



DOI: <http://dx.doi.org/10.23857/dc.v8i3>

Ciencias Técnicas y Aplicadas  
Artículo de Investigación

*Detección y digitalización de datos de interés en documentos de identificación*

*Detection and digitization of data of interest in identification documents*

*Detecção e digitalização de dados de interesse em documentos de identificação*

Erick Alejandro Ayala-Churo <sup>I</sup>  
[eaayalac07@est.ucacue.edu.ec](mailto:eaayalac07@est.ucacue.edu.ec)  
<https://orcid.org/0000-0001-9633-5742>

Juan Pablo Cuenca <sup>II</sup>  
[jcuenca@ucacue.edu.ec](mailto:jcuenca@ucacue.edu.ec)  
<https://orcid.org/0000-0001-5982-634X>

Andrés Sebastián Quevedo-Sacoto <sup>III</sup>  
[asquevedos@ucacue.edu.ec](mailto:asquevedos@ucacue.edu.ec)  
<https://orcid.org/0000-0001-5585-0270>

**Correspondencia:** [eaayalac07@est.ucacue.edu.ec](mailto:eaayalac07@est.ucacue.edu.ec)

\***Recibido:** 29 de julio del 2022 \***Aceptado:** 12 de agosto de 2022 \* **Publicado:** 15 de septiembre de 2022

- I. Estudiante de la Universidad Católica de Cuenca, Cuenca, Ecuador.
- II. Docente de la Carrera de Software, Universidad Católica de Cuenca, Ecuador.
- III. Magíster en Geomática, Docente de la Universidad Católica de Cuenca, Cuenca, Ecuador.

## Resumen

La automatización de la detección y contextualización de campos en documentos de identificación como requerimiento principal de las instituciones públicas y privadas ha motivado la implementación de arquitecturas de redes neuronales convolucionales con el fin de mejorar aspectos como el tiempo, la calidad y la cantidad de documentos digitales analizados. En el presente trabajo, se propone un modelo de detección que permita la extracción y contextualización de campos dentro de documentos de identificación. Con el fin de ejecutar el entrenamiento de los modelos de detección, fue necesaria la construcción de un dataset que contenga imágenes reales de documentos previamente etiquetado por personas expertas en el área. Se procedió a realizar el etiquetado de imágenes mediante plataformas, lo que permitió la identificación y contextualización de campos mediante cuadros delimitadores. Durante el entrenamiento del modelo se obtuvo un resultado de la Precisión Media de Entrenamiento (mAP) que alcanzó el 100%.

**Palabras clave:** Aprendizaje Profundo; detección de campos; creación del dataset; reconocimiento óptico de caracteres; red neuronal convolucional.

## Abstract

The automation of the detection and contextualization of fields in identification documents as a main requirement of public and private institutions has motivated the implementation of convolutional neural network architectures in order to improve aspects such as time, quality and quantity of digital documents. analyzed. In the present work, a detection model is proposed that allows the extraction and contextualization of fields within identification documents. In order to execute the training of the detection models, it was necessary to build a dataset containing real images of documents previously labeled by experts in the area. The labeling of images was carried out through platforms, which allowed the identification and contextualization of fields through bounding boxes. During the training of the model, a result of the Average Training Accuracy (mAP) that reached 100% was obtained.

**Keywords:** Deep Learning; field detection; dataset creation; optical character recognition; convolutional neural network.

## Resumo

A automatização da detecção e contextualização de campos em documentos de identificação como principal requisito de instituições públicas e privadas tem motivado a implementação de arquiteturas de redes neurais convolucionais a fim de melhorar aspectos como tempo, qualidade e quantidade dos documentos digitais analisados. No presente trabalho, é proposto um modelo de detecção que permite a extração e contextualização de campos dentro de documentos de identificação. Para executar o treinamento dos modelos de detecção, foi necessário construir um conjunto de dados contendo imagens reais de documentos previamente rotulados por especialistas na área. A rotulagem das imagens foi realizada por meio de plataformas, que permitiram a identificação e contextualização dos campos por meio de caixas delimitadoras. Durante o treinamento do modelo, obteve-se um resultado de Precisão Média de Treinamento (mAP) que atingiu 100%.

**Palavras-chave:** Aprendizagem Profunda; detecção de campo; criação do conjunto de dados; reconhecimento óptico de caracteres; rede neural convolucional.

## Introducción

Desde el inicio de 2020 se ha incrementado la necesidad de obtener documentos digitalizados a consecuencia de la pandemia debida al COVID-19 que ha obligado a la suspensión de actividades presenciales. Por otro lado, la precaria infraestructura digital disponible, los escasos recursos y el desconocimiento de las tecnologías emergentes han sido algunas de las carencias existentes en ciertas organizaciones públicas y privadas (Molina-Pérez J., 2021). En este contexto, surge la necesidad de digitalizar documentos y de extraer de los mismos la información necesaria, la cual luego de ser validada sea empleada para los diferentes procesos administrativos.

El uso extendido de documentos digitales como parte fundamental de los datos empresariales, la extracción de la información referente a identificación personal y la asignación de una función para esta información dentro de dichos documentos se han convertido en requerimientos funcionales muchas veces obligatorios que permiten cumplir requerimientos tanto legales como de veracidad de la información.

