Universidad de Lima

Facultad de Ingeniería

Carrera de Ingeniería de Sistemas

SISTEMA DE RECONOCIMIENTO FACIAL EN TIEMPO REAL BASADO EN CNN PARA EL CONTROL DE ACCESO DE LOS RESIDENTES DE UN CONJUNTO HABITACIONAL

Tesis para optar el Título Profesional de Ingeniero de Sistemas

Javier Hassler Guevara Burgos

Código 20160650

Asesor

Lennin Paul Quiroz Villalobos

Lima – Perú

Noviembre de 2024

Sistema de reconocimiento facial en tiempo real basado en CNN para el control de acceso de los residentes de un conjunto habitacional

Javier Hassler Guevara Burgos
20160650@aloe.ulima.edu.pe
Universidad de Lima

Resumen: En los últimos años, el reconocimiento facial se ha convertido en un modelo de referencia en el campo de la seguridad debido a su universalidad, singularidad, no intrusividad y facilidad de implementación siendo utilizado en sistemas orientados al monitoreo y control de acceso a instalaciones, sistemas de video vigilancia externos, búsqueda de personas desaparecidas, prófugos de la ley, entre otros. El reconocimiento facial, pertenece a los métodos biométricos, consiste en examinar las características fisiológicas de un individuo para determinar su identidad. No obstante, al paso que los desarrollos tecnológicos incrementan la eficiencia del reconocimiento facial, éste aún enfrenta retos como las oclusiones, el costo de las licencias de software propietario y factores como cambios en las expresiones faciales, iluminación, orientación del rostro, vello facial y cambios inherentes como el envejecimiento de las personas que dificultan un adecuado proceso de identificación. En el campo del Deep Learning, las Redes Neuronales Convolucionales (CNN) tienen múltiples aplicaciones en el campo de Visión Computacional (CV) ya que permiten extraer características propias, genéricas de los objetos, las cuales luego podrán ser utilizadas para reconocer objetos similares en otras imágenes. Las técnicas de reconocimiento facial utilizan diversos tipos de algoritmos (PCA, CNN, Transformers) para automatizar las tareas de reconocimiento e identificación de personas a partir de imágenes o videos provenientes de sistemas de videovigilancia. El sistema de reconocimiento facial implementado en este trabajo está basado en Redes Neuronales Convolucionales (CNN) utilizando la arquitectura VGG-Face, realizando el proceso de reconocimiento en tiempo real y de manera no intrusiva por lo que lo convierte en una opción óptima para instituciones, organismos, entidades donde se requieran instalar mecanismos de identificación eficientes a un costo mínimo. Con el sistema implementado se logró una exactitud de 84% para 50 casos de prueba de reconocimiento en video de 5 personas en un entorno no controlado con iluminación diurna y a rostro descubierto en el conjunto residencial Belo Horizonte North (BHN) ubicado en Av. Brasil 1636 Distrito de Pueblo Libre.

Palabras Clave: Reconocimiento facial, CNN, visión computacional, VGG-Face, Transfer-learning, detección de personas, PubFig83

Abstract: In recent years, facial recognition has become a reference model in the field of security due to its universality, uniqueness, non-intrusiveness and ease of implementation, being used in systems aimed at monitoring and controlling access to facilities, external video surveillance systems, search for missing persons, fugitives from the law, among others. Facial recognition, which belongs to the biometric methods, consists of examining the physiological characteristics of an individual to determine his or her identity. However, as technological developments increase the efficiency of facial recognition, it still faces challenges such as occlusions, the cost of proprietary software licenses and factors such as changes in facial expressions, lighting, facial orientation, facial hair and inherent changes such as the aging of people that hinder an adequate identification process. In the field of Deep Learning, Convolutional Neural Networks (CNN) have multiple applications in the field of Computer Vision (CV) since they allow extracting generic characteristics of objects, which can then be used to recognize similar objects in other images. Facial recognition techniques use various types of algorithms (PCA, CNN, Transformers) to

automate the tasks of recognition and identification of people from images or videos from video surveillance systems. The facial recognition system implemented in this work is based on Convolutional Neural Networks (CNN) using the VGG-Face architecture, performing the recognition process in real time and in a non-intrusive way, which makes it an optimal option for institutions, organizations, entities where efficient identification mechanisms need to be installed at a minimum cost. With the implemented system an accuracy of 84% was achieved for 50 test cases of video recognition of 5 people in an uncontrolled environment with daylight and open face in the residential complex Belo Horizonte North (BHN) located at Av. Brasil 1636 District of Pueblo Libre.

Keywords: Facial recognition, CNN, computer vision, VGG-Face, Transfer-learning, people detection, PubFig83

1. INTRODUCCIÓN

En la actualidad, los sistemas de reconocimiento facial tienen una gran importancia debido a que son la base de una gran cantidad de aplicaciones en los campos de videovigilancia, interacción hombre máquina y sistemas de seguridad para dispositivos IoT (Internet Of Things). El reconocimiento facial se ha convertido en un área de gran interés en el campo de la visión computacional por la gran cantidad de aplicaciones que esta tecnología tiene.

Inicialmente fue destinado para el uso de las fuerzas de seguridad (verificación de datos de pasaporte, DNI, licencias de conducir, etc.), luego surgió en aplicaciones comerciales con el objetivo de controlar accesos, identificación y vigilancia (Sims, 1994). Para implementar sistemas de reconocimiento facial se han propuesto varias técnicas, entre ellas: Principal Component Analysis (PCA) (Simancas, 2019), Linear Discriminant Analysis (LDA) (Shlens, 2003), Redes Transformers (Grágeda, 2023) y Redes Neuronales Convolucionales (CNN) (Cao et al., 2018).

Las redes transformers destacan por su capacidad de paralelización, lo que reduce el tiempo de entrenamiento. Sin embargo, necesitan grandes volúmenes de datos para identificar patrones locales. Por otro lado, las redes CNN, gracias a su estructura multicapa, detectan patrones locales durante el proceso de entrenamiento. Aunque esto puede ralentizar el entrenamiento, hace que las CNN sean más adecuadas para aplicaciones con conjuntos de datos pequeños (Vegas & Merayo, 2023).

La imagen de entrada es pieza fundamental puesto que una imagen de baja calidad puede ocasionar un mecanismo de reconocimiento deteriorado (Cook et al., 2019). En escenarios de aplicaciones reales es de suma importancia validar que la calidad de la imagen se encuentre apta para el reconocimiento de tal manera que el algoritmo pueda reconocer a la persona correctamente. Este tópico enfrenta diversos desafíos por superar ya que en un entorno no controlado a diferencia de un entorno controlado se dificulta la capacidad de reconocimiento de la persona debido a la variación de la iluminación, pose, expresión, sombra y enfoque. En particular las condiciones de iluminación pueden alterar la apariencia del rostro y los rasgos faciales para una correcta detección con resultados menos al 50% precisión (Ho Chan et al., 2018). Es decir, la calidad de la imagen de entrada y otros factores como expresiones faciales, orientación del rostro, vello facial, envejecimiento, deficiente presentación tal como la persona cabizbaja y uso de accesorios como anteojos, bufandas y sombreros provocan oclusiones en el rostro con limitaciones en el reconocimiento determinarán la exactitud del sistema de reconocimiento (Grother et al., 2019). En este aspecto, otra ventaja de las redes neuronales convolucionales es que han demostrado una tolerancia a imágenes por debajo del estándar, es decir, aquellas fotografías que se desvían de la ISO / IEC 39794-5 el cual es el estándar oficial para retratos.

En la actualidad, la tecnología ha optimizado nuestra calidad de vida al ahorrar tiempo y recursos en diversas tareas. Sin embargo, es raro encontrar tecnologías en el hogar que realmente

contribuyan a mejorar la seguridad y el confort en el interior de la vivienda. Por ello, se desarrolló la Domótica el cual permite la integración de sistemas de seguridad, comunicaciones y bienestar en la automatización de una vivienda. En efecto, la automatización en el hogar trae consigo una característica fundamental que es la seguridad el cual permite que el usuario posea un tipo de confort al saber que sus pertenencias tecnológicas y puertas principales están resguardados gracias a un sistema de seguridad de acceso como es el reconocimiento facial el cual permite la interacción personalizada del usuario con el sistema de seguridad en la activación o desactivación de una alarma teniendo un acceso personalizado (Castaño & Alonso, 2019).

También, la mayor parte de los sistemas biométricos actuales enfrentan el problema del reconocimiento facial múltiple dado que funcionan para una cara a la vez, dado que, cuando se implementa en múltiples rostros tiene un alto costo computacional y es difícil de implementar para un hardware con recursos limitados. En un futuro cercano, el objetivo principal sería implementar mejores arquitecturas en un conjunto de datos más grande y complejo (Chaturvedi & Vishwakarma, 2020).

