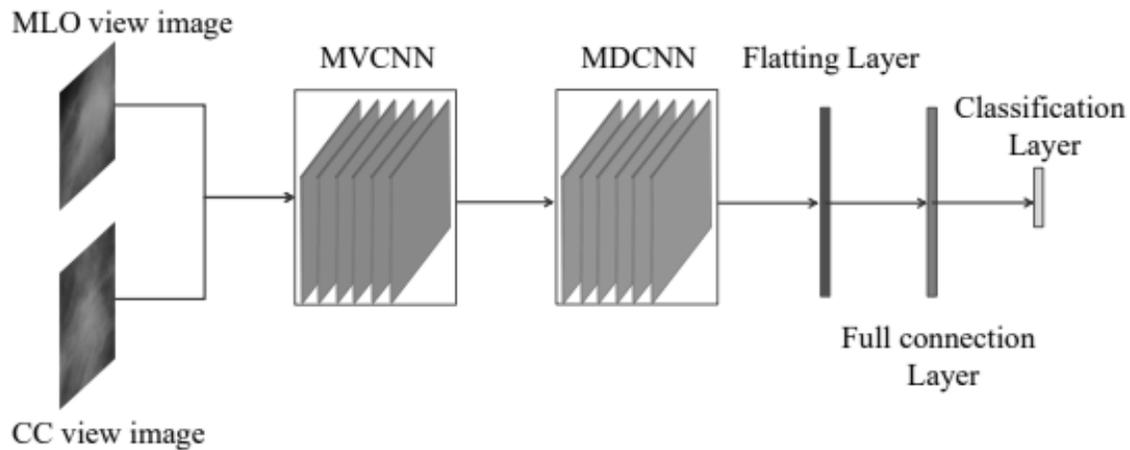
En la Figura 10, se muestra como es el proceso en el cual las imágenes son analizadas en dos redes neuronales convolucionales, el resultado entra al modelo de multiview fusión el cual analiza las imágenes y define cuáles puede entender y cuáles no, posteriormente esta información se le presenta a una persona la cual hace una retroalimentación a las redes neuronales convolucionales, esto con la finalidad de mejorar los resultados que se obtienen en cada una de las redes neuronales convolucionales.



**Figura 10.** Muestra del proceso de multiview learning active.

**Fuente:** [14].

En sí el concepto de redes convolucionales utilizadas en conjunto con el MVL es algo que se ha ido trabajando pero que a su vez ha tenido implementaciones interesantes en ámbitos como la medicina uno de ellos es la clasificación de imágenes mamográficas como se presenta en el artículo “Multi-View Convolutional Neural Networks for Mammographic Image Classification” [15] ,aquí se presenta como utilizar dos técnicas para obtener las vistas de las imágenes mamográficas (mediolateral oblique (MLO) y Craniocaudal (CC)). Una vez obtenidos estos dos tipos de imágenes proceden a utilizar una arquitectura MVMDCNN en la cual aparecen dos subredes una convolucional y una vista oblicuo medio lateral para poder extraer las características necesarias de las imágenes y así al final utilizar funciones de pérdida de entropía cruzada para el proceso de entrenamiento profundo del modelo. Esto se muestra en la Figura 11.



**Figura 11.** Muestra de la arquitectura MVDCNN.

**Fuente:** [15].

Es importante recalcar que detrás de estas estructuras de multiview hay una parte teórica y algorítmica que se basa en diversos modelos es por ello que se puede dar lugar a adentrarnos en diversos algoritmos frameworks que se trabajan en el MVL, un artículo que trabaja el multiview con un framework de Discriminación por máxima entropía es el trabajo denominado: “Multi-View Maximum Entropy Discrimination” [16], en dicho trabajo se manejan conceptos como Maximum entropy discrimination (MED) implementados con support vector machines (SVMs) para que al usar las múltiples vistas se pueda establecer un margen de clasificación idéntico entre las múltiples vistas, es por ello que realizan tres pruebas con sets de pruebas y al final se observa como el modelo denominado MVMEED logra una mejoría con respecto a otros métodos.

Otro artículo que se enfoca en un teorema para ser implementado en conjunto con el MVL es “PAC-Bayes analysis of multi-view learning” [17], en el cual se toma como base los límites PAC-Bayes, posteriormente se realizan diversas variaciones del algoritmo con máquinas de soporte

vectorial SVM semi supervisadas para posteriormente evaluar estas variaciones en unos sets de datos y encontrar los rangos en los cuales se comporta mejor una variación que otra.

MVL es una rama de investigación que al integrarse en el aprendizaje automático genera nuevas mejoras en diversas temáticas, una de ellas es la temática del aprendizaje de múltiples fuentes, tal como se denota en el artículo “Multi-source Transfer Learning with Multi-view Adaboost” [18], en el cual MVL es una parte fundamental para complementar el algoritmo Adaboost, y de esta manera lograr que las diversas fuentes de información que recibe el algoritmo se puedan manejar generando el nuevo algoritmo MsTL-MvAdaboost.

Es amplio el trabajo que se ha realizado en los últimos años donde interviene MVL, uno de los trabajos académicos en los cuales se busca adaptar los dominios que no se encuentran etiquetados a dominios que ya se encuentran adaptados tomando como base el aprendizaje de multiview que ya se tenga, es el trabajado en el artículo “Robust Multi-view Representation: A Unified Perspective from Multi-view Learning to Domain Adaption” [19], en dicho artículo se presenta una revisión de algoritmos enfocados a los clústeres de múltiples vistas, el aprendizaje de múltiples vistas (Multi-view) y como se pueden adaptar los dominios mediante algoritmos de transferencia del aprendizaje.

Una idea que se trabajó con MVL es el manejo de diccionarios de aprendizaje, en donde se utilizan estos dichos diccionarios para definir y encontrar características y así poder resolver problemas discriminando las imágenes como tal, esta técnica al ser usada tiene algunos aspectos a mejorar como el rendimiento el manejo del ruido en múltiples escenarios etc, es aquí donde se centra el artículo titulado: “Multi-view low-rank dictionary learning for image classification” [20], ya que realizan un trabajo con MVL supervisado que aprende sub diccionarios y aplica una fórmula de

optimización para denotar la data presente en el mapa de la imagen y posteriormente correlacionarla discriminando posibles errores pero a un nivel bajo ya que regulariza algunos términos y haciendo más precisa la correlación.

Como se ha ido viendo en esta revisión es usual que el MVL se relacione con otras técnicas de aprendizaje automático, para el caso del artículo “Weighted Multi-view Clustering with Feature Selection” [21], se manejó técnica de Clusterizado junto con una ponderación que se le dan a las instancias de cada una de las vistas para poder utilizar Multi-view k-means y hacer un proceso de optimización sobre las características que se van a agrupar y así obtener una vista representativa del conjunto de vistas y lo que buscan representar.