En este contexto nos podemos referir como ejemplo los requisitos obligatorios que tienen que cumplir las instituciones públicas y privadas de educación superior del Ecuador cuando se encuentran en procesos de acreditación, estas deben entregar documentos de identificación

## Detección y digitalización de datos de interés en documentos de identificación

---

digitalizados de sus estudiantes, adicionalmente estos datos deben ser procesados con el fin de verificar su veracidad, significado y el aporte que proporciona para ejecutar el proceso. Para cumplir con este requerimiento las instituciones crean formularios donde se les solicita a los estudiantes ingresar la información necesaria. Sin embargo, esta medida no asegura la automatización de la verificación y la digitalización de la información enviada puesto que los estudiantes realizan esta actividad de manera autónoma y en muchos casos no se garantiza la fiabilidad del contenido de los datos enviados por los estudiantes.

La automatización del proceso de validación de los datos no es una tarea sencilla, puesto que uno de los problemas principales es la basta cantidad de documentos que ingresan las instituciones durante procesos de acreditación. Por otro lado, este proceso no solo consiste en la digitalización de la información, adicionalmente es necesario contextualizar y comprender el área de la imagen que ha podido ser identificada.

Con el fin de automatizar los procesos de identificación, verificación y digitalización es necesario implementar estrategias que permitan garantizar la calidad de los datos sin la revisión y supervisión de un ser humano. En este artículo presentamos la implementación y desarrollo de un modelo de aprendizaje profundo puesto en producción en una arquitectura de microservicios que permite detectar y extraer los puntos de interés de los documentos de identificación de una manera autónoma. El modelo fue entrenado en un conjunto de datos de 600 imágenes de documentos de identidad ecuatorianos previamente etiquetados con los puntos de interés que permiten identificar elementos semánticos. Adicionalmente con tecnologías de reconocimiento óptico de caracteres se realizó la transformación de las imágenes a caracteres alfanuméricos. Así garantizando la digitalización y fiabilidad de los datos.

Este artículo se encuentra estructurado de la siguiente manera: en la sección 2, se refiere a los trabajos previamente realizados sobre Modelos de aprendizaje profundo y redes neuronales convolucionales; sección 3, presentamos la metodología y herramientas utilizadas; sección 4, se realiza la discusión y presentación de los resultados del desarrollo; sección 5, nuestras conclusiones.

### Estado del arte

La detección e identificación de la semántica de la información, es una tarea de entrenamiento y análisis de dimensiones de archivos. Tiene como meta principal reducir el tiempo de reconocimiento y mejorar los procesos de digitalización. Para el cumplimiento de estos procesos es

## Detección y digitalización de datos de interés en documentos de identificación

---

necesario el análisis de las diferentes tecnologías de reconocimiento de textos en documentos con el fin de identificar y verificar si estos datos proporcionan un valor para el cumplimiento de requerimientos durante los procesos que realicen las instituciones (B Parvathi Sangeetha, 2022).

La pandemia de COVID-19 ha proporcionado una presión importante en sectores gerenciales con respecto a la recepción de documentos en físico para iniciar con procesos administrativos. A consecuencia, se ha adoptado a la digitalización como una medida necesaria para el control de datos y la eficacia operativa. Por lo cual (Karthikeyan, Seco de Herrera, Doctor, & Mirza, 2021) propone en su estudio una técnica de preentrenamiento orientado a la autosupervisión en redes neuronales, al cual denominó Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa) la cual fue entrenada mediante un conjunto de datos específicos extraídos desde documentos médicos los cuales muestran una tasa de error reducida con el fin de mejorar la precisión de la detección durante el proceso de entrenamiento. En este contexto, el algoritmo permitiera ubicar la información y darle un contexto dentro de los documentos que pueda generarse posteriormente.

El estudio anterior evidencia que los modelos de aprendizaje y automatización requieren recursos para el aprendizaje con una curva de precisión elevada, según (Rivas, Paul, Hristidis, Papalexakis, & Roy-Chowdhury, 2022) una de las plataformas favoritas para la recolección de datos destinados a estudios como el análisis de sentimientos, etiquetado de imágenes y predicción de ubicación de objetos es Twitter, el cual alberga una vasta cantidad de datos en distintos tipos como lo son, sólo texto, o texto e imagen.

Estudios realizados para el entrenamiento de algoritmos basados en Deep Learning se ha detectado que uno de los problemas principales para el aprendizaje consiste en la detección de texto irregular en imágenes naturales, (Prabu & Sundar K., 2023) propone una solución basa en un marco de reconocimiento que consiste en una arquitectura de codificación/codificación. Esta arquitectura se encuentra dividida en cuatro etapas, la primera consistente una transformación de texto irregular a texto legible por medio de TPS (Thin Plate Spline), SAM (Módulo de Atención Espacial) para obligar al modelo a concentrarse en las áreas con texto y extraer un mapeado característico, Bi-LSTM (Red Profunda de memoria a corto plazo bidimensional.) para el mapeado de caracteres textuales partiendo de una red neuronal convolucional común y por último DSAM (Mecanismo de atención de dos pasos) en conjunto con CTC (Clasificación temporal Conexionista) que tienen

## DetECCIÓN Y DIGITALIZACIÓN DE DATOS DE INTERÉS EN DOCUMENTOS DE IDENTIFICACIÓN

---

como propósito integrar las características visuales de la imagen y el texto para generar secuencias de caracteres con mejor precisión.

(Khan, Uddin, Parvez, & Nahar, 2022) manifiesta que ciertos documentos digitalizados poseen partes las cuales necesariamente deben estar hechos a mano como lo son las firmas de los participantes del proceso. Esto hace de la automatización de la lectura de estos documentos una tarea complicada por la forma, el tamaño y la variación de colores. Sin embargo, ha propuesto un modelo de red neuronal convolucional CNN, por las iniciales en inglés convolutional neural network, en conjunto con SE-ResNet y los denominados bloques de comprensión y excitación que permite el reconocimiento de caracteres escritos a mano.