La presente investigación tiene como objetivo implementar un sistema de reconocimiento facial en tiempo real basado en CNN para el control de acceso de los residentes del conjunto habitacional Belo Horizonte North (BHN) ubicado en Pueblo Libre. La investigación posee 2 etapas en la primera se realizó pruebas con imágenes fijas utilizando el dataset PubFig83 en un entorno controlado y la segunda corresponde al sistema en tiempo real usando imágenes de video y un dataset propio en condiciones de iluminación diurna en un entorno no controlado y con/sin la presencia de elementos de oclusión como gorro, lentes y audífonos.

En la sección 2 del presente documento se examinan los resultados obtenidos y las bases de datos empleados por los autores de diversas investigaciones previas relacionadas. Luego en la sección 3 se explican los conceptos fundamentales, principios y técnicas en los que se basa la investigación para la implementación del sistema de reconocimiento facial. En la sección 4 se expone la metodología seguida para la investigación. En la sección 5 se muestra el proceso de experimentación. En la sección 6 se muestran los resultados obtenidos y las técnicas aplicadas para su medición. En la sección 7 se realiza la discusión de los resultados obtenidos. Finalmente, en la sección 8 se presenta la validación y en la sección 9 las conclusiones.

2. ESTADO DEL ARTE

Las redes neuronales convolucionales (CNN) han demostrado que tienen un nivel de acierto y rendimiento muy alto respecto al reconocimiento y clasificación de imágenes, donde domina el campo de la biometría. Una propuesta realizada fue diseñada en base a la arquitectura CNN con versión Beta23, realizan una evaluación empírica con base datos representado por 15 imágenes con restricciones de fondo desordenado y el objetivo fue demostrar la superioridad de CNN respecto al análisis discriminante lineal (LDA) y el análisis de componentes principales (PCA) demostrando una tasa de precisión más alta con un promedio de 98% de similitud (Coskun et al., 2017).

La extracción de la representación facial de dos modelos distintos de aprendizaje profundo basados en Lightened CNN y VGG-Face, centrado en el diseño estructural de VGG Net. Ambos modelos son robustos para diferentes tipos de desalineación y toleran errores de localización de la distancia intraocular. Al emplear este algoritmo de aprendizaje profundo se puede obtener resultados prometedores mayores al 50% de similitud (Qawaqneh et al., 2017). Una etapa importante de los modelos mencionados anteriormente es el preprocesamiento, por ejemplo, la normalización de la pose y la iluminación para lograr un mejor rendimiento en diversas condiciones. La alineación de imágenes basados en un modelo 3D y uso conjunto de CNN se puede obtener precisiones de 97.35% y 91.4% en los conjuntos de datos LFW e YTF. Cabe

destacar que la localización errónea de la función facial es el principal responsable de diferentes fuerzas y desalineación por la iluminación, la oclusión de la cara superior e inferior e incluidos los diferentes ángulos de postura de la cabeza (Mehdipour & Kemal, 2016). De este modo, también es considerable tomar en cuenta la escala de la imagen ya que se puede agrupar por la cantidad de píxeles en sus 3 escalas: pequeño(10-50 píxeles), mediano(50-300 píxeles) y grande(más de 300 píxeles) puesto que las tasas de detección varían de acuerdo a la escala y cantidad de imágenes propuestas; y un rostros es considerado parcialmente ocluido cuando representa entre 1% y 30% del total del rostro pero si supera el límite de 30% es considerado con muy ocluido (Yang et al., 2016).

En el campo de la biometría es muy usual el uso de la aplicación de Deep Learning pero con visión a solucionar problemas relacionados a las oclusiones del rostro, desalienación e imágenes en diferentes ángulos. El uso de Lightened CNN y VGG-Face demuestran ser modelos robustos para diferentes tipos de desalineación y con tolerancias de errores de localización de la distancia intraocular. Sobre la base de las ideas expuestas, los sistemas biométricos aplicados al reconocimiento facial son afectados por la fiabilidad y seguridad de estos por el intervalo de tiempo de respuesta. Gracias al aprendizaje profundo VGG-Face se puede realizar la extracción automática de características produciendo características altamente discriminatorias e interoperables que son robustas para las variaciones antiguas. Los resultados de clasificación para los métodos de autenticación globales propuestos en los desafiantes conjuntos de datos de la red de reconocimiento de rostros y gestos (FG-NET) y base de datos longitudinal MORPH son competitivos puesto que la prueba realizada al periodo del rendimiento biométrico dio como resultado para FG-NET de 92.0% y MORPH 80.6%. MORPH obtuvo un rendimiento superior que FG-NET ya que este presentó una variabilidad inferior de interclase en la base de datos (El Khiyari & Wechsler, 2016).

Cabe decir que un enfoque relevante es la comparación entre dos conjuntos de datos: el conjunto de datos FERET y el conjunto de datos Caras etiquetadas en la naturaleza (LFW). El primero utiliza la distancia angular con el objetivo de identificar mientras que el segundo conjunto de datos es verificar una coincidencia dada gracias máquina de vectores de soporte (SVM) y aprendizaje métrica de similitud (SML) (Lumini et al., 2017). Por otro lado, diversos autores han buscado poder mejorar una red neuronal convolucional para mejorar efectivamente las tasas de reconocimiento para evitar incremento de falsos positivos. Tal es el caso de la red clásica LeNet LeNet-5 que puede realizar automáticamente la reducción de dimensionalidad de datos, extracción de características y resultados de clasificación. Este algoritmo fue puesto a prueba en diversos datasets como ORL Face, AR Face y rostros etiquetados en la naturaleza (LFW) con una alta tasa de reconocimiento en la base de datos facial en comparación de los métodos red neuronal artificial (ANN) y redes neuronales convolucionales (CNN) comúnmente utilizados (Lan et al., 2019).

En relación a la problemática expuesta se evidencia que las imágenes y videos comunes están orientados primordialmente en las personas. En esta perspectiva las relaciones entre los objetos, proporcionan claves contextuales complementarias para el reconocimiento. En consecuencia, otra propuesta de investigación estuvo basada en un modelo CNN que predice ubicaciones probables y escalas de objetos mediante la producción de un mapa de calor a escala múltiple para toda la imagen. Los modelos usados fueron el modelo local Faster R-CNN y un modelo global con un conjunto de datos denominado HollywoodHeads extraído de 21 películas de Hollywood, teniendo resultados favorables respecto a las detecciones que tienen una alta relación de superposición con la verdad básica se consideran verdaderos positivos (Vu et al., 2015).

De otro modo el reconocimiento facial también es aplicado al ámbito de detección confiable siendo de suma importancia. En igual forma también existe la propuesta de reconocedor de expresiones faciales en tiempo real usando una red neuronal profunda. Para identificar las

expresiones faciales se debe evaluar condiciones de individualidad y visión, donde un sistema basado en CNN es robusto a la variabilidad y posee una detección confiable con una tasa de reconocimiento de 97,6%. Esto demostraría que utilizar CNN tiene una tasa alta de reconocimiento en la aplicación de detección de rostros (Matsugu et al., 2003). En consecuencia, (Jeon et al., 2016) sustentan la propuesta de un modelo de red neuronal convolucional fundamentando en Tang para explotar el trabajo del reconocimiento de expresiones en el rostro. Dicho modelo obtuvo un tasa de reconocimiento de 69.77%, Tang utiliza el modelo de red neuronal convolucional utilizó como clasificador a SVM lineal y el histograma de gradientes orientado (HOG) para la detección de rostros con la intención de construir un reconocedor de expresiones faciales en tiempo real utilizando una red neuronal profunda que es invariable para el sujeto. En este sentido se comprende que Las aplicaciones de reconocimiento rostros del mundo real presentan diversos tipos de degradaciones ocasionados por el desenfoque del movimiento, ruido, artefactos de compresión, distorsiones de color y oclusión. El rendimiento de una red neuronal se ve afectado por el desenfoque, el ruido y la oclusión pero un modelo CNN profundo es resistente a la distorsiones, como las distorsiones de color y el cambio en el balance de color (Karahan et al., 2016).

Un nuevo método moderno es el transfer-learning el cual permite el ahorro de tiempo de un modelo mediante el uso de modelos pre-entrenados. La finalidad del aprendizaje por transferencia es emplear en un entorno nuevo las nociones aprendidas de un entorno ya existente. El modelo Inception-v3 de la plataforma TensorFlow empleó técnicas de transferencia de aprendizaje para reentrenar el conjunto de datos de expresión facial, que puede mantener la precisión del reconocimiento y reducir en gran medida el tiempo de entrenamiento, con resultados de la precisión de clasificación del modelo es del 97%, que es más alta que la de MLP, LBP y el modelo de red de bajo nivel (Xia et al., 2017).