MVL no solo se usa para la clasificación sino también para reconstrucción de imágenes digitales tal como se presenta en el artículo: “Automatic 3D building reconstruction from multi-view aerial images with deep learning” [22], en el cual mediante Deep learning y un método de matching de multiview logran identificar las áreas de las construcciones, y mediante múltiples vistas de la imagen logran establecer variaciones en su forma y en su diseño para establecer posibles relieves que en conjunto con el Deep learning logran establecer las características de la construcción para después construir el modelo en 3d.

Y Continuando con las aplicaciones del MVL en imágenes geográficas se encontró el trabajo denominado “Forest Fire Recognition Based on Feature Extraction from Multi-View Images” [23], en donde se reconocen incendios extrayendo características como los componentes que hay en la imagen (humo y arboles), se extraen características de color como el rojo y amarillo del fuego y todo a partir de imágenes de las cuales se hacía un proceso dinámico de extracción, para posteriormente con redes neuronales gráficas y convolucionales depurar las imágenes que

contenían incendios y establecer las dos capas que se manejan en el proceso de segmentación y clasificación.

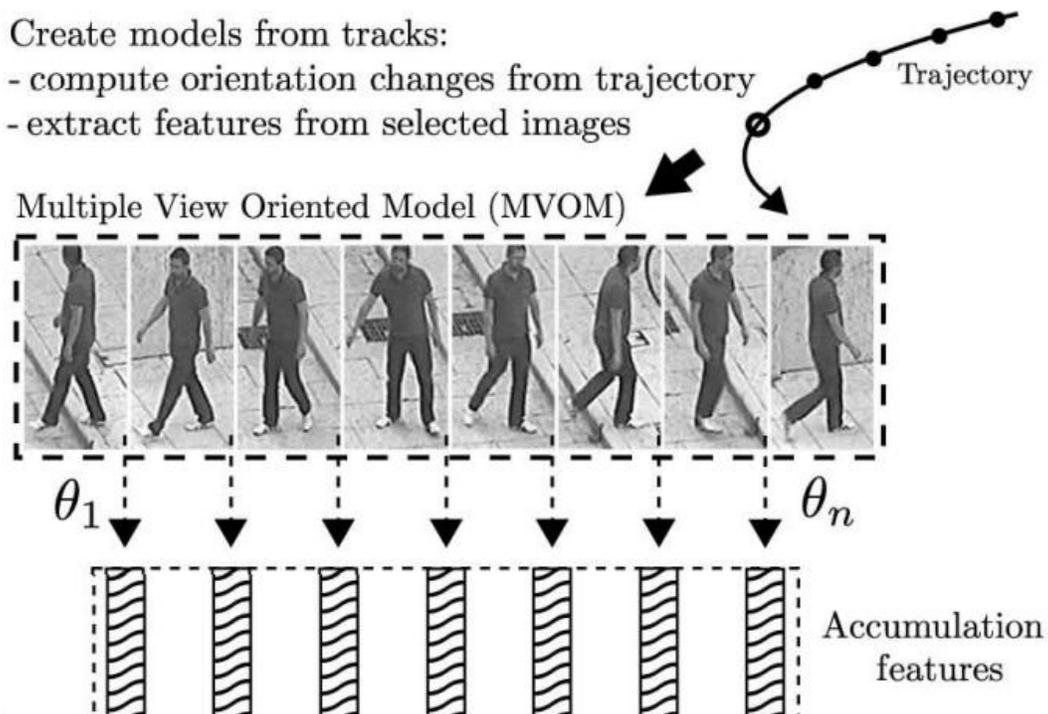
MVL conforme va avanzando su utilización se van encontrando temáticas a mejorar una de ellas es poder ser optimizada y utilizada en la reconstrucción de escenas que son más complejas de lo usual para los casos de aprendizaje y entrenamiento de un sistema clasificador de imágenes, en este caso el artículo “Differentiable Diffusion for Dense Depth Estimation from Multi-view Images” [24], siendo trabajado el concepto de Depth Diffusion para posteriormente diferenciar los cambios de los píxeles en la imagen y así ser optimizados mediante Multi-view Stereo a varios niveles de las imágenes.

Un artículo que es interesante revisar a nivel conceptual es el denominado “Deep multi-view learning methods: A review” [25], ya que en este hace una revisión de como el MVL ha tenido un gran impacto sobre todo utilizando técnicas de aprendizaje profundo, como análisis profundo con correlación, factorización de matrices, análisis de cuello de botella entre otros, siendo un trabajo en los que se encuentran algunas comparaciones y muestran algunas aplicaciones de Deep MVL.

También se revisó un trabajo de grado doctoral titulado “Multi-View Imaging: Depth Estimation And Enhancement” [26], en el cual se aborda el manejo de múltiples vistas para soportar el “Stereo Problem” el cual es un problema de correspondencia entre dos imágenes separadas por una distancia horizontal mínima, para ello se utilizan las redes neuronales junto con Multi-View Super-Resolution para lograr una mejora de las imágenes al hacer las correspondencias entre las imágenes.

Como se ha ido mencionando MVL no solo se unas para la clasificación de imágenes sino también se ha adentrado en el reconocimiento de múltiples imágenes para poder determinar a partir de un

conjunto de características similares, si un elemento de una imagen va variando con respecto a otra, pero sigue siendo el mismo, en este trabajo titulado “Multiple View Oriented Matching Algorithm for People Reidentification” [27], se evidencia como a partir de múltiples vistas en diferentes imágenes se busca obtener la identificación de las personas, su cambio de dirección y una actualización de la perspectiva actual, un ejemplo se encuentra en la Figura 12, en la cual se ejemplifica como es la idea base del modelo para su caracterización.



**Figura 12.** Muestra de Multiple View Oriented Matching Algorithm.

**Fuente:** [27].

MVL en su versión de Deep MVL se ha enfocado en mejorar tiempos y optimizar proceso, en el artículo “Deep Multiview Learning for Hyperspectral Image Classification” [28], se trabaja un método de aprendizaje profundo de múltiples vistas para abordar la clasificación de imágenes hiperespectrales (HSI), optimizando la cantidad de etiquetas que se usan para su respectiva

clasificación, esto extrayendo características invariantes en cada vista para posteriormente entrar a una red no supervisada para su entrenamiento y posterior clasificación con Deep MVL.

El aspecto hiper espectral ha tomado una relevancia en MVL ya que al manejar múltiples dimensionalidades se busca poder realizar una mejora en la clasificación y por ende en su rendimiento, es por ello que en el artículo denominado “Hyperspectral Image Classification via Spatial Window-Based Multiview Intact Feature Learning” [29], se propone el modelo spatial window-based multiview intact feature learning method (SWMIFL) para la clasificación de estas imágenes, extrayendo características de las imágenes y aplicando MMVL en vistas intactas para hacer el proceso de aprendizaje junto con Multiple morphological component analysis (MMCA) para lograr extraer las características necesarias para el sistema de entrenamiento con SVM.