La identificación y extracción de texto es una tarea requerida incluso en documentos digitalizados no planos. Según (Colter, y otros, 2022) la necesidad por la extracción de texto en tablas ubicadas en documentos digitalizados con fines para la minería de datos. No obstante, las diferentes herramientas de identificación y extracción de texto para tablas guardadas como imágenes son ineficientes, por lo cual presentan Tablext la cual es una herramienta que combina técnicas de visión por computador y métodos de aprendizaje profundo por medio de una red neuronal convolucional usando la personalización proporcionada por el modelo de detección de objetos YOLO y posteriormente aplicar la tecnología OCR para la lectura de la información detectada en las tablas gracias al red neuronal convolucional.

El siguiente estudio (Arslan, 2022) centra sus esfuerzos en el desarrollo de un modelo basado en CNN con el fin de automatizar la identificación de datos usables en documentos digitales e imágenes orientados a la facturación. Propone un modelo de detección de textos de interés en facturas digitalizadas por medio del procesamiento de imágenes basado en Yolov5, obteniendo los textos de las celdas y las imágenes legibles de cada tabla para después ser procesadas por medio la tecnología OCR.

(Almutairi & Almashan, 2019) propone una solución basada en OCR orientada a la extracción semántica de los principales elementos ubicados en páginas de periódicos tales como artículos, anuncios y cabeceras. Todo este proceso está impulsado mediante el uso de Mask R-CNN para poder crear el modelo sin limitación del lenguaje ubicado en el documento.

Con el fin de identificar zonas de interés (MRZ) para la verificación de datos e identidad en pasaportes y visados, (Liu, James, Gupta, & Raviv, 2022) en su estudio establece que la extracción del MRZ no es una tarea que asegure una precisión aceptable. Por lo tanto, presenta un modelo

## Detección y digitalización de datos de interés en documentos de identificación

---

específicamente diseñado con redes neuronales convulsionales, que permite una identificación y lectura de pasaportes con una tasa de detección del 100% y un reconocimiento del 99,25% demostrado en una gran variedad de simulaciones. En base al anterior estudio se determinó que es una solución que da buenos resultados al realizar el entrenamiento de redes totalmente convulsionales FCN. Sin embargo, según (Sun & Wang, 2018) este metodo tiene una desventaja notable la cual es la perdida de información específica dada por las operaciones de downsampling en FCN, con el fin de dar solución a este problema recurrente han optado por ejecutar una combinación máxima de la información semántica de las capas profundas y las superficiales. Además, proponen el desarrollo de un robusto Backend con el fin de mejorar los resultados de FCN alcanzando una precisión general del 90,6%. Por otro lado, en Liao et al. (Liao, Wan, Yao, Chen, & Bai, 2020) desarrolló un componente para realizar procesos de bancarización en una red subdivida. Este componente permite la integración de módulos de tal forma que tenga una orientación particular y comunicación eficiente sin la perdida de datos durante su traslado.

Se han estudiado tecnologías basadas en OCR Detección Óptica de Texto (Smith, 2017) que han sido usadas como prototipo para investigaciones de precisión para la digitalización de documentos, entre las tecnologías estudiadas se encuentra Tesseract el cual fue objeto principal de estudio para el reconocimiento de documentos digitalizados de Indonesia tal como el e-KTP (Kartu Tanda Penduduk Elektronik) documento nacional de identidad por medio de una red generativa adversa (Zhang & Putra, 2022) que permite identificar y dar un significado de valor a esta información. Con el fin de mejorar la detección óptica de texto en imágenes, (Sherrah, 2016) establece que, en lugar de centrarse en el contenido espectral de la imagen, es necesario tomar en cuenta la apariencia general de la imagen. Para ello, la adición de redes neuronales profundas (CNN) al etiquetado semántico de datos proporcionado por imágenes de alta resolución proporcionando una precisión de vanguardia para detección de los datos.

Con respecto a la modificación de imágenes durante la detección del texto, (Long, y otros, 2018) e (He, Gkioxari, Dollar, & Girshick, 2017) implementaron una arquitectura orientada a Mask CNN, en donde se procede a realizar un enmascaramiento de imágenes por medio de divisiones de pixeles mejorando la arquitectura en un tiempo considerable. Con respecto a la tecnología de detección óptica Tesseract, se han realizados pruebas con todos regulares de color de imagen, que permite una detección individual de los caracteres mejorando la identificación de letras/palabras. Una de las

## Detección y digitalización de datos de interés en documentos de identificación

---

pruebas iniciales realizadas por un estudio general del algoritmo (Smith, 2017), separa las palabras y las refragmenta obteniendo un todo fijo en la figura 1.



**Figura 1:** Palabra Fragmentada en tono Fijo (Smith, 2017).

Según (Lin, Tu, & Lin, 2021) propone un esquema para la detección de palabras de interés entrenadas por medio de redes neuronales convolucionales profundas, este algoritmo permite detectar información determinada en imágenes de documentos. Este esquema está estructurado por medio de componentes de detección de datos claves y análisis de la sintaxis de las palabras. En resumen, gracias a estos dos componentes permite separar el proceso de extracción y detección en dos procesos donde se determina la ubicación y espacio del dato en el documento y posteriormente se determina su significado y posición en el proceso.