3. ANTECEDENTES

El propósito de un sistema de reconocimiento facial es reconocer de forma automática a una persona en una imagen de video en tiempo real el cual puede clasificarse en dos tipos de escenarios (i) verificación facial (o autenticación) y (ii) identificación facial (o reconocimiento). (Lu et al., 2003). Un punto por destacar es el costo computacional que genera el reconocimiento facial que varía de acuerdo a la entrada de píxeles tanto en ancho y alto puesto que normalmente una imagen de un rostro de sólo 50 píxeles de ancho y 50 de alto tendría una dimensionalidad de 2500.Esto hace el problema de reconocimiento de rostros altamente costoso desde el punto de vista computacional. (Fuentes, 2011).

Asimismo, el rostro humano otorga gran cantidad de información de diferenciación de un sujeto pudiendo distinguir y reconocer de forma sencilla diferentes personas. Los rasgos más significativos del rostro son los ojos, nariz, cejas ,boca y orejas, pero el rasgo con mayor variabilidad es el pelo ya que permite esconder ciertos rasgos y transformar el aspecto de un individuo mediante los elementos de oclusión tales como el uso de gafas, gorras y chalinas (Ríos, 2013).

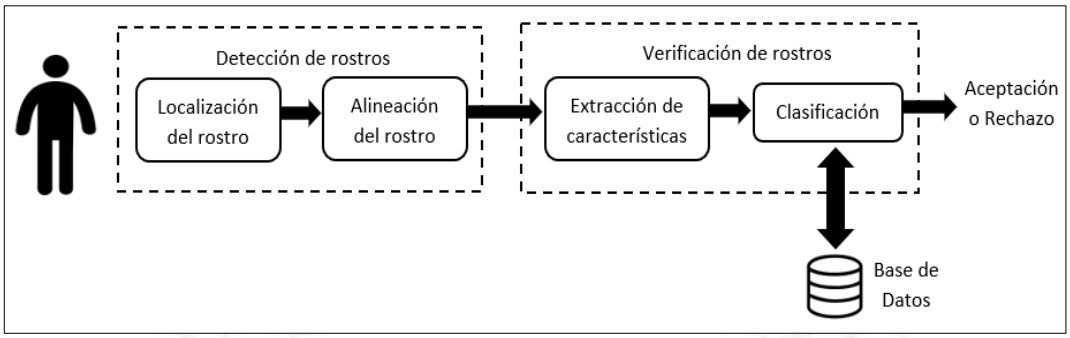
3.1 Etapas de un sistema de reconocimiento facial

En la Figura 1 se muestran las principales etapas del proceso de un sistema de reconocimiento facial estándar, pero en la presente investigación solo será tomado como referencia. La primera etapa consiste en la captura de la imagen que se puede realizar mediante cámaras, videos o webcams que faciliten una imagen que tenga un rostro a analizar. La segunda etapa consiste en la detección del rostro donde se va a realizar la localización de la cara de la persona y una posible alineación de acuerdo a las condiciones de la imagen. La detección de regiones de interés del rostro se lleva a cabo mediante la extracción de características de Haar, el clasificador (Haar Features) extrae características distintivas de los rostros, como la forma de los ojos, la nariz y la boca y hace

uso de métodos de aprendizaje automático para procesar gran cantidad de imágenes en corto tiempo. El algoritmo de Haar está basado en Adaboost asegurando que sea un clasificador eficiente (Viola & Jones, 2001).

Figura 1

Etapas de un sistema de reconocimiento facial estándar

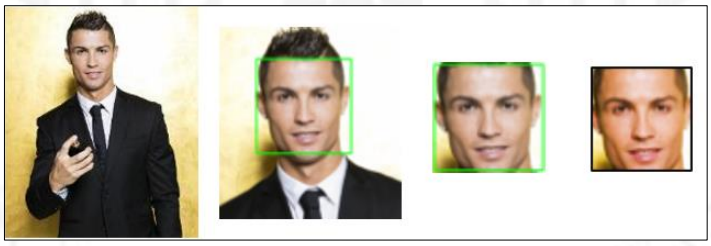


Nota. Adaptada de "Diseño e implementación de una herramienta de detección facial", por M. García, 2009 (http://tesis.ipn.mx/bitstream/handle/123456789/6111/DISENOIMPLEMFACIAL.pdf).

En la tercera etapa, se lleva a cabo la normalización y el escalado de la imagen. Esta fase prepara la imagen para la siguiente etapa, que consiste en la extracción de características mediante una serie de transformaciones geométricas. La normalización se realiza en cuatro fases, que también incluyen el alineamiento si es necesario. La Figura 2 ilustra el proceso de detección y normalización del rostro en la imagen ingresada, enfocándose únicamente en el área de interés y descartando el resto de la imagen.

Figura 2

Proceso de detección y normalización



En la cuarta etapa se realiza la extracción de características para conseguir solo información notable de la imagen. Esta etapa tiene el objetivo de la extracción de características principales de la imagen y se almacenan los rostros en la base de datos a partir de un conjunto de imágenes capturadas para poder realizar posteriormente la comparación respectiva. En la siguiente subsección se describen los principales métodos o técnicas para reconocimiento facial.

3.2 Técnicas para Reconocimiento Facial

3.2.1 Métodos basados en Redes Neuronales Convolucionales (CNN)

A. VGG-Face

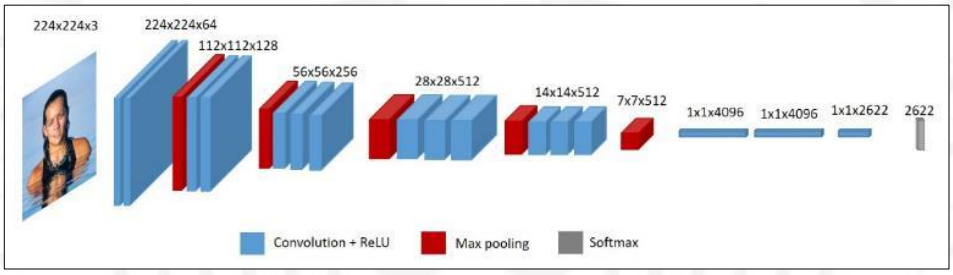
El modelo VGGFace fue desarrollado por el Grupo de Geometría Visual en Oxford el cual utiliza redes neuronales convolucionales. Los descriptores de VGG-Face se sustentan en la arquitectura VGGNet conformado por una secuencia de capas convolucionales, capas agrupadas y capas conectadas completamente. La primera capa utiliza filtros de dimensión 3 mientras que la capa de agrupación realiza un submuestreo con factor 2 (Parkhi et al., 2015). Este modelo planteó recoger

el mayor conjunto de datos de internet utilizando su red VGGNet. Dicho conjunto de datos fue entrenado, posteriormente se realizó un ajuste a las redes neuronales mediante una función de pérdida de tripletes obteniendo un resultado de precisión de 98.95% (Cao et al., 2018).

Asimismo, la arquitectura VGGNet se destaca por la flexibilidad que posee para asimilar el aprendizaje de mapeos no lineales a través de arquitecturas profundas tales como la red de 16 y 19 capas de peso teniendo como entrada la imagen de dimensión de 224x224 píxeles que permite un estándar en el conjunto de datos y estabilidad en el algoritmo de optimización (Simonyan & Zisserman, 2014). En otros términos, en su naturaleza toma la imagen de entrada y clasifica la región de interés del objeto mediante un conjunto de clases predefinidas, donde se obtiene un nivel de precisión alto, pero esto implica un mayor tamaño del modelo. De esta forma también es aplicado en el ámbito del reconocimiento facial, puesto que calcula la incrustación de un rostro en uno nuevo de prueba, luego se compara dicha incrustación con las demás incrustaciones de rostros de la base de datos y se determina un veredicto de similitud. A continuación, en la Figura 3 se muestra la visualización de la arquitectura del modelo VGG-Face. VGG-Face utiliza Transfer Learning y Data Augmentation, de manera opcional.

Figura 3

Visualización VGG-Face



Nota. Adaptada de "*Deep face recognition", por* Parkhi et al., Proceedings of the British Machine Vision Conference (pp. 41.1–41.12), 2015 (https://doi.org/10.5244/c.29.41)

B. Transfer-learning

Es una metodología de diseño dentro del aprendizaje automático, pero no un modelo o técnica de aprendizaje. Consiste en aprovechar el conocimiento aprendido de las tareas existentes que tienen diversos datos etiquetados disponibles aplicados en entornos de pocos datos etiquetados, es sustancial beneficiarse de forma óptima la base de datos existente. El objetivo del aprendizaje de transferencia es comenzar el proceso de generalización de patrones aprendidos para diferentes tareas (Christanti et al., 2018). Por otro lado, es común que la clasificación de las imágenes represente un problema relevante en el área de visión computacional, que busca optimizar lo aprendido en un nuevo entorno (Gao & Mosalam, 2018). Por ello, se busca reducir la función de costo de las redes neuronales convolucionales para poder extraer las características recopiladas de forma automática gracias a capas apiladas. Por último, la última etapa es la comparación con la base de datos, aquí se aplicará los algoritmos de reconocimiento y se tomará una decisión de acuerdo con el porcentaje de precisión obtenido.