Otra de las aplicaciones de Deep MVL ha sido la segmentación de múltiples clases para el manejo de ecografías como se presenta en “Multiview and Multiclass Image Segmentation using Deep Learning in Fetal Echocardiography” [30], en donde se hace un proceso de entrenamiento para la detección de cardiopatía congénita (CHD), esto a partir de una mejora en la V-Net con abandono espacial, normalización de grupos y supervisión profunda en segmentación de la información que a su vez se integra con las múltiples vistas para completar cualquier información sobre anomalías que no se esté teniendo en cuenta.

En el artículo “Multiview Similarity Learning for Robust Visual Clustering” [31], se plantea un método robusto de aprendizaje de similitud de múltiples vistas (RMvSL) el cual busca implementar rangos bajos en las muestras de las imágenes para eliminar ruido y valores atípicos, de tal forma que puedan establecer un esquema de múltiples vistas que se pueda trabajar bajo una función de

optimización que realiza la reconstrucción de los datos y sirva para el aprendizaje a la hora de fusionar diferentes gráficos con diversas vistas logrando un buen desempeño.

Como se ha visto a lo largo de esta revisión siempre se está en busca de una mejora con respecto al MVL, en el artículo “Multi-View Clustering and Semi-Supervised Classification with Adaptive Neighbours” [32], se presenta un modelo supervisado adaptativo denominado: Adaptive MULTiview SEMI-supervised model (AMUSE), el cual busca adaptarse a las ponderaciones en la regularización para así optimizar el algoritmo frente a un nuevo modelo, esto mediante la toma de los vecinos más cercanos a la matriz que se está trabajando para ser considerados en la función de minimización para así adaptarse a los vecinos cercanos, y posteriormente extienden el multiview con un modelo de clasificación semi supervisada.

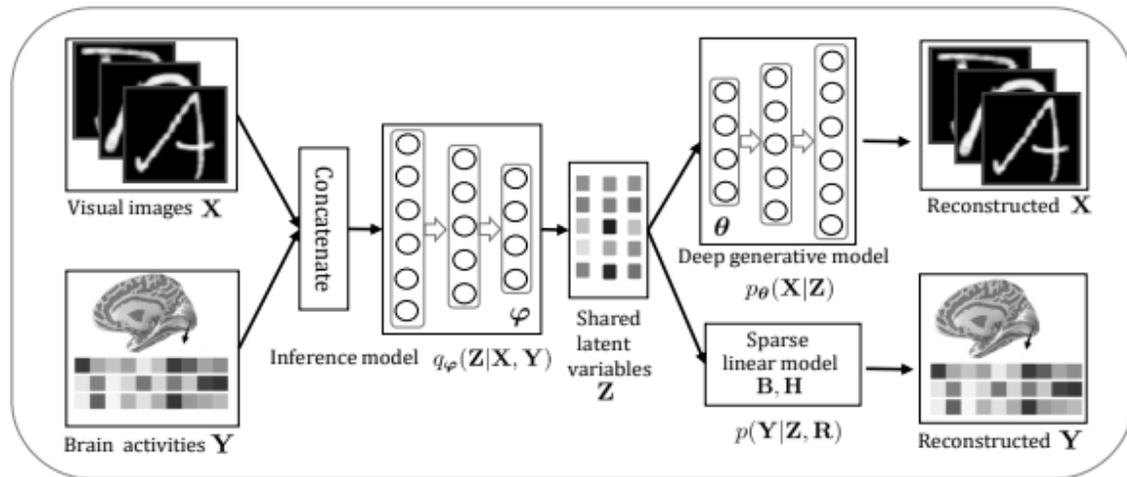
Otro ámbito al cual se ha aplicado MVL ha sido las tecnologías basadas en el internet, tal como se muestra en el artículo “Multiview Hessian Regularization for Image Annotation” [33], en donde se basan en el trabajo de la regularización laplaciana que se ha aplicado al aprendizaje automático semi supervisado, de ahí es donde se desarrolla la idea de regularización hessiana multivista (mHR), esto con la finalidad de realizar la combinación de forma óptima de múltiples regularizaciones Hessianas las cuales cada una es una vista particular de las instancias, y así al optimizar esa combinación se puede mejorar la clasificación de los datos.

En el trabajo titulado “Multi-view metric learning for multi-instance image classification” [34], se aborda la clasificación de imágenes utilizando la extracción de características de las imágenes, generando 3 vistas de la información las cuales son complementarias entre sí y se genera una función para hacer el cálculo entre las vistas y su respectivo aprendizaje, es aquí donde se propone

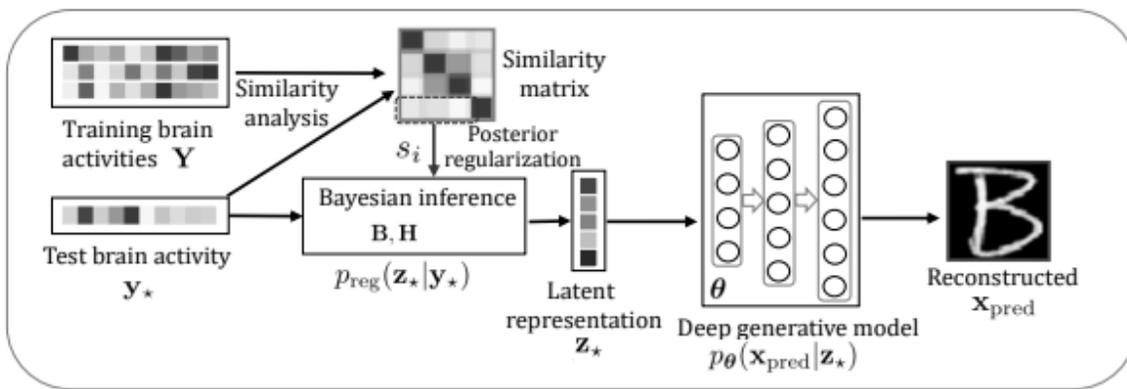
el MVML para optimizar las métricas de dichas distancias y así mejorar el cálculo de distancias en las imágenes para así tener este aspecto en cuenta a la hora de la clasificación de las imágenes.

Como parte de la optimización de la clasificación en el artículo “Auto-Weighted Multi-View Learning for Image Clustering and Semi-Supervised Classification” [35], se plantea el modelo de aprendizaje de vista múltiple que realiza la clasificación de agrupamiento/semisupervisada y el aprendizaje de estructura local simultáneamente, logrando asignar a cada vista un peso óptimo si la necesidad explícita de parámetros y así lograr optimizar el modelo de clasificación de imágenes.