Durante un estudio realizado por (Ramdhani, Budi, & Purwandari, 2021) se determinó y planteo las ventajas y desventajas de diferentes motores OCR, entre los motores de reconocimientos estudiados están Foxit, PDF2GO y Tesseract. Como punto clave se determina que Tesseract tiene una mejor extracción de reconocimiento de caracteres específicos y mejores tiempos para la conversión en conjunto con una precisión proporcional al mismo.

La propuesta de desarrollo dada por (Salma, y otros, 2021) implementó el motor de detección Tesseract en conjunto un modelo de detección de datos entrenado por YoloV5 con el fin de extraer información proporcionada por imágenes de matrículas de vehículos en Paquistán. Esta investigación proporciona un análisis comparativo para determinar la precisión de trabajar con Tesseract en conjunto con redes neuronales convolucionales con el fin de obtener mejores resultados en tiempo y contextualización de los datos.

Según (Clausner, Antonacopoulos, & Pletschacher, 2020) propone una metodología para el eficaz y eficiente entrenamiento de los motores OCR, mediante un análisis de documentos Aletheia, procesos usados para el entrenamiento, reconocimiento de texto manual y evaluación cuantitativa

## Detección y digitalización de datos de interés en documentos de identificación

---

de cada motor estudiado. El estudio permitió generar un sistema de evaluación completo con el fin de justificar el enfoque de la investigación proporcionando resultados que permiten validar el enfoque del entrenamiento propuesto.

Para la extracción de información relevante en fondos complejos y difíciles de detectar caracteres, (Akinbade, Ogunde, Odim, & Oguntunde, 2020) realizaron un estudio en donde se recopila datos de imágenes e implementa distintas técnicas usadas para la contextualización de la información. Para ello se empleó un algoritmo de umbralización adaptable para ciertas imágenes con el fin de delimitar y confinar la información de textual oculto en las imágenes con un entorno complejo.

Por otra parte, (Chernyshova, Sheshkus, & Arlazorov, 2020) muestra una metodología para el reconocimiento de texto mediante dispositivos móviles o integrados. La solución propuesta consiste en dos redes neuronales convolucionales separadas por medio de programación dinámica que permiten el entrenamiento y procesamiento de imágenes por medio de segmentación de parámetros entrenables individuales lo cual permite detectar texto en documentos con fondos complejos, imágenes de baja calidad y diferentes tipos de letras e idiomas.

Es importante conocer las metodologías implementadas para el correcto entrenamiento de datos en conjunto con motores de reconocimiento de caracteres, para ello (Dergachov, Krasnov, Bilozerskyi, & Zymovin, 2021) presentan un estudio en donde se presenta un modelo moderno para la mejora de la calidad de sistemas de reconocimiento óptico. Para ello, se sintetizaron algoritmo de procesamiento de imágenes a criterio del usuario para poder compensar ciertos factores negativos como la forma de los objetos y la distorsión del texto, mala iluminación, efectos de ruido, etc.

En conclusión, en el Ecuador no se han implementado tecnologías con el propósito de automatizar los procesos de identificación, validación y digitalización con el fin de detectar la información usable dentro de documentos de identificación y así poder darles un valor y contexto durante los diferentes procesos que solicitan las instituciones públicas y privadas.

### Metodología

La problemática principal consiste en la detección, identificación, validación y extracción de datos de interés que se encuentran en documentos de identificación ecuatorianos, para ello entrenamos un modelo de detección de Objetos basado en enfoques de aprendizaje profundo con el fin de identificar estos datos para posteriormente extraerlos y procesarlos. En este contexto en las siguientes subsecciones se detallan los siguientes aspectos:

## Detección y digitalización de datos de interés en documentos de identificación

---

- Las herramientas utilizadas para implementar el desarrollo de la propuesta;
- El uso de Yolov5 para el desarrollo del entrenamiento;
- Las características del modelo usado;
- El proceso implementado para la creación y preparación del dataset usado en el entrenamiento;
- El proceso realizado para la construcción del modelo de detección; y,
- La arquitectura del sistema.

### Herramientas

#### Hardware

El equipo de entrenamiento usado para el desarrollo del algoritmo consiste en una estación de trabajo Lambda con las siguientes características: 132Gb de RAM, con una tarjeta gráfica NVIDIA RTX-3070 de 8Gb y otra de tarjeta RTX-3090 de 24 Gb para las GPU. El entrenamiento fue ejecutado en el lenguaje Python 3.8 y el framework de aprendizaje de máquina Pytorch.

#### Plataforma para la gestión del desarrollo

La administración de proyecto de software se gestionó por medio de Jira Software *Plataforma para la Gestión de Proyectos Ágiles*. Esta plataforma permite la gestión de proyectos por medio de la metodología Ágil SCRUM. El proyecto se encuentra separado por cuatro etapas/sprint que se detallan en la siguiente sección.

#### Framework usado para la construcción de la Interfaz del usuario

Se eligió Angular como framework de desarrollo orientado a la web puesto que provee de herramientas que permite la comunicación con documentos HTML estáticos orientado al contenido dinámico.

## Gestor de versionamiento

Mediante un gestor de versiones Git se realizó la presentación de los sprints y avances de las tareas con los colaboradores durante el desarrollo. GitHub es una plataforma que permite gestionar cambios y controlar errores, mediante un repositorio Online.