C. Data Augmentation

Debido a que el número de imágenes de rostros originales es corto para el entrenamiento de un modelo CNN profundo, un método frecuente es incrementar el conjunto de entrenamiento gracias al método de aumento de datos generando múltiples imágenes virtuales partiendo de cada imagen original mediante transformación geométrica, alteración al brillo de la imagen y filtrado (Pei et al., 2019). De esta forma, al aplicar técnicas de aumento de datos se realizan transformaciones que

conservan las etiquetas y generalmente empleado a las imágenes de entrenamiento. El objetivo de las técnicas de aumento de datos es mejorar la precisión de los métodos basados en CNN y evitar el sobreajuste (Masi et al., 2019).

D. Métodos Holísticos

a. PCA (Principal Component Analysis)

El método de Análisis de Componentes Principales (PCA) es un método estadístico empleado para disminuir la dimensionalidad, es decir, describe al conjunto de datos en nuevas variables no correlacionadas entre sí denominados componentes. El algoritmo PCA realiza una proyección de las variables donde estas queden interpretados en términos de mínimos cuadrados con el objetivo de que un conjunto de observaciones de variables correlacionadas entre sí, se transformen en nuevo conjunto de valores sin correlación lineal denominado componentes principales (Simancas, 2019).

b. LDA (Linear Discriminant Analysis)

Es también denominado análisis discriminante de Fisher, es una técnica de reducción de la dimensionalidad. Esta técnica se emplea para la selección de características mediante la distancia euclidiana siendo un método supervisado, y en comparación de PCA, utiliza la información de la etiqueta para mejorar la separabilidad entre las distintas clases y pretenden minimizar la variación dentro de una misma clase. De esta forma, este método no aborda las dependencias estadísticas de alto orden, por ejemplo, la relación entre varios píxeles. No obstante, LDA presenta una limitación relevante la cual es que requiere de matrices de dispersión no-singulares. Este problema es conocido como "singularity problem" donde la alta dimensionalidad de las imágenes es mucho mayor que el número de imágenes, pero para evitar este efecto se aplica diferentes aproximaciones como "pseudo-inverse" LDA (Shlens, 2003).

E. Redes Transformer

La red neuronal tipo transformer es un tipo de modelo deep learning que surgió por su capacidad para gestionar datos secuenciales. Este tipo de red posee un diseño codificador-decodificador el cual ha sido utilizado en una extensa gama de tareas de procesamiento de lenguaje natural, así como en ámbitos para el reconocimiento del habla y predicción de series temporales. Además, la arquitectura transformer se ha empleado en modelo preentrenados tales como BERT, GPT-2 y GPT-3 (Grágeda, 2023).

La estructura de red transformer está compuesta por el Input embedding, la capa codificadora y la capa decodificadora donde se inicia con la muestra de datos de la red en forma de números para la creación de un vector para cada palabra de los datos de entrada, luego procesa la secuencia de entrada mediante una serie de capas proporcionando un conjunto de datos ocultos que se transfieren a la capa decodificadora obteniendo la secuencia de salida y por último, se genera como resultado un conjunto de salidas a partir de un token de inicio junto a determinadas palabras de la secuencia de entrada y la salidas de la capa decodificadora. Sin embargo, la red transformer posee limitaciones tales como la alta carga computacional durante el entrenamiento por su arquitectura avanzada puesto que es intensiva en recursos y consume una cantidad importante de tiempo y potencial computacional lo que origina que limite su aplicabilidad en entornos con recursos acotados o donde se solicite una respuesta en tiempo real; dificultad para parametrizar y la capacidad de interpretar los resultados generados (Vegas & Merayo, 2023).

4. METODOLOGÍA

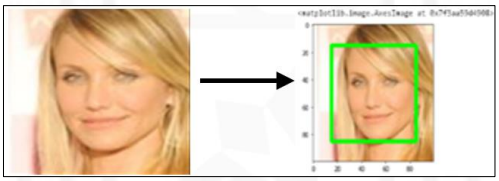
Esta sección contiene la descripción a detalle de los pasos realizados para obtener el reconocimiento facial en base a la coincidencia más cercana. Inicialmente, en la primera parte del presente proyecto, para verificar el correcto funcionamiento del modelo VGG Face se utilizó el

subconjunto mejorado de datos faciales PubFig83 del conjunto de datos principal PubFig propuesto por la Universidad de Columbia (Kumar et al., 2009). El conjunto de datos PubFig83 incluye 100 imágenes de cada una de 83 celebridades. No obstante, para reducir el costo computacional en este estudio, se utilizaron únicamente las imágenes de 49 celebridades, con 70 imágenes por persona (50 para entrenamiento y 20 para pruebas). Estas imágenes se capturan en situaciones totalmente caóticas con personas que no colaboran. Por lo tanto, hay una gran variabilidad en la pose, la iluminación, la expresión, la escena, la cámara, las condiciones y los parámetros de la imagen, entre otros.

Por otro lado, el conjunto de datos PubFig es un conjunto de datos faciales de personas del mundo real está compuesto por 58,797 imágenes de 200 personas recopiladas de internet. Estas imágenes fueron tomadas en situaciones completamente no controladas con sujetos no cooperativos. Por ende, se encontrará gran variación en pose, iluminación, expresión, escena, condiciones de imagen, parámetro, etc. Este conjunto de datos PubFig es análogo al conjunto de datos de caras etiquetadas en la naturaleza (LFW) pero poseen diferencias ya que la base de datos LFW posee 13,233 imágenes de 5,749 personas siendo más amplio que PubFig. No obstante, es menos profundo al poseer menor cantidad de imágenes por persona en promedio y más pequeño en comparación con PubFig (Kumar et al., 2009). En la Figura 4, se muestra la captura de la imagen de entrada y posterior la detección de caras dentro de la imagen donde se realizará el recorte del área de interés, el cuadrado de color verde, y se usará una red neuronal convolucional capaz de mapear las características de una cara humana en una representación numérica (embedding).

Figura 4

Pruebas con el dataset PubFig83



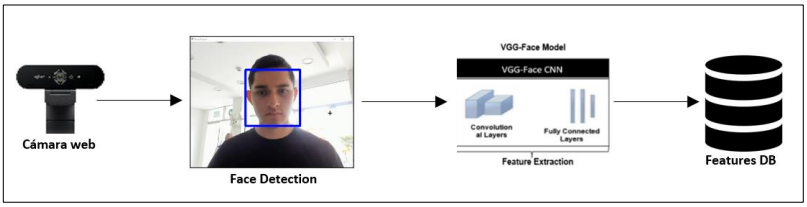
Luego de confirmar que el modelo funcionaba adecuadamente con el dataset PubFig83, se procedió a entrenarlo utilizando nuestro propio conjunto de datos, tal como se indica a continuación:

4.1 Fase de Entrenamiento

En la Figura 5, se aprecia la “fase de entrenamiento” la cual inicia con el ingreso de la imagen del conjunto de datos al algoritmo de detección de rostros, luego pasa a la etapa de extracción de características usando el VGG Face Model, las cuales finalmente se almacenan en la base de datos de features. En la etapa de entrenamiento, inicialmente se creó una base de datos con 5 sujetos de prueba con 15 imágenes por persona con resolución de imagen 1280x720 píxeles. Luego, a esta base de datos se añadió 10 personas de la base datos FEI la cual es un conjunto de datos de rostros brasileños elaborado por el Laboratorio de Inteligencia Artificial de FEI en São Bernardo do Campo, São Paulo (Tenorio & Thomaz, 2011), con resolución de 640x480 píxeles y 15 imágenes por persona. Posteriormente, se añadió los rostros de 15 personas de la base de datos de Georgia Tech con 15 imágenes por persona con fondo desordenado y una resolución de 640x480 píxeles (Nefian & Hayes, 2000). La base de datos de entrenamiento quedó compuesta por 30 personas, con 15 imágenes por persona y una resolución uniforme de 640x480 píxeles. Posteriormente se usó Data Augmentation, para tener un total de 75 imágenes por cada persona en la etapa de entrenamiento. Es preciso indicar que usar más de un dataset para validar un modelo de machine learning ayuda a garantizar que el modelo sea generalizable, robusto y fiable en una variedad de situaciones y datos, lo cual es fundamental para aplicaciones prácticas y para asegurar un buen desempeño en el mundo real.