Con MVL se han abierto la posibilidad de realizar la concatenación de información que a primera vista no se relacionan, un ejemplo de ello es el trabajo “Reconstructing Perceived Images From Human Brain Activities With Bayesian Deep Multiview Learning” [36], en el cual se crea un módulo multivista enfocado a la reconstrucción de imágenes visuales, pero esto se realiza a partir de la actividad cerebral registrada en una resonancia magnética funcional que en el entrenamiento se logró relacionar estadísticamente las dos vistas (la resonancia y la imagen que es estímulo visual), de tal forma que al momento de obtener una resonancia se pueda obtener el estímulo visual que se aplicó mediante una red neuronal profunda. En la Figura 13, se expresa la idea general del desarrollo de este artículo, siendo la primera parte de entrenamiento como se relacionan las dos vistas y como en la prueba de predicción se debe comportar las actividades cerebrales para posteriormente reconstruir el estímulo esperado.



(a) Training



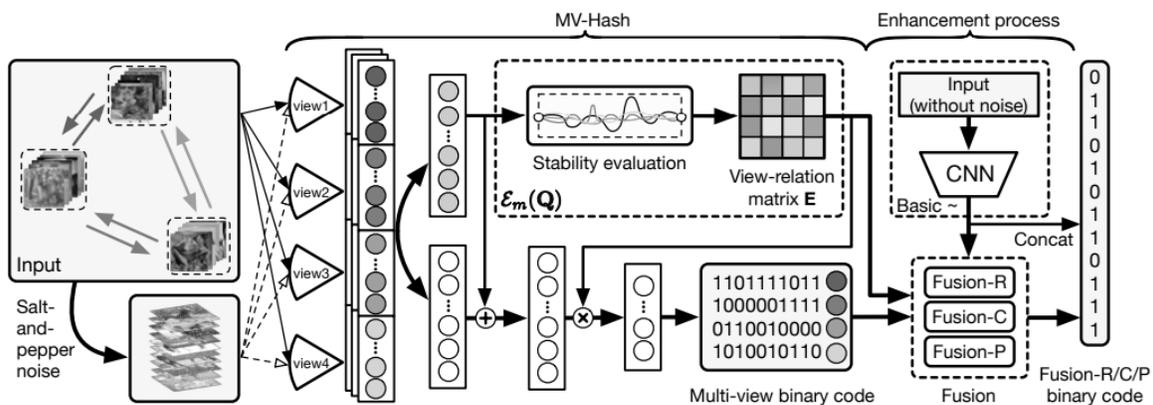
(b) Prediction

**Figura 13.** Muestra de Entrenamiento y predicción del modelo de decodificación Neuronal.

**Fuente:** [36],

Un artículo que es base para la parte de clasificación de imágenes de manera supervisada es “Experiments with Multi-view Multi-instance Learning for Supervised Image Classification” [37], ya que en este trabajo usan un clasificador por cada vista para así poder aplicar el aprendizaje de las múltiples vistas al momento de clasificar las imágenes siendo este aplicado en varios conjuntos de imágenes.

Una ventaja que se ha encontrado utilizando MVL para el manejo de imágenes es la capacidad de preservar las características de los datos al momento de ser extraídas de la fuente original, esto lo usan en el trabajo “Deep Multi-View Enhancement Hashing for Image Retrieval” [38], en donde se crea una red neuronal profunda multivista, que se enfoca en el aprendizaje del hash (del vecino más cercano en un espacio de dato) para así lograr recuperar la información de una forma eficiente y así estar evaluando activamente la relación entre las vistas ya que tiene en cuenta el ruido como parte de la entrada. En la Figura 14 se evidencia como es el proceso de tratamiento de la información en este trabajo, en donde dicho marco propuesto de hash de mejora de múltiples vistas profundas (D-MVE-Hash).

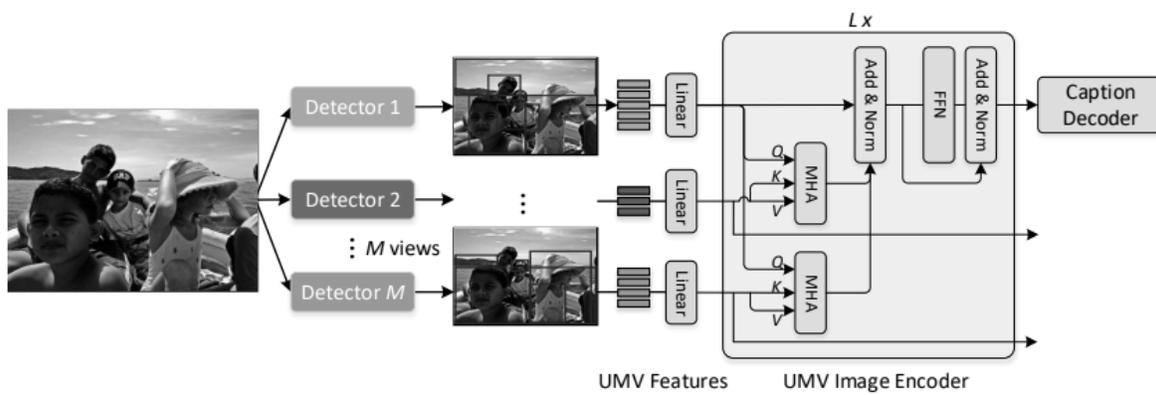


**Figura 14.** Muestra de (D-MVE-Hash) framework.

**Fuente:** [38].

Un tema recurrente en el manejo de imágenes es cuando dichas imágenes contienen alguna etiqueta, esto conlleva a que se hayan desarrollado trabajos enfocados a la comprensión de dichos etiquetas, pero estos descuidan las interacciones intra modales que pueden haber entre dichas etiquetas, es por ello que en el artículo “Multimodal Transformer with Multi-View Visual Representation for Image Captioning” [39], lo que plantean es un modelo Transformador

multimodal (MT) el cual permita la comparación de las etiquetas que han sido asignadas a las imágenes preexistentes y combina las interacciones intra modales e intermodales con un razonamiento multimodal complejo para generar etiquetas precisas y acordes al contexto que se está presentando. Tal como se muestra en la Figura 15 en donde para una imagen se usa un detector que asigna las etiquetas las cuales generan  $m$  vistas que posteriormente entraran a un codificador para así lograr asignar correctamente la etiqueta.