## Justificación del Modelo

La detección de áreas específicas en una imagen puede variar dependiendo de distintas características como anomalías, cambios bruscos del contraste, brillo, opacidad o factores externos que hayan alterado la imagen. Existen ciertos modelos que usan el concepto de IoU “Intersection over Union” lo cual permite obtener un acierto del área de predicción del bounding-box real, requeridos para la detección. Han surgido varios CNN usados en el entrenamiento para la detección de objetos en imágenes, como es el caso de Fast R-CNN, Faster R-CNN (Ren, He, Girshick, & Sun, 2015), y Mask R-CNN (Almutairi & Almashan, 2019).

En 2016 se creó una propuesta conocida como YOLO (Arslan, 2022) (Colter, y otros, 2022) “You Only Look Once”, que a diferencia de sus predecesores realiza un solo procesamiento de la detección para la cual fue entrenada, es decir, no realiza iteraciones. Esta propiedad permite obtener una mejor velocidad de resultados en las ordenes que se le ha solicitado y, agregar recursos externos para aplicar el proceso de entrenado como un video en tiempo real y la ejecución en dispositivos móviles.

## Modelo

Para determinar la metodología que mejor se ajusta a la propuesta fue necesario entrenar dos CNN profundas para la detección de puntos de interés en imágenes en los documentos de identificación: a) Yolov5 (Colter, y otros, 2022), una red neuronal convolucional para la detección de objetos, en conjunto con Res-Net 50 como pilar principal; y, b) Faster R-CNN (Ren, He, Girshick, & Sun, 2015). Estos modelos CNN se entrenaron mediante un dataset compuesto por imágenes que contienen de una a dos caras del documento de identificación, generando porcentajes con respecto a la precisión de la detección del campo durante el entrenamiento. Por otra parte, durante el entrenamiento surgieron particularidades que obligaron a ejecutar este proceso varias veces, generando un total de diez experimentos con diferentes cantidades de imágenes procesadas y su

## DetECCIÓN Y DIGITALIZACIÓN DE DATOS DE INTERÉS EN DOCUMENTOS DE IDENTIFICACIÓN

---

correspondiente información etiquetada, para lo cual se aplicaron distintas técnicas de aumento de datos y procesamiento de limpieza de los mismos.

### **Dataset**

Para el entrenamiento del algoritmo, se realiza la construcción de un conjunto de datos proporcionados por estudiantes, que se conforman 600 cédulas de identidad, para posteriormente ser etiquetadas por profesionales. En este Dataset se pueden identificar un total de 11 puntos de interés, que representan datos necesarios para el cumplimiento de procesos en instituciones públicas y privadas. En su totalidad, la información real recopilada fue proporcionada por instituciones de Cuenca. Dando un total de 600 imágenes en las cuales para poder identificar y contextualizar los diferentes puntos de interés por medio de rectángulos, detallando la semántica de los datos capturados en las imágenes en formatos JPG y PNG.

Los datos etiquetados constan de 11 campos identificados y contextualizados descritos como: (0) ID, (1) Names, (2) Birth place, (3) Birth date, (4) Nationality, (5) Sex, (6) Civil status, (7) Picture, (8) Expiration date, (9) Fingerprint, (10) Signature.

### **Construcción del Modelo de detección**

Durante la primera etapa (Primer Sprint: Análisis y desarrollo del algoritmo.) del desarrollo se estableció el entrenamiento del algoritmo, en el cual se encuentra dividido en cuatro actividades principales. Estas actividades se encuentran distribuidas con el fin de obtener resultados más precisos para que el algoritmo de identificación sea robusto.

### **Recopilación de datos**

Durante esta etapa se procedió a realizar la solicitud para el uso de información de documentos de identificación. Para lo cual se realizó la recolección de 600 firmas dadas por estudiantes con el fin de que podamos tener acceso a los datos para poder comenzar el entrenamiento del algoritmo. Una vez obtenidos los datos se procedió a hacer la limpieza de las imágenes que son útiles para el etiquetado.

## **Etiquetado**

La anotación de los datos o etiquetado es el paso más importante antes de la creación del modelo de detección de datos. En esta etapa es donde principalmente intervinieron expertos humanos para el etiquetado de los datos completos y personalizados mediante el uso de la plataforma MakeSense. Esta herramienta nos permite realizar el etiquetado personalizado y nos devuelve información correspondiente a la posición del dato que hemos etiquetado y de paso podemos exportar esta información para poder proceder a hacer el entrenamiento.

## **Entrenamiento del Modelo de Aprendizaje Profundo**

Entre los diferentes modelos de aprendizaje profundo que se probaron para el entrenamiento del algoritmo se decidió utilizar YOLOv5, puesto que permite la el ingreso de Dataset y entrenamiento personalizado. Durante esta etapa se usa las imágenes etiquetadas en conjunto con los labels generados por la plataforma de etiquetado. Por medio de línea de comando se procede a la ejecución del aprendizaje del algoritmo.

## **Resultados**

Como resultado del entrenamiento, el modelo de detección YOLOv5 nos arrojó ciertos resultados con respecto al aprendizaje del algoritmo. Con respecto a la precisión durante el aprendizaje tenemos 93.7%, recall del 99.2% y con respecto al mAP obtenemos 1.00%.

## **Arquitectura**

En esta sección apartado se detallará el contenido de la estructura, módulos y funcionalidades de la arquitectura del proyecto desarrollado mostrado en la figura 2. En la sección (a) detallamos los componentes que se usaron para el desarrollo del Backend; por último, en la sección (b) se habla sobre los componentes del modelo de entrenamiento.