Figura 5

Fase de Entrenamiento (I) - Diagrama de bloques del sistema de reconocimiento facial basado en CNN del conjunto de datos en tiempo real

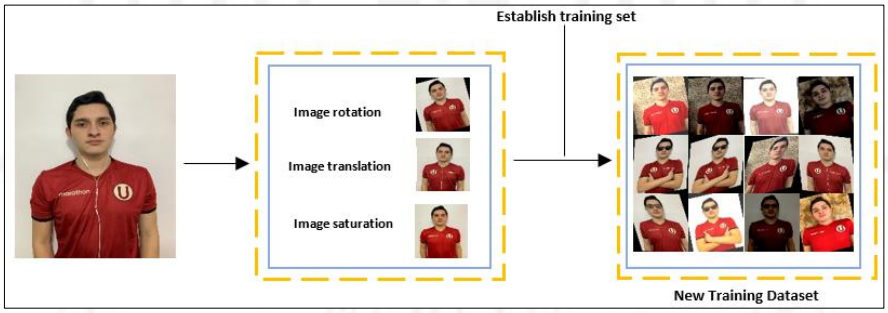


4.2 Data Augmentation

La data augmentation comúnmente es aplicada a las imágenes de entrenamiento mejorando la precisión de los métodos basados en CNN y evitar el sobreajuste. Conforme a ello, en la segunda parte del proyecto se aplicó data augmentation a la base de datos propia con 30 sujetos de prueba el cual cada sujeto tenía 15 imágenes por persona. Luego, se aplicó la técnica de aumento de datos y se obtuvo 75 imágenes por cada persona del dataset propio en tiempo real donde se aplicó los siguientes filtros a las imágenes de prueba: cambio de orientación de la imagen, variabilidad de la posición en 15° grados y saturación de +-25%. Para mayor detalle, la Figura 6 muestra el proceso que se realizó para llevar a cabo el proceso de data augmentatiton desde el reconocimiento del rostro de la persona el cual sirve para aumentar la cantidad de imágenes al conjunto de datos, luego la red neuronal convolucional (CNN) genera un modelo que sirve para reconocer a la persona.

Figura 6

Workflow del proceso de data augmentation con dataset propio en tiempo real



4.3 Detección del rostro

A continuación, se expone el algoritmo de Detección Facial utilizando Haar Cascade a través de la biblioteca OpenCV. Este algoritmo está diseñado para identificar rostros en imágenes digitales utilizando técnicas de Deep Learning. En la Figura 7, se ilustra la detección de rostros en sujetos del propio Dataset mediante el algoritmo de Detección Facial. Este algoritmo se aplica para la detección y reconocimiento de rostros en una imagen específica. A continuación, en el Algoritmo I se detalla cada etapa del proceso:

Primero, se define la ruta de la imagen, especificando su ubicación para el procesamiento. En este caso, la imagen de Mark Hamill se encuentra en una carpeta específica en Google Drive. A continuación, se carga un archivo XML que contiene un clasificador Haar, un algoritmo basado en características utilizado para la detección de objetos, con el fin de identificar rostros en la imagen. Luego, la imagen se lee desde la ruta especificada y se convierte a escala de grises, lo que simplifica el procesamiento y mejora la eficiencia en la detección de rostros. Posteriormente, se aplica el clasificador Haar a la imagen en escala de grises para detectar los rostros, y la función detectMultiScale devuelve las coordenadas de los rostros encontrados. Para cada rostro detectado,

se dibuja un rectángulo verde alrededor en la imagen original, lo que permite visualizar los rostros. Después, se abre la imagen original utilizando la biblioteca PIL (Python Imaging Library) y se calculan las coordenadas del centro del primer rostro detectado. Estas coordenadas se usan para definir una región de recorte alrededor del rostro, asegurando que el tamaño de esta región no exceda las dimensiones de la imagen original. Se define una caja de recorte alrededor del rostro, y se recorta la imagen original para obtener solo la región que contiene el rostro, la cual se redimensiona a 224x224 píxeles. Finalmente, se emplea un modelo de reconocimiento facial para realizar predicciones sobre la imagen recortada, tanto con transformación como sin transformación. Este algoritmo es especialmente útil en tareas de reconocimiento facial donde se requiere identificar y procesar rostros en imágenes.

Algoritmo I

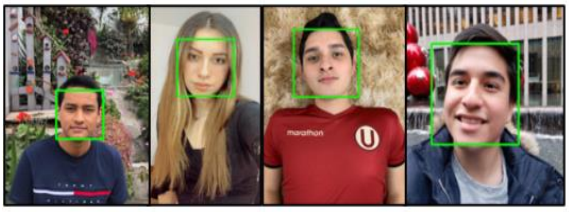
Algoritmo Face Detection

```
Importar librerías numpy, cv,matplotlib.pyplot
##Definición de la ruta de la imagen
rutadeimagen = '/content/drive/My Drive/Mark Hamill/Mark_Hamill1.jpg'
##Carga del clasificador de rostros
faceCascade = cv2.CascadeClassifier("/content/drive/My Drive/haarcascade_frontalface_default.xml")
##Lectura y conversión de la imagen a escala de grises
image = cv2.imreadrutadeimagen
gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
rostros = faceCascade.detectMultiScale(      ##Detección de rostros en la imagen
   gray,
   scaleFactor=1.1,
   minNeighbors=5,
   minSize=(30, 30))
rostros = faceCascade.detectMultiScale(gray, 1.2, 5)
Imprimir("Found {0} faces!".format(len(faces)))  ##Impresión del número de rostros encontrados
Para (x, y, w, h) en rostros:            ##Dibujo de rectángulos alrededor de los rostros detectados
   cv2.rectangle(image, (x, y), (x+w, y+h), (0, 255, 0), 2)
FinPara
plt.imshow(image)   ##Se utiliza para mostrar datos como una imagen
im = Image.open(rutadeimagen)  ##Apertura de la imagen original
(x, y, w, h) = faces[0] ##Cálculo del centro del primer rostro detectado
center_x = x + w / 2
center_y = y + h / 2
##Cálculo de las dimensiones del recorte
b_dim = min(max(w, h) * 1.2, im.width, im.height)
##Definición de la caja de recorte y recorte de la imagen:
box = (center_x - b_dim / 2, center_y - b_dim / 2, center_x + b_dim / 2, center_y + b_dim / 2)
crpim = im.crop(box).resize((224, 224))
plt.imshow(np.asarray(crpim))
##Predicción utilizando el modelo de reconocimiento facial
pred(facemodel, crpim, transform=False)
pred(facemodel, crpim, transform=True)
FinAlgoritmo
```

Algoritmo I. Pseudocódigo del algoritmo Face Detection

Figura 7

Face Detection con sujetos de prueba del dataset propio

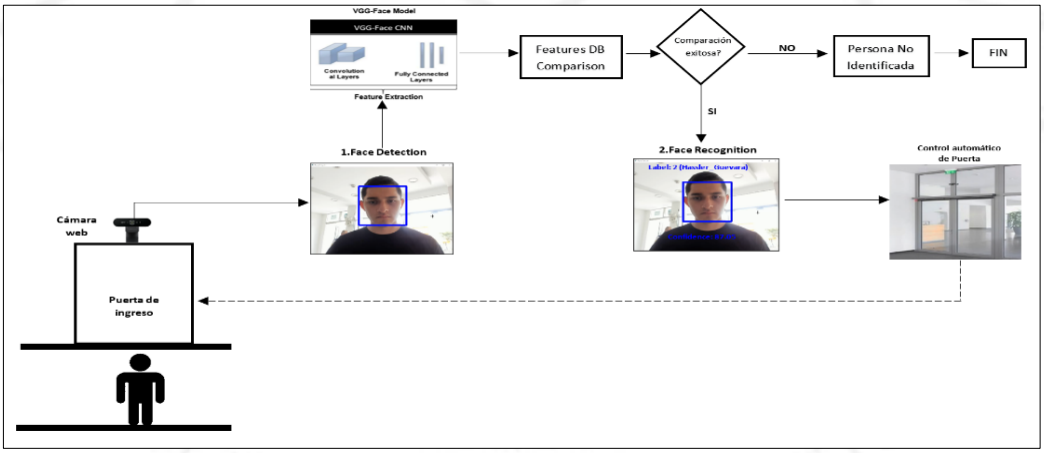


4.4 Fase de Reconocimiento

En la Figura 8, se aprecia la “fase de reconocimiento”. Se muestra el diagrama de arquitectura del sistema de reconocimiento en tiempo real, así como el sistema de control en lazo cerrado para la puerta. Este diagrama inicia comienza con la entrada de la cámara web. Luego, se detectan los rostros en la imagen capturada por la cámara. Enseguida, se extraen las características faciales utilizando el Modelo VGG-Face y descriptores CNN de VGG-Face y las características extraídas se comparan con las almacenadas en una Base de Datos de Características (Features DB). En este punto se realiza la verificación de la comparación, si es exitosa se procede al Reconocimiento Facial y se controla automáticamente la puerta de ingreso, permitiendo el acceso, en caso no sea existosa el proceso termina indicando “Persona No Identificada”. Finalmente, el proceso finaliza ya sea con la identificación exitosa y el control de acceso, o con la no identificación de la persona.