**Figura 15.** Muestra de la estructura del manejo de etiquetas Multimodal.

**Fuente:** [39].

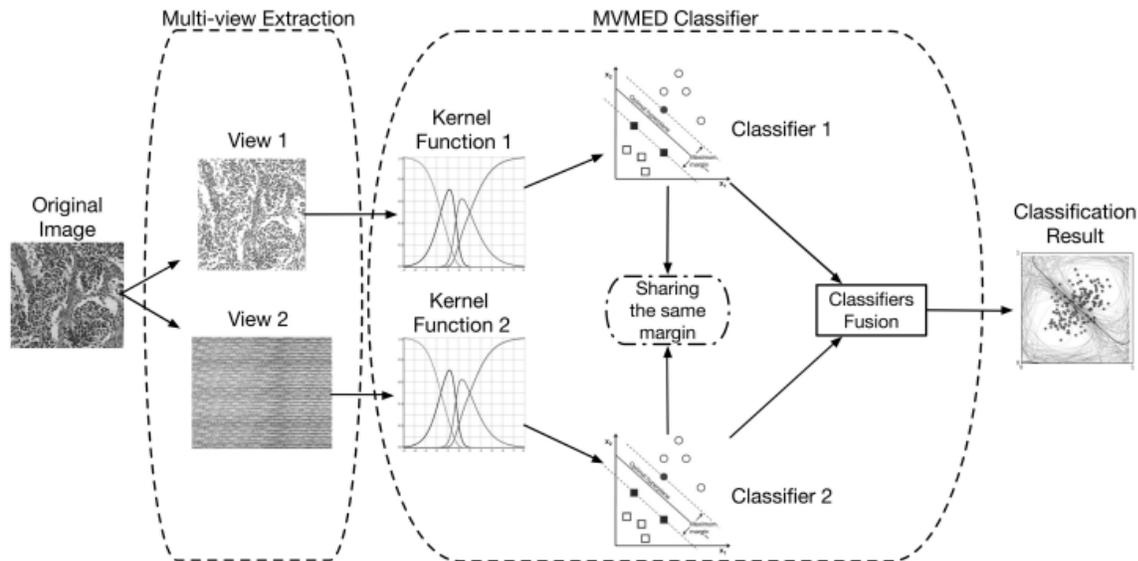
Continuando con los procesos de etiquetado en imágenes con MVL se ha planteado un modelo en el trabajo de conferencia titulado: “Joint Multi-View Representation Learning and Image Tagging” [40], en el cual se basan en el reto del manejo semántico de las etiquetas que tienen cada imagen, para ello trabajan dos subespacios en los cuales definen localmente imágenes que ya se han definido previamente y también imágenes con diferentes etiquetas basadas en el mismo ejemplo, con estas geometrías de múltiples vistas, se establecen unos puntos vecinos para modelar la

estructura semántica de las imágenes, entendiendo que una imagen con  $x$  características geométricas pueden corresponder a una etiqueta basada en la geometría que comparten entre las múltiples vistas, posteriormente se entra a un proceso de aprendizaje para entrenar el predictor que es una función de optimización que a su vez será entrenada por una máquina de soporte vectorial.

Un tema relevante para la investigación conforme al MVL enfocado a las imágenes son el manejo de las múltiples características que puede tener cada una de las múltiples vistas, este tema es trabajado en el artículo “Generative multi-view and multi-feature learning for classification” [41], en el cual al darse cuenta de esta particularidad de las múltiples vistas se propone un modelo bayesiano generativo para además de relacionar la característica que se está trabajando con la vista que se está examinando, pueda aprender una representación discriminante de las características en diversas categorías, esto a partir de un modelo condicional de clases que alterna los parámetros y variables para que el sistema vaya aprendiendo con base en resultados del mundo real.

MVL ha sido incorporado en muchas áreas, una de ellas es la medicina, tal es el caso del trabajo titulado: “Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis” [42], en el cual se parte de la idea de poder diagnosticar una patología a partir de una imagen esto para la detección de varios tipos de cáncer, enfocándose en el neuroblastoma, el tumor sólido extracraneal más común de la infancia, esta clasificación se debe realizar en unos subtipos de neuroblastoma, subtipo indiferenciado (UD), subtipo pobremente diferenciado (PD), subtipo diferenciador (D) y el tipo normal de tejidos no neoplásicos (UN), para lo cual se usan redes neuronales en el análisis de estas imágenes que contienen la patología. Comienzan haciendo una extracción de las características para generar las múltiples vistas con algoritmos de bajo nivel como lo son LBP y DSIFT, para posteriormente hacer una función de clasificación basada en una función discriminante de las múltiples vistas en la parte de MVMED Classifier donde se hace la

fusión de las clasificaciones respectivas y se da un resultado final de la clasificación tal cual como se muestra en la Figura 16.

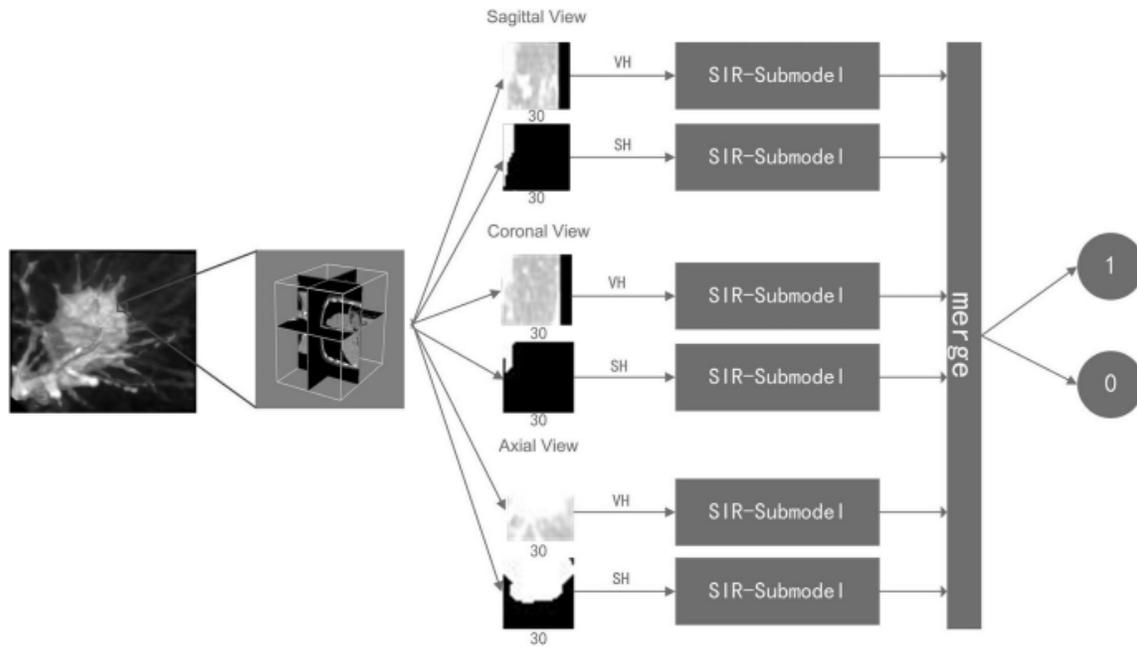


**Figura 16.** Muestra de la estructura de la arquitectura del MVMED

**Fuente:** [42].

Continuando con el análisis de imágenes enfocadas a la detección, clasificación y diagnósticos médicos aparece el artículo titulado “Multi-view secondary input collaborative deep learning for lung nodule 3D segmentation” [43], en el cual se enfoca en tomar imágenes y datos de una tomografía computarizada de tórax para detectar los cubos de nódulos pulmonares, estas imágenes en 3d de los pulmones son las que entran ser segmentados siendo este un desafío que se trabaja en dicho artículo, ya que usualmente se manejan imágenes en 2D, es por ello que se propone “multi-view secondary input residual (MV-SIR)” en donde se toma la imagen 3d y se convierte en una imagen 2d para su clasificación generando múltiples vistas de esta imagen en vistas como sagital, coronal y axial y estas se pasan en los módulos de reconocimiento de los nódulos para