## Detección y digitalización de datos de interés en documentos de identificación

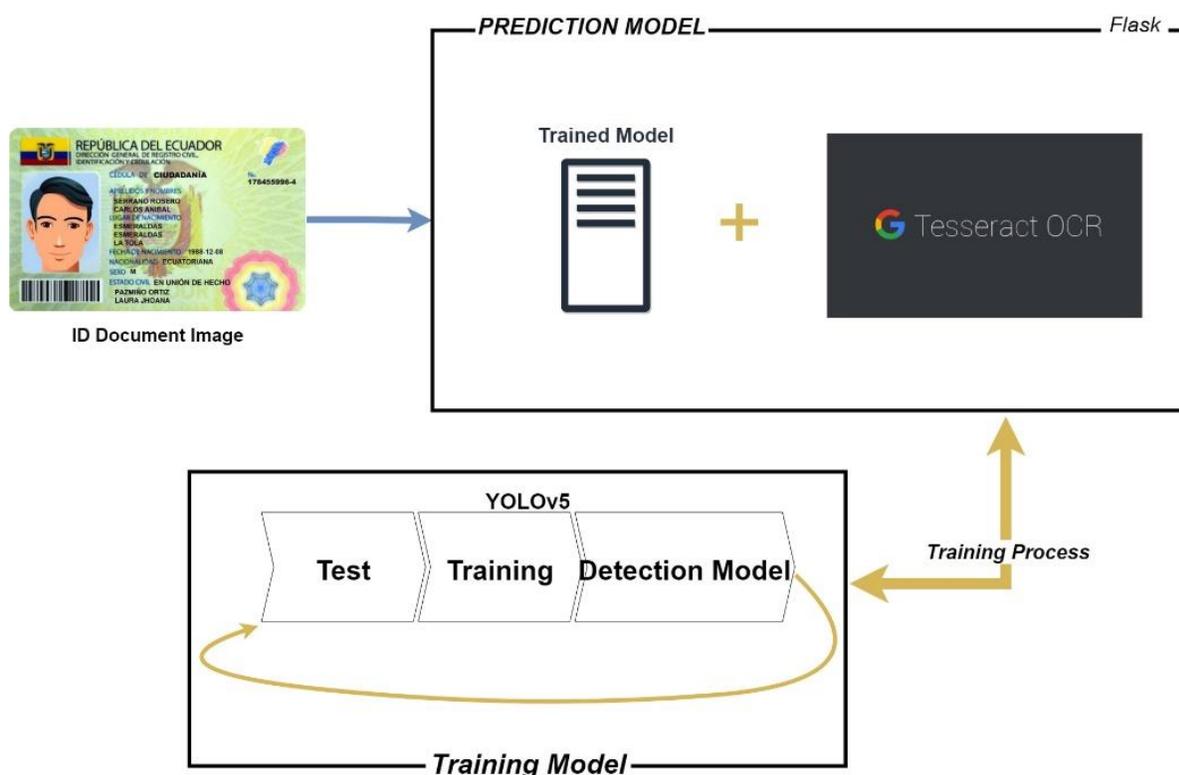


Figura 2: Arquitectura General del Modelo

### Modelo de Predicción

Se encuentra subdivido en dos componentes indispensables como lo son: Trained model y Tesseract OCR. Estos componentes permiten al modelo de predicción recibir documentos dados por el usuario y poder realizar la detección de los datos de interés para realizar la lectura de las imágenes por medio de tecnología OCR. Ambos componentes están contruidos mediante el framework de desarrollo Flask.

### Modelo de Entrenamiento

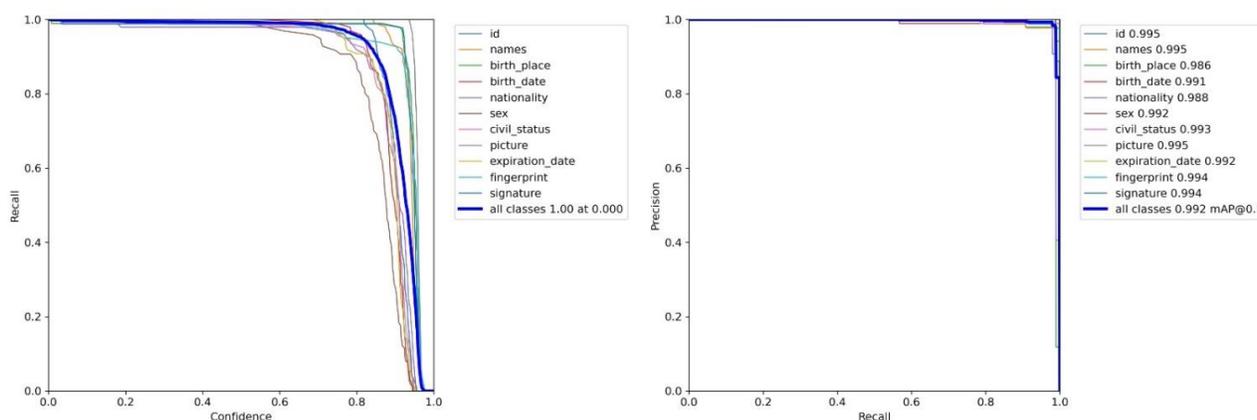
El modelo de entrenamiento y detección de objetos Yolov5 divide 3 tareas principales que permiten obtener un modelo robusto y de calidad de acuerdo a la configuración, labels e imágenes etiquetadas proporcionado. Estas tareas se dividen en test, entrenamiento, detección de objetos; las cuales son eterizadas con el fin de generar experimentos los cuales proporcionan datos importantes para obtener un resumen sobre los porcentajes con respecto al proceso de entrenamiento, para luego

## DetECCIÓN Y DIGITALIZACIÓN DE DATOS DE INTERÉS EN DOCUMENTOS DE IDENTIFICACIÓN

en base a esto usar generar el modelo personalizado en la detección y contextualización de la información requerida.