Figura 8

Fase de Reconocimiento (II) - Diagrama de arquitectura del sistema de reconocimiento facial en tiempo real como un controlador en lazo cerrado



4.5 Identificación de la persona

En esta etapa se realiza la comparación de las features generadas a partir del nuevo frame de la persona no identificada con las features existentes en la base de datos. El Algoritmo Face Recognition **s**e encarga de determinar de quién es el rostro analizado. En el Algoritmo II, se muestra cómo el algoritmo captura imágenes en tiempo real, detecta caras, y utiliza un modelo de reconocimiento facial entrenado para identificar a las personas en las imágenes. De este modo, en el bucle principal se captura un fotograma de la webcam, luego se detectan las caras en dicho fotograma utilizando una función de detección facial (fr.faceDetection). A continuación, se dibujan rectángulos alrededor de las caras detectadas para visualización, y la imagen se redimensiona y se muestra en una ventana llamada "Face Detection". Para cada cara detectada, se extrae la región de interés (ROI) en escala de grises. Luego, el reconocedor facial (face_recognizer.predict) predice la etiqueta y la confianza de la cara. Si la confianza es menor a 39, se dibuja un rectángulo y se muestra el nombre predicho en la imagen. Además, en la Figura 9, se muestra un ejemplo de reconocimiento usando una figura del data set propio.

Algoritmo II

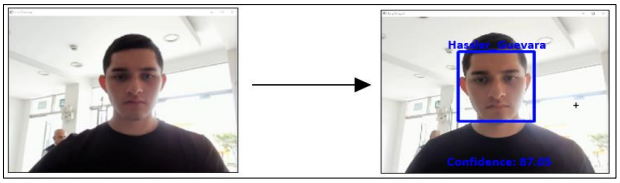
Algoritmo Face Recognition

```
##Importación de Módulos
Importar os
Importar cv2
Importar numpy as np
Importar faceRecognition as fr
##Creación del Reconocedor Facial
face_recognizer = cv2.face.VGGFaceRecognizer_create()
face_recognizer.read('trainingData.xml')#Load saved training data
##Diccionario de Nombres
name = {0 : "Priyanka", 1 : "Cesar_Delgado" , 2 : "Hassler_Guevara" , 3 : "Kevin_Zapata", 4 : "Renzo_Alvarado", 5
: "Stephanie_Guevara", 6:"Other"}#creating dictionary containing names for each label
##Captura de Video
cap=cv2.VideoCapture(0)
##Bucle Principal
Mientras True:
   ret,test_img=cap.read()# captures frame and returns boolean value and captured image
   faces_detected,gray_img=fr.faceDetection(test_img)
   Para (x,y,w,h) en faces_detected:
     cv2.rectangle(test_img,(x,y),(x+w,y+h),(255,0,0),thickness=7)
   resized_img = cv2.resize(test_img, (1000, 700))
   cv2.imshow('Face Detection',resized_img)
   cv2.waitKey(10)
   FinPara
   ##Reconocimiento Facial
   Para face en faces_detected:
      (x,y,w,h)=face
      roi_gray=gray_img[y:y+w, x:x+h]
      label,confidence=face_recognizer.predict(roi_gray) #predicting the label of given image
      Imprimir("confidence:",confidence)
      Imprimir("label:",label)
      fr.draw_rect(test_img,face)
      predicted_name=name[label]
      if confidence < 39:#Si la confianza es inferior a 37 no se imprime en pantalla el texto de la cara prevista
        fr.put_text(test_img,predicted_name,x,y)
   resized_img = cv2.resize(test_img, (1000, 700))
   cv2.imshow('Face Detection',resized_img)
   if cv2.waitKey(10) == ord('q'):#wait until 'q' key is pressed
      break
##Finalización
cap.release()
cv2.destroyAllWindows
FinAlgoritmo
```

Algoritmo II. Pseudocódigo del algoritmo Face Recognition

Figura 9

Ejemplo de reconocimiento de una persona del data set propio



4.6 Entorno de pruebas

En el presente trabajo de investigación se tomó en cuenta los siguientes criterios de medición: la cantidad de iluminación que acepta el sistema, la distancia máxima de la cámara con la persona para su reconocimiento, pruebas con elementos de oclusión y diferentes atuendos.

A. Condiciones de Iluminación

La obtención de resultados apropiados depende en gran medida de poseer una conveniente fuente de iluminación. Los resultados pueden verse afectados de forma relevante con efectos no deseados en imágenes que posean bajo brillo, sombras, contraste y detalles falsos de oclusión en el rostro. Al utilizar un sistema de luces adecuado se logra conseguir más información de calidad respecto a la detección y extracción, esto trae como consecuencia que los algoritmos de corrección en el proceso de visión no sean necesarios. Comúnmente se pretende que las condiciones de iluminación se mantengan invariantes gran cantidad de tiempo, pero cuando es aplicado en entornos no estructurados complica la aplicación de procesos y técnicas de visión (Xu et al., 2018).

Los sistemas de reconocimiento facial en diferentes aplicaciones buscan obtener el mejor rendimiento con precisión superior al 50% pero en un entorno no controlado se ve afectado por alteración de la iluminación, expresión, postura del rostro, etc. Un dilema desafiante es la variación de la iluminación que provoca alteraciones en la forma del rostro (Zhao et al., 2003). De modo particular, la distribución de la energía de la iluminación ambiental y la dirección variable unido con la estructura 3D del rostro origina considerables diferencias en las sombras del rostro y las sombras, aunque esto varía acorde a la identidad personal (Moses et al., 1994). Según Universidad Veracruzana (2010), la norma NOM-025-STPS-2008 establece un análisis lumínico con certificación para evaluar la cantidad de iluminación adecuada respecto a cada ambiente ya que al tener un bajo nivel de iluminación afecta en gran medida en la capacidad sensorial disminuyendo la productividad de la persona. Así mismo, establece niveles mínimos de iluminación por área de trabajo y en relación a la tarea visual del puesto de trabajo.

Además, en aplicaciones de exteriores se debe tomar en consideración la intensidad y dirección de la luz del sol en un entorno no controlado donde la recomendación acerca del reconocimiento facial de un rostro está dentro del rango de [300-500] lux y la fidelidad del color va acorde a la fuente de luz en uso debido a que solo se mantiene constante si se aplica una función automática de balance de blancos (Axis Communication, 2020).

B. Resolución requerida

Actualmente, las cámaras de video poseen gran variedad de resoluciones lo cual se debe elegir la cámara más óptima acorde a la situación. Los sistemas de videovigilancia fueron creados para vigilar el comportamiento humano, por este motivo en la Tabla 1 se presentan distintas situaciones donde está dividido por subcategorías relacionado con la altura del cuerpo humano en el campo de visión.

Tabla 1

Transición a la tecnología IP

Categoría	Requisito Operativo	Altura del cuerpo
Identificación	Posibilidad de determinar con un nivel alto de certeza si una persona es la misma que otra vista anteriormente o no.	100%
Reconocimiento	Posibilidad de determinar con un nivel alto de certeza si una persona es la misma que otra vista anteriormente o no.	50%

Detección	Posibilidad de determinar con bastante certeza si hay una persona o no.	50%

Nota. Adaptada de "*CCTV Operational Requirements Manual", por* Cohen et al., 2009 (http://www.focus-security.com/_includes/docs/pdf/cctv/cctv-fit-for-purpose.pdf)

En un entorno real se debe tomar en cuenta factores como la dirección de la luz, la dinámica e intensidad. Por ello, los requisitos operativos son determinados con un punto referencial a 90° grados del eje de la cámara, siendo una línea imaginaria dentro del campo de visión. La Tabla 2 muestra los requisitos operativos establecidos para un rostro común y cálculos de densidad de píxel por centímetro de la gran mayoría de los sistemas CCTV elaborados para controlar comportamientos humanos (International, A., 2020). Además, el cálculo en densidad del píxel de la imagen permite la representación del área de interés que en este caso es el rostro humano. Por otro lado, la Tabla 3 presenta ejemplos de distancias máximas para poder identificar a un sujeto donde la captura de la resolución de una escena es afectada por el tamaño de la escena y la resolución de la cámara (International, A., 2020).