posteriormente hacer una combinación de los resultados de cada vista y establecer si hay o no un nódulo pulmonar y esto se realiza en conjunto con una red neuronal para la mejora de la segmentación de las imágenes que tienen los nódulos pulmonares. Esto se puede ver representado en la Figura 17.



**Figura 17.** Muestra de la estructura del modelo MV-SIR y proceso de entrenamiento del modelo.

**Fuente** [43].

Un reto que se ha visto en la clasificación de imágenes de múltiples vistas es cuando se trata de relacionar múltiples vistas tomadas desde diferentes ángulos ya que se debe cruzar coincidencias entre las diferentes vistas, para ello se propone un modelo en el trabajo “Multi-view Deep Network for Cross-view Classification” [44], este modelo es llamado “multi-view deep network (MvDN)” en el cual se busca un discriminante no lineal e invariable de una sola vista para las múltiples vistas que se trabajan, para ello se plantearon dos subredes que intentan eliminar las variaciones de las

vistas y establecer una representación común y compartida de todas las vistas y así a partir de múltiples muestras poder guiar el aprendizaje en estas redes.

En el artículo “On Deep Multi-View Representation Learning: Objectives and Optimization” [45], los autores hacen una revisión de varias técnicas que ellos han ido trabajando como variantes de las representaciones de las múltiples vistas a las cuales les han agregado un etiquetado, entre ellas se caracterizan varias técnicas de redes neuronales entre otras técnicas y algoritmos de coherencia, aproximación y correspondencia, entre información de múltiples vistas para así obtener indicadores de evaluación y clasificación, estos posteriormente son comparados contra el modelo propuesto por los autores denominado “deep canonically correlated autoencoders (DCCA-E)” en el cual se hace el estudio de una optimización no lineal de algoritmo CCA (Canonically correlated autoencoders) para disminuir tiempos de ejecución obtener una mejora en comparación con los algoritmos trabajados en todo el artículo.

Conforme a la clasificación de imágenes, algunas de ellas se encuentran pseudoetiquetadas es por ello que en el trabajo titulado “Improving Image Classification Quality Using Multi-View Learning” [46], utilizan el principio de máxima entropía para poder asignar una etiqueta de categoría adecuada a una imagen que no tiene etiqueta usando clasificadores específicos de vista que han sido entrenados de manera óptima, haciendo que no se requiera tener una etiqueta en específico para asignar sino que de forma automática los clasificadores asignan la etiqueta que consideran acorde a la vista.

También los conjuntos de imágenes a clasificar se enfocan en una temática como lo pueden ser los rostros, lo cual es una rama de clasificación como tal ya que esta representa retos para establecer con precisión la identidad de alguien basado en el rostro, esto se trabaja en el artículo “Multi-View

Perceptron: a Deep Model for Learning Face Identity and View Representations” [47], en el cual se propone una red neurona profunda la cual tenga un perceptrón de vista múltiple en el cual se puede identificar características de identidad para cada vista y poder concatenarla en una sola vista logrando interpolar y predecir la identidad de la persona a pesar de que se trabajen diferentes puntos de vista de una persona.

Al realizar la revisión se encontró un trabajo reciente titulado “A Multi-view Convolutional Neural Network Approach for Image Data Classification” [48] , el cual centra su investigación en una red convolucional de múltiple vista enfocado a las múltiples vistas para la extracción y clasificación de las características de las imágenes “Multi-view convolutional neural network (MvCNN)” el cual busca una mejora en el rendimiento de las redes neuronales convolucionales que se han trabajado de la mano con MVL esto se comparó con “single-view learning of CNN (SvCNN)” y se establece la metodología trabajada para la mejora y comparación entre MvCNN vs SvCNN.

Como se ha presentado a lo largo de esta revisión se suelen usar etiquetas para las múltiples vistas, esto genera que al hacer los procesos se exploren subespacios en común para combinar las vistas y obtener la información que sea más útil, en el trabajo denominado “Multi-View Multi-Label Learning with View-Specific Information Extraction” [49], lo que realizan es la utilización de la información específica que se extraen de los subespacios de cada vista para posteriormente minimizar la pérdida de esta información y que las etiquetas múltiples contengan más información que pueda ser útil, esto imponiendo una serie de restricciones para así utilizar esta información discriminatoria de la vista que se está evaluando.

Una gran implementación de MVL enfocada a la clasificación de imágenes es la que se realizó en el artículo “Cross-View Representation Learning for Multi-View Logo Classification with

Information Bottleneck” [50], en la cual se enfrentan al problema de las pequeñas variaciones que puede tener una apariencia de un logotipo, y los algoritmos usuales de MVL se enfocan en hallar la máxima correlación entre las diferentes vistas, es aquí donde los autores se enfocan investigando los posibles cuellos de botella en la información a la hora de extraer las diferentes características comunes de la vista y posteriormente maximizaron la información mutua que había entre dos vistas para preservar las características clave de la clasificación, y al mismo tiempo realizan una eliminación de la información redundante que no se comparte entre las dos vistas.

Y ya para finalizar la revisión de artículos con temáticas enfocadas al MVL en imágenes y sus temáticas asociadas, se revisó el artículo “Embedding Regularizer Learning for Multi-View Semi-Supervised Classification” [51], en el cual su propósito es establecer un marco de trabajo para la integración de un regularización de información esto para poder integrar información complementaria a las vistas y de esta forma basarse en una matriz de etiquetas para la clasificación de múltiples vistas, esta técnica se denominó “Embedding regularizer learning scheme for multi-view semi-supervised classification (ERL-MVSC)”, e incluyen un aspecto de optimización conjunta de un regularización de incrustación de información y la matriz de información predicha mediante un método de manejo de coordenadas por descenso.

#### **4. Conclusiones**

La investigación estructurada que se realizó basada en los planteamientos de la metodología propuesta logró establecer las bases conceptuales más relevantes para tener en cuenta al momento de la selección de los artículos relacionado con MVL en imágenes gracias a la clasificación y depuración de los datos extraídos de cada una de las fuentes académicas consultadas.