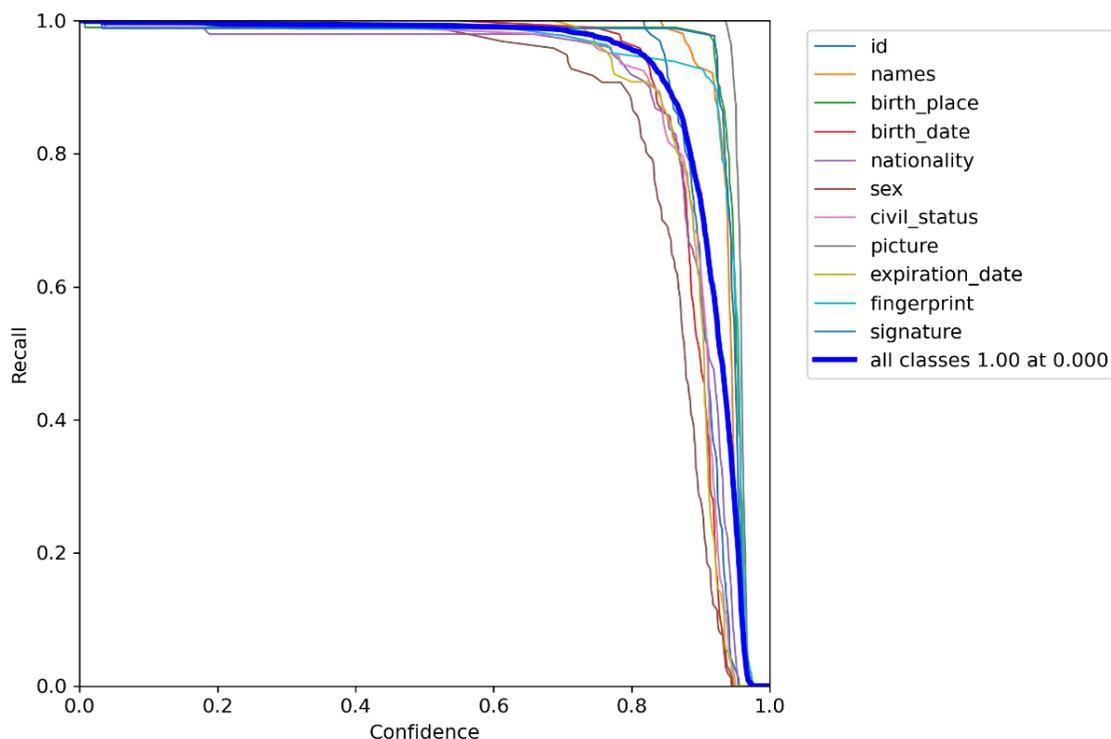
### Resultados

El algoritmo fue entrenado por medio de conjunto de imágenes previamente etiquetadas por los autores de este artículo usando la plataforma Make-Sense. Se experimentó con 600 imágenes en diferentes dimensiones, espacios y modelos de documentos. Se procesan los datos mediante YOLOv5. Este modelo de aprendizaje arrojó resultados que corresponden a una precisión de 93.7%, recall del 99.2% y un mAP de 100% según los experimentos ubicados en el servidor. En la figura 3 se muestra la curva de precisión y confianza en el aprendizaje del modelo de predicción y la relación entre la precisión y la recuperación de los datos relevantes en los documentos (*recall*) por último la figura 4 proporciona un vistazo entre la relación entre la confianza para la detección y la recuperación de los datos. Cada grafica muestra con qué porcentaje de precisión el modelo pudo entrenar la detección del campo con respecto a la confianza del modelo. Como resultado del entrenamiento obtendremos un modelo de detección estable y robusta el cual ofrece cumplir con las expectativas esperadas para la automatización de la detección de los parámetros de interés dentro de los documentos de identificación antiguos y actuales.



**Figura 3:** Relación de la precisión entre la confianza de los datos y la recuperación de los datos relevantes en los documentos.

## Detección y digitalización de datos de interés en documentos de identificación



**Figura 4:** Relación entre la confianza de la detección y la recuperación de los datos.

### Conclusiones

A diferencia de Tablex (Colter, y otros, 2022) y la propuesta dada por (Arslan, 2022). La plataforma desarrollada para la lectura de información en Documentos Ecuatorianos resultó ser una tecnología eficaz en la detección específica de estos datos en cualquier tipo de imagen, versión del documento y tamaño de la imagen a su vez que permite extraer la semántica detectada por el algoritmo y ser usada en procesos de validación institucionales. El modelo de entrenamiento arrojó resultados satisfactorios con respecto al aprendizaje del algoritmo y mediante la Liberia Tesseract-OCR nos permite mostrar la información en texto plano para poder ser usada por los usuarios. En comparación con otras plataformas como JAIDED AI, consiste en una herramienta que trabaja con aprendizaje y OCR integrado sin embargo este algoritmo detecta todos los textos de una imagen sin tomar en cuenta las posiciones, imágenes, dimensiones y tamaño de las imágenes, en comparación con esta plataforma nuestro algoritmo está orientado a una necesidad específica como lo es la

## Detección y digitalización de datos de interés en documentos de identificación

---

detección de datos relevantes en documentos Ecuatorianos los cual permitirá agilizar procesos de verificación de identidad en distintas entidades y así solventar estas necesidades.