Tabla 2

Requisitos operativos

Requisito operativo	Píxeles horizontales/cara	Píxeles/cm
Identificación (condiciones difíciles)	80 px/cara	5 px/cm
Identificación (buenas condiciones)	40 px/cara	2,5 px/cm
Reconocimiento	20 px/cara	1,25 px/cm
Detección	4 px/cara	0,25 px/cm

Nota. Adaptada de Airfal Soluciones con pasión para mejorar el mundo, *por* A. International, 2020 (https://www.airfal.com/sin-categorizar/calculadora-de-lumenes-lumen-a-lux-a-candela-15835/)

Tabla 3

Ejemplos de distancias máximas para identificación (80px/cara)

Resolución horizontal de la cámara	Longitud focal	Distancia máxima	Anchura de la escena máxima
2592 píxeles	2,8-8 mm	9 m	5.2 m
1280 píxeles	3,3-12 mm	6 m	2.6 m
1920 píxeles	5,1-51 mm	41 m	3.8 m
736 píxeles	3,3-119 mm	50 m	1.5 m

Nota. Adaptada de Airfal Soluciones con pasión para mejorar el mundo, *por* A. International, 2020 (https://www.airfal.com/sin-categorizar/calculadora-de-lumenes-lumen-a-lux-a-candela-15835/)

5. EXPERIMENTACIÓN

El sistema en tiempo real diseñado fue elaborado utilizando Anaconda, una suite de código abierto con Python que involucra un conjunto de librerías y aplicaciones y se realizó bajo un procesador Intel i7-8750H de 2.20GHz, 16 GB de RAM en Windows. Así mismo, para las pruebas se usó una Cámara Web Logitech modelo Brio Ultra HD Pro, resolución máxima de 1920 x 1080 píxeles a 30 o 60 fps, campo visual diagonal ajustable (dFOV) de 65°, 78° o 90° y un Zoom digital de 5 ampliaciones (Full HD) (https://www.logitech.com/). Para la etapa de entrenamiento, posterior a

la fase de captura de la imagen, se procedió al preprocesamiento. En esta etapa, se realizó 2 pasos, que fueron: data augmentation y el tamaño uniforme de las imágenes (640x480 pixels) donde se usó la herramienta Roboflow realizando la homologación de la experimentación con el trabajo de investigación de Pei et al. (2019). Los conjuntos de datos se dividieron de la siguiente forma: 80% de las imágenes por sujeto para entrenamiento, 20% para validación.

Por otro lado, las pruebas en tiempo real se llevaron a cabo con las siguientes configuraciones: la cámara se ubicó a una altura de 2.5 metros, considerando que estará destinada al control de acceso en un complejo residencial. La cámara tiene una resolución en alta definición (HD) de 1920x1080 píxeles, lo que asegura imágenes claras. Las pruebas se realizaron en dos momentos diferentes del día, a las 10 a.m. con luz diurna (Hora de inicio) y a las 6 p.m. con luz nocturna (Hora fin), utilizando la aplicación móvil Lightmeter, un luxómetro, para medir el nivel de iluminación en el entorno de prueba, que no estaba controlado.

El sistema implementado, fue puesto a prueba en un entorno real donde aspectos tanto la variabilidad del entorno como de iluminación juegan un rol fundamental en la detección del rostro del sujeto. Si la iluminación es diurna y uniforme (luz solar) el fondo de la imagen de captura no afectará en el resultado, pero con iluminación nocturna es totalmente distinto.

En la Figura 10 se ilustra el entorno operativo del sistema implementado, con la cámara situada a 1.9 metros sobre el suelo. El ángulo de inclinación respecto a la horizontal es de 30°, y la distancia para el reconocimiento de las personas es de 2.5 metros. Este ángulo de 30° fue determinado experimentalmente, ya que se considera el valor óptimo para posicionar la cámara y capturar el rostro de las personas a una distancia de 2.5 metros.

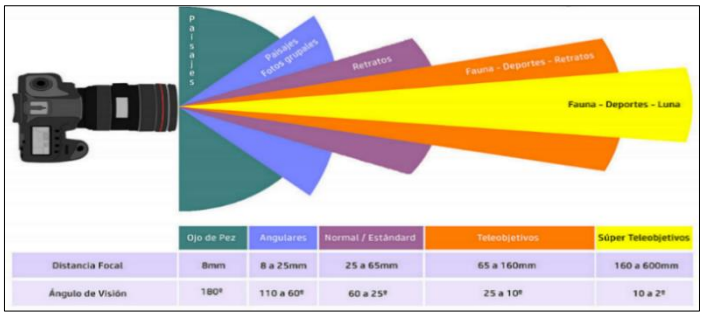
Figura 10

Escenario de operación del sistema implementado

Además, para las pruebas se estableció un campo visual de 90° grados, con el propósito de tener una distancia focal más corta entre el centro del objetivo y el sensor de imagen, de esta manera se tendrá un campo de visión más amplio, tal como se muestra en la Figura 11 (https://www.dzoom.org.es/). Asimismo, en la Tabla 4 se muestra que la luz solar durante el día proporciona una iluminación significativamente más intensa en comparación con la luz artificial durante la noche a la misma distancia.

Figura 11

Ángulo de Visión vs Distancia Focal



Nota. De Fotografía digital para todos, por dZoom, 2024 (https://www.dzoom.org.es/).

Tabla 4

Medidas de iluminación en LUX del ambiente de pruebas para luz diurna y nocturna

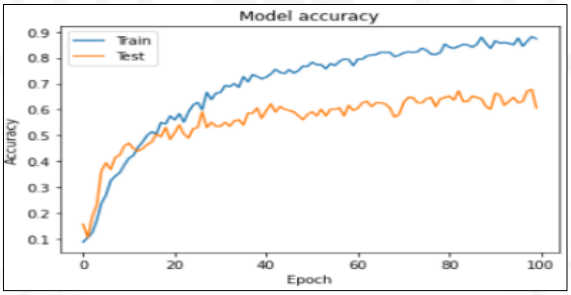
Tipo de Iluminación	Distancia diagonal del objeto a la cámara	Lux
Diurna 10:00 AM (Luz solar)	2.5 metros	350
Nocturna 18:00 PM (Luz artificial)	2.5 metros	10

6. RESULTADOS

Para las pruebas iniciales realizadas con la base de datos PubFig83, se utilizó 49 clases, 70 imágenes por persona, de las cuales 2450 imágenes fueron de entrenamiento y 980 de prueba, logrando una precisión de 91.06%.

Figura 12

Modelo de precisión dataset PubFig83



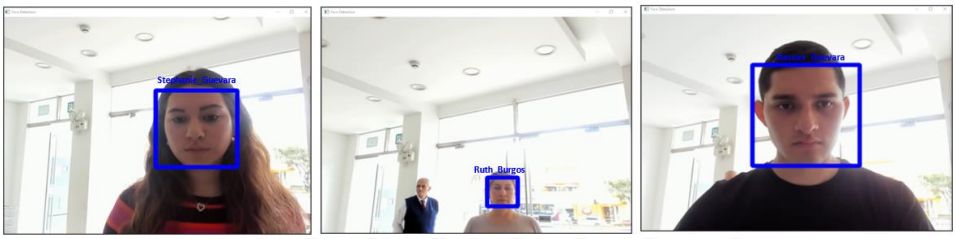
Respecto a los resultados obtenidos para el sistema en tiempo real desde una fuente en video fueron de 87.18% de precisión para 5 personas, expuestas a la cámara 10 veces cada una de forma alternada, con rostro descubierto, iluminación diurna realizada la prueba a horas de 10 am del día y 350 lux de iluminación. En caso la persona no sea identificada se envía un mensaje de alerta que la persona no está registrada en la base de datos. En la figura 13, se muestra el reconocimiento de 3 sujetos de la base datos propia en tiempo real, para diferentes distancias de la cámara en el rango de 1 metro hasta 2.5 metros. Es preciso indicar que en el Dataset propio de 5 personas con 75 imágenes de cada uno, 60 fueron para entrenamiento (80%) y 15 (20%) fueron para validación

No obstante, la prueba realizada para el mismo conjunto de prueba en horas de la noche 6 pm con cantidad de iluminación de 10 lux (iluminación artificial) dio un resultado de 45.35% de precisión y trajo consigo falsos positivos y falsos negativos al momento de identificar a la persona. Asimismo, el sistema mantiene la tendencia cuando se usan elementos de oclusión como lentes,

audífonos y gorro durante el día, con un 78% de exactitud y disminuye a 33% por la noche. En el caso que tenga una mascarilla puesta, el sistema no reconoce al sujeto sin importar las condiciones de iluminación (día o noche).

Figura 13

Prueba en tiempo real



7. DISCUSIÓN

En esta sección se detallarán los resultados obtenidos de las pruebas realizadas para diferentes condiciones de iluminación, distancia máxima de la cámara y el nivel de precisión del sistema propuesto mediante las métricas generadas a partir de la matriz de confusión. Se va a precisar acerca del proceso de evaluación de los resultados de cada experimento realizado tanto con la base de datos PubFig83 y la base datos en tiempo real.