En la revisión de los artículos de MVL se recopiló suficiente información para establecer una muestra de los temas más relevantes, cuáles son las temáticas a las que se está enfocando, cuales de ellas son las más actuales y que se están trabajando en el ámbito académico enfocado al MVL en cuestión de clasificación de imágenes.,

## **5. Agradecimientos**

En reconocimiento al grupo de investigación GIIRA, grupo al cual pertenecen los autores.

Este artículo se deriva de la investigación de la tesis doctoral MODELO COMPUTACIONAL COGNITIVO PARA LA PREDICCIÓN DE PREFERENCIAS DE USUARIO, BASADO EN ANÁLISIS DE EMOCIONES Y PRECLASIFICACIÓN DE CONTENIDO, ORIENTADO A LA VISUALIZACIÓN DE DATOS, de igual manera sirve como base para la tesis de maestría MODELO DE SOFTWARE PARA EL MANEJO DE IMÁGENES CULTURALES BASADO EN MULTIVIEW LEARNING., la cual es también derivada de la tesis doctoral anteriormente nombrada.

Para más detalles de la metodología sobre la cual se estructuró el presente artículo, revisar el artículo denominado “Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización” [52].

## **Referencias**

- [1] J. Zhao, X. Xijiong, X. Xin y S. Shiliang, “Multi-view learning overview: Recent progress and new challenges and new challenges” Information Fusion, pp. 43-54, 2017. <https://doi.org/10.1016/j.inffus.2017.02.007>
- [2] Ministerio de Ciencias de Colombia, “Modelo de clasificación de revistas científicas nacionales” 2021. [En línea]. Available: <https://minciencias.gov.co/pagina/modelo-de-clasificacion-de-revistas-cientificas-nacionales>

- [3] Scimago Lab, “Scimago Journal & Country Rank” 10 07 2021. [En línea]. Available: <https://www.scimagojr.com/>
- [4] L. Zhang, H. Shum, L. Liu, G. Guo y L. Shao, “Multiview discriminative marginal metric learning for makeup face verification”, *Neurocomputing*, pp. 339-350, 2019. <https://doi.org/10.1016/j.neucom.2018.12.003>
- [5] S. Sun, Y. Liu y L. Mao, “Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features”, *Information Fusion*, pp. 43-53, 2018. <https://doi.org/10.1016/j.inffus.2018.10.004>
- [6] C. Xu, T. Dacheng y C. Xu, “A Survey on Multi-view Learning” arXiv.org, 2013.
- [7] C. Zhang, J. Cheng y Q. Tian, “Multi-View Image Classification with Visual, Semantic and View Consistency”, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 29, pp. 617 - 627, 2019. <https://doi.org/10.1109/TIP.2019.2934576>
- [8] S. Zhu, X. Sun y D. Jin, “Multi-view semi-supervised learning for image classification”, *Neurocomputing*, pp. 136-142, 2016. <https://doi.org/10.1016/j.neucom.2016.02.072>
- [9] X. Zhang, J. Cheng, C. Xu, H. Lu y S. Ma, “Multi-view multi-label active learning for image classification”, 2009 IEEE International Conference on Multimedia and Expo, 03 07 2009.
- [10] M. Liu, Y. Luo, D. Tao, C. Xu y Y. Wen, “Low-Rank Multi-View Learning in Matrix Completion for Multi-Label Image Classification”, *Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 29, n° 1, 2015. <https://doi.org/10.1609/aaai.v29i1.9547>
- [11] H. Su, S. Maji, E. Kalogerakis y E. L. Miller, “Multi-view Convolutional Neural Networks for 3D Shape Recognition”, arXiv, 2015. <https://doi.org/10.1109/ICCV.2015.114>
- [12] M. Seeland y P. Mäder, “Multi-view classification with convolutional neural networks” *PLOS ONE*, 2020. <https://doi.org/10.1371/journal.pone.0245230>
- [13] K. Jia, J. Lin, M. Tan y D. Tao, “Deep Multi-View Learning using Neuron-Wise Correlation-Maximizing Regularizers”, 2019. <https://doi.org/10.1109/TIP.2019.2912356>
- [14] Y. Tuozhong, W. Wenfeng y G. Yuhong, “A Deep Multiview Active Learning for Large-Scale Image Classification”, *Mathematical Problems in Engineering*, 2020. <https://doi.org/10.1155/2020/6639503>
- [15] L. Sun, J. Wang, Z. Hu, Y. Xu y Z. Cui, “Multi-View Convolutional Neural Networks for Mammographic Image Classification” 2019. <https://doi.org/10.1109/ACCESS.2019.2939167>

- [16] M. Ángel Barrera Pérez, N. Y. Serrato Losada, E. Rojas Sánchez, G. Mancilla Gaona, “State of the art in software defined networking (SDN)”, *Visión electrónica*, vol. 13, no. 1, pp. 178–194, 2019. <https://doi.org/10.14483/22484728.14424>
- [17] S. Sun, J. Shawe-Taylor y L. Mao, “PAC-Bayes analysis of multi-view learning”, *Information Fusion*, pp. 117-131, 2016. <https://doi.org/10.1016/j.inffus.2016.09.008>
- [18] Z. Xu y S. Sun, “Multi-source Transfer Learning with Multi-view Adaboost”, de *International Conference on Neural Information Processing*, Doha, Qatar, 2012. [https://doi.org/10.1007/978-3-642-34487-9\\_41](https://doi.org/10.1007/978-3-642-34487-9_41)
- [19] Z. Ding, M. Shao y Y. Fu, “Robust Multi-view Representation: A Unified Perspective from Multi-view Learning to Domain Adaption”, de *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018. <https://doi.org/10.24963/ijcai.2018/767>
- [20] F. Wu, X. Y. Jing, X. You, D. Yue, R. Hu y J. Y. Yang, “Multi-view low-rank dictionary learning for image classification,” *Pattern Recognition*, pp. 143-154, 2015. <https://doi.org/10.1016/j.patcog.2015.08.012>
- [21] Y.-M. Xu, C.-D. Wang y J.-H. Lai, “Weighted Multi-view Clustering with Feature Selection”, *Pattern Recognition*, pp. 25-35, 2015. <https://doi.org/10.1016/j.patcog.2015.12.007>
- [22] D. Yu, S. Ji, J. Liu y S. Wei, “Automatic 3D building reconstruction from multi-view aerial images with deep learning”, *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 155-170, 2021. <https://doi.org/10.1016/j.isprsjprs.2020.11.011>
- [23] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu y Y. Xu, “Forest Fire Recognition Based on Feature Extraction from Multi-View Images”, *Traitement du Signal*, pp. 775-783, 2021. <https://doi.org/10.18280/ts.380324>
- [24] N. Khan, M. Kim y J. Tompkin, “Differentiable Diffusion for Dense Depth Estimation from Multi-view Images”, arXiv, 2021. <https://doi.org/10.1109/CVPR46437.2021.00880>
- [25] X. Yan, S. Hu, Y. Mao, Y. Ye y H. Yu, “Deep multi-view learning methods: A review”, *Neurocomputing*, pp. 106-129, 2021. <https://doi.org/10.1016/j.neucom.2021.03.090>
- [26] J. N. Oliver, “Multi-View Imaging: Depth Estimation and Enhancement”, 2019. [En línea]. Available: <https://dspace.uib.es/xmlui/handle/11201/150580>
- [27] J. García, A. Gardel, I. Bravo y J. L. Lázaro, “Multiple View Oriented Matching Algorithm for People Reidentification”, *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, pp. 1841 - 1851, 2014. <https://doi.org/10.1109/TII.2014.2330976>