### Agradecimientos

Los autores agradecemos al CITT de la Universidad Católica de Cuenca por brindar el dispositivo de entrenamiento y el área para las pruebas para el algoritmo. En especial a los estudiantes que nos proporcionaron su consentimiento para el uso de su información con el fin de realizar nuestro estudio.

### Referencias

1. Akinbade, D., Ogunde, A. O., Odim, M., & Oguntunde, B. O. (2020). An adaptive thresholding algorithm-based optical character recognition system for information extraction in complex images. *Journal of Computer Science*.
2. Almutairi, A., & Almashan, M. (2019). Instance Segmentation of Newspaper Elements Using Mask R-CNN. In 2019 18th IEEE international Conference On Machine Learning And Applications (ICMLA)(18), 1371-1375.
3. Arslan, H. (2022). End to End Invoice Processing Application Based on Key Fields Extraction. *EEE Access*, 10:78398-78413.
4. B Parvathi Sangeetha, E. K. (2022). Artificial intelligence based handwritten text recognition system. In *AIP Conference Proceedings*, 2393, 020095.
5. Chernyshova, Y. S., Sheshkus, A., & Arlazorov, V. V. (2020). Two-step CNN framework for text line recognition in camera-captured images.
6. Clausner, C., Antonacopoulos, A., & Pletschacher, S. (2020). Efficient and effective OCR engine training. *International Journal on Document Analysis and Recognition (IJ DAR)*, 73-88.
7. Colter, Z., Fayazi, M., Youbi, Z. B.-E., Kamp, S., Yu, S., & Dreslinski, R. (2022). Tablext: A combined neural network and heuristic based table extractor. *Array*, 15:100220.
8. Dergachov, K., Krasnov, L., Bilozerskyi, V., & Zymovin, A. (2021). Data pre-processing to increase the quality of optical text recognition systems. *Radioelectronic and Computer Systems*, 193-190.

Detección y digitalización de datos de interés en documentos de identificación

---

9. He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask r-cnn. Proceedings of the IEEE international conference on, 2961-1969.
10. Karthikeyan, S., Seco de Herrera, A. G., Doctor, F., & Mirza, A. (2021). An ocr post-correction approach using deep learning for processing medical reports. IEEE Transactions on Circuits and Systems for Video Technology, V(32), 2574-2581.
11. Khan, M. M., Uddin, M. S., Parvez, M. Z., & Nahar, L. (2022). A squeeze and excitation ResNeXt-based deep learning model for Bangla handwritten compound character recognition. Journal of King Saud University-Computer and Information Sciences, VI(34), 3356-3364.
12. Liao, M., Wan, Z., Yao, C., Chen, K., & Bai, X. (2020). Real-time scene text detection with differentiable binarization. In Proceedings of the AAAI conference on artificial intelligence(34), 11474-11481.
13. Lin, G.-S., Tu, J.-C., & Lin, J.-Y. (2021). Keyword Detection Based on RetinaNet and Transfer Learning for Personal Information Protection in Document Images. Applied Sciences, 9528.
14. Liu, Y., James, H., Gupta, O., & Raviv, D. (2022). MRZ code extraction from visa and passport documents using convolutional neural networks. International Journal on Document Analysis and Recognition (IJ DAR), I(25), 29-39.
15. Long, S., Ruan, J., Zhang, W., He, X., Wu, W., & Yao, C. (2018). TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes. Proceedings of the European conference on computervision (ECCV), 20.36.
16. Molina-Pérez J., P.-M. C. (2021). Covid-19 and "improvised" digitization in secondary education: Emotional tensions and challenged professional identity. Scopus, 181.
17. P. S., & Sundar K., J. A. (2023). Enhanced Attention-Based Encoder-Decoder Framework for Text Recognition. Intelligent Automation and Soft Computing, 2(35), 2071-2086.
18. Ramdhani, T., Budi, I., & Purwandari, B. (2021). Optical Character Recognition Engines Performance Comparison in Information Extraction. International Journal of Advanced Computer Science and Applications.
19. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in neural information processing systems(28).

Detección y digitalización de datos de interés en documentos de identificación

---

20. Rivas, R., Paul, S., Hristidis, V., Papalexakis, E. E., & Roy-Chowdhury, A. K. (2022). Task-agnostic representation learning of multimodal twitter data for downstream applications. *Journal of Big Data*, I(9), 1-19.
21. Salma, Saeed, M., Ur Rahim, R., Gufran Khan, M., Zulfiqar, A., & Bhatti, M. T. (2021). Development of ANPR Framework for Pakistani Vehicle Number Plates Using Object Detection and OCR. *Complexity*.
22. Sherrah, J. (2016). Fully convolutional networks for dense semantic labelling of highresolution aerial imagery. *arXiv preprint arXiv:1606.02585*.
23. Smith, R. (2017). An overview of the tesseract ocr engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*, II, 629-633.
24. Sun, W., & Wang, R. (2018). Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, III(15), 474-478.
25. Zhang, E., & Putra, V. A. (2022). Improving optical character recognition accuracy for indonesia identification card using generative adversarial network. *Journal of Theoretical and Applied Information Technology*, VIII(100).