- [28] B. Liu, A. Yu, X. Yu, R. Wang, K. Gao y W. Guo, "Deep Multiview Learning for Hyperspectral Image Classification", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, pp. 7758 - 7772, 2021. <https://doi.org/10.1109/TGRS.2020.3034133>
- [29] Y. Zhao, Y.-m. Cheung, X. You, Q. Peng, J. Peng, P. Yuan y Y. Shi, "Hyperspectral Image Classification via Spatial Window-Based Multiview Intact Feature Learning", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, pp. 2294 - 2306, 03 03 2021. <https://doi.org/10.1109/TGRS.2020.3004858>
- [30] K. Wong, E. Sinkovskaya, A. Abuhamad y T. Mahmood, "Multiview and Multiclass Image Segmentation using Deep Learning in Fetal Echocardiography" Proceedings of SPIE - The International Society for Optical Engineering, 2021. <https://doi.org/10.1117/12.2582191>
- [31] A. Li, J. Chen, D. Chen y G. Sun, "Multiview Similarity Learning for Robust Visual Clustering" de Computer Vision - ACCV 2020 Workshops: 15th Asian Conference on Computer Vision, Kyoto, Japan, Tokio, 2020. [https://doi.org/10.1007/978-3-030-69756-3\\_12](https://doi.org/10.1007/978-3-030-69756-3_12)
- [32] F. Nie, G. Cai y X. Li, "Multi-View Clustering and Semi-Supervised Classification with Adaptive Neighbours" IEEE Transactions on Knowledge and Data Engineering, pp. 2389 - 2400, 2020. <https://doi.org/10.1109/TKDE.2019.2920985>
- [33] W. Liu y D. Tao, "Multiview Hessian Regularization for Image Annotation" IEEE Transactions on Image Processing, 2019.
- [34] D. Li y Y. Tian, "Multi-view metric learning for multi-instance image classification" Expert Systems with Applications, 2021.
- [35] F. Nie, G. Cai, J. Li y X. Li, "Auto-Weighted Multi-View Learning for Image Clustering and Semi-Supervised Classification" IEEE Transactions on Image Processing, pp. 1501 - 1511, 2018. <https://doi.org/10.1109/TIP.2017.2754939>
- [36] C. Du, C. Du, L. Huang y H. He, "Reconstructing Perceived Images from Human Brain Activities with Bayesian Deep Multiview Learning" IEEE Transactions on Neural Networks and Learning Systems, pp. 1-14, 2018.
- [37] M. Mayo y E. Frank, "Experiments with Multi-view Multi-instance Learning for Supervised Image Classification", 2011.
- [38] C. Yan, B. Gong, Y. Wei y Y. Gao, "Deep Multi-View Enhancement Hashing for Image Retrieval", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

- [39] J. Yu, J. Li, Z. Yu y Q. Huang, “Multimodal Transformer with Multi-View Visual Representation for Image Captioning”, IEEE Transactions on Circuits and Systems for Video Technology, p. 4467-4480, 2020. <https://doi.org/10.1109/TCSVT.2019.2947482>
- [40] Z. Xue, G. Li y Q. Huang, “Joint Multi-View Representation Learning and Image Tagging”, de Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016. <https://doi.org/10.1609/aaai.v30i1.10147>
- [41] J. Li, B. Zhang, G. Lu y D. Zhang, “Generative multi-view and multi-feature learning for classification”, Information Fusion, pp. 215-226, 2019. <https://doi.org/10.1016/j.inffus.2018.02.005>
- [42] Y. Liu, S. Sun y M. Yin, “Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis”, de Pacific Rim International Conference on Artificial Intelligence, 2018. [https://doi.org/10.1007/978-3-319-97304-3\\_42](https://doi.org/10.1007/978-3-319-97304-3_42)
- [43] X. Dong, S. Xu, Y. Liu, A. Wang, M. I. Saripan, L. Li, X. Zhang y L. Lu, “Multi-view secondary input collaborative deep learning for lung nodule 3D segmentation”, Cancer imaging, 2020. <https://doi.org/10.1186/s40644-020-00331-0>
- [44] M. Kan, S. Shan y X. Chen, “Multi-view Deep Network for Cross-view Classification”, de 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. <https://doi.org/10.1109/CVPR.2016.524>
- [45] W. Wang, R. Arora, K. Livescu y J. Bilmes, “On Deep Multi-View Representation Learning: Objectives and Optimization”, de ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015.
- [46] B. Neva Camacho, O. Najar Sánchez, "Comprensión de la lectura crítica con el edublog Leanet", Revista Vínculos, vol. 16, no. 2, pp. 219–231, 2019. <https://doi.org/10.14483/2322939X.15750>
- [47] Z. Zhu, P. Luo, X. Wang y X. Tang, “Multi-View Perceptron: A Deep Model for Learning Face Identity and View Representations”, de NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2015.
- [48] M. T. Alam, V. Kumar y A. Kumar, “A Multi-view Convolutional Neural Network Approach for Image Data Classification”, de 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021. <https://doi.org/10.1109/ICCICT50803.2021.9509943>

[49] X. Wu, Q.-G. Chen, Y. Hu, D. Wang y X. Chang, “Multi-View Multi-Label Learning with View-Specific Information Extraction”, de IJCAI'19: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019. <https://doi.org/10.24963/ijcai.2019/539>

[50] J. Wang, Y. Zheng, J. Song y S. Hou, “Cross-View Representation Learning for Multi-View Logo Classification with Information Bottleneck”, 2021. <https://doi.org/10.1145/3474085.3475704>

[51] A. Huang, Z. Wang, Y. Zheng, T. Zhao y C.-W. Lin, “Embedding Regularizer Learning for Multi-View Semi-Supervised Classification”, pp. 6997-7011, 2021. <https://doi.org/10.1109/TIP.2021.3101917>

[52] E. Gómez Luna, D. Fernando Navas, G. Aponte Mayor y L. A. Betancourt Buitrago, “Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización”, Dyna, vol. 81, n° 184, pp. 158-163, 2014. <https://doi.org/10.15446/dyna.v81n184.37066>