La Figura 12 muestra que el modelo usando la base de datos PubFig83 de imágenes fijas requiere un poco más de entrenamiento puesto que la tendencia de precisión tanto para el entrenamiento y test sigue incrementando en las últimas épocas. Es importante resaltar que la aplicación de Data Augmentation y transfer learning mejora la habilidad del modelo para generalizar frente a datos nuevos, disminuyendo el riesgo de sobreajuste, incluso cuando se dispone de un conjunto de datos limitado.

A. Matriz de Confusión

Para realizar la evaluación de los resultados de las pruebas identificadas se utilizó la matriz de confusión, y posteriormente obtener las métricas de exactitud, sensibilidad, especificidad y precisión. Al mismo tiempo también se realizó la implementación de la identificación de falsas clasificaciones probado con las 49 clases, la Figura 14 muestra la clase con nombre Adam Sandler donde en la primera fila se observa los top n falsos negativos, en la segunda fila los top n falsos positivos y en la tercera fila los top n verdaderos positivos, tal como se muestra en Figura 14. Esto sirvió como primer peldaño para tomar en cuenta la existencia de falsos positivos y negativos en el sistema.

Figura 14

Ejemplo de falsas clasificaciones database PubFig83

Así mismo, para PubFig83, al realizar las pruebas correspondientes se generaron gráficos que permitieron comparar los resultados de entrenamiento y los de validación. En el primer caso, se realizó la primera prueba con 10 épocas, cantidad de veces que el algoritmo de aprendizaje interactuará de inicio a fin en el dataset, y se obtuvo lo mostrado en la figura 15, con eje x el número de épocas y eje y valor de pérdida. Se observa que los valores de pérdida, tanto para entrenamiento y validación , al inicio fueron elevados y uniformes, pero al entrenarlos fueron disminuyendo su valor mientras la figura 16, muestra que los valores de precisión fueron incrementando su valor de forma uniforme a lo largo de las épocas.

En el segundo caso tal como se muestra en figura 17 al incrementar el número de épocas a 100 , muestra como han ido variando los valores de pérdida tanto en el entrenamiento como en la validación donde los valores de la validación han sido más abruptos respecto al entrenamiento. Por otro lado, la figura 18 muestra que el modelo muestra un buen aprendizaje con los datos de entrenamiento, pero se observa una diferencia significativa en la precisión al aplicarlo a los datos de validación, lo que sugiere un posible sobreajuste.

Figura 15

Pérdida de entrenamiento vs pérdida de validación con 10 épocas

Figura 16

Precisión de entrenamiento vs precisión de validación con 10 épocas

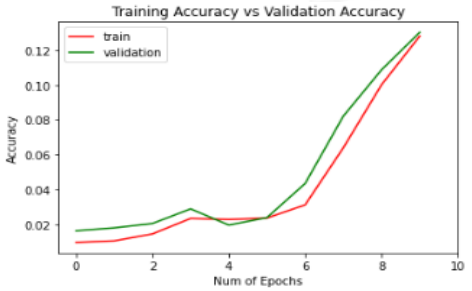
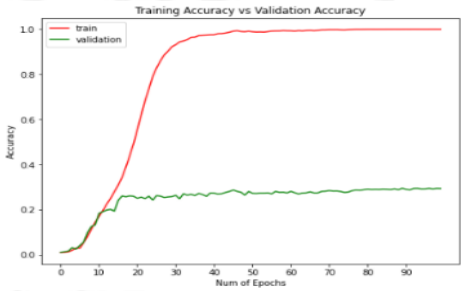


Figura 17

Pérdida de entrenamiento vs pérdida de validación con 100 épocas

Figura 18

Precisión de entrenamiento vs precisión de validación con 100 épocas



La Tabla 5 muestra los valores alcanzados para la Matriz de Confusión de la BD PubFig83, lo que nos permite corroborar los resultados obtenidos respecto a la calidad del modelo.

Tabla 5

Matriz de confusión Pubfig83

	Predicción	
Observación	Positivos	Negativos
Positivos	662	65
Negativos	59	223

Con los resultados obtenidos previamente del número total de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Es fundamental resaltar que el Eje X (Predicción) representa lo que el sistema predice (si reconoce a la persona como conocida o no) mientras el Eje Y (Observación) representa lo que realmente es (la verdad de la situación, conocido o desconocido). Asimismo, la matriz de confusión permite observar y comprender el rendimiento de un sistema de reconocimiento facial, mostrando con claridad dónde y cómo se producen los errores.

De esta forma, se procedió a realizar las métricas correspondientes para evaluar la calidad del modelo:

1. Métrica 1: Nivel de exactitud (Accuracy)

$$Exactitud = \frac{\text{VP} + \text{VN}}{\text{Total}} = \frac{662 + 223}{(662 + 65 + 59 + 223)} * 100\% = 87.71\%$$

Ecuación 1. Métrica de exactitud

2. Métrica 2: Nivel de sensibilidad (Recall)

$$Sensibilidad = \frac{\text{VP}}{\text{Total de positivos}} = \frac{662}{662 + 59} * 100\% = 91.82\%$$

Ecuación 2. Métrica de sensibilidad

3. Métrica 3: Nivel de especificidad

$$Especificidad = \frac{\text{VN}}{\text{Total de negativos}} = \frac{223}{223 + 65} * 100\% = 77.43\%$$

Ecuación 3. Métrica de especificidad

4. Métrica 4: Nivel de precisión

$$Precision = \frac{\text{VP}}{\text{Total clasificados positivos}} = \frac{662}{662 + 65} * 100\% = 91.06\%$$

Ecuación 4. Métrica de precisión

5. Métrica 5: F1-score

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 2 * \frac{87.71\% * 91.82\%}{87.71\% + 91.82\%} = 89.72\%$$

Ecuación 5. Métrica de F1-score

La precisión alcanzada en esta investigación con respecto a la base de datos Pubfig83 es del 87.71%, lo que se interpreta como la habilidad del sistema para reconocer a una persona. La métrica 2 de sensibilidad se obtuvo 91.82% el cual nos indica que el sistema reconoce y asigna el valor de verdadero a un rostro dentro de la base datos.

La tercera métrica es un indicador con un valor de 77.43% el cual indica que el sistema tiene la capacidad de reconocer una imagen ajena a la base de datos. La métrica 4 de precisión obtuvo un valor de 91.06% teniendo en cuenta la existencia de falsos positivos. De este modo, un falso positivo ocurre cuando el sistema reconoce incorrectamente a una persona no incluida en la base de datos como si fuera alguien registrado. Por último, el indicador de F1-score tiene un resultado de 89.72% que indica que existe un equilibrio entre la precisión y el recall donde se combina ambos valores para obtener un solo valor.

En las pruebas del sistema en tiempo real, se realizaron 50 evaluaciones para cada tipo de iluminación con la base de datos de 5 sujetos en un entorno no controlado, y se detectaron también falsas clasificaciones. Como ejemplo en la figura 19 se muestran los casos de falsas clasificaciones para luz diurna, en la primera fila se observa los top n falsos negativos, en la segunda fila los top n falsos positivos y en la tercera fila los top n verdaderos positivos.

Figura 19

Ejemplo de falsas clasificaciones database propia en tiempo real

A continuación, en la Tabla 6 se muestra la matriz de confusión del sistema en tiempo real, para luz diurna.

Tabla 6

Matriz de confusión dataset tiempo real (luz diurna)

	Predicción	
Observación	Positivos	Negativos
Positivos	34	5
Negativos	3	8

1. Métrica 1: Nivel de exactitud (Accuracy)

$$Exactitud = \frac{\text{VP+VN}}{\text{Total}} = \frac{34+8}{(34+5+3+8)} * 100\% = 84.00\%$$

Ecuación 6. Métrica de exactitud

2. Métrica 2: Nivel de sensibilidad (Recall)

$$Sensibilidad = \frac{\text{VP}}{\text{Total de positivos}} = \frac{34}{34+3} * 100\% = 91.89\%$$

Ecuación 7. Métrica de sensibilidad

3. Métrica 3: Nivel de especificidad

$$Especificidad = \frac{\text{VN}}{\text{Total de negativos}} = \frac{8}{5+8} * 100\% = 61.54\%$$

Ecuación 8. Métrica de especificidad

4. Métrica 4: Nivel de precisión

$$Precision = \frac{\text{VP}}{\text{Total clasificados positivos}} = \frac{34}{34+5} * 100\% = 87.18\%$$

Ecuación 9. Métrica de precisión

5. Métrica 5: F1-score

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 2 * \frac{87.18\% * 91.89\%}{87.18\% + 91.89\%} = 89.47\%$$

Ecuación 10. Métrica de F1-score

De acuerdo con las pruebas realizadas en tiempo real con luz diurna con la base de datos propia se obtuvo un nivel de exactitud de 84.00%. Respecto al nivel de sensibilidad, siendo la métrica 2, se tuvo como resultado 91.89% donde esta métrica indica la capacidad del estimador para discernir los casos positivos. La métrica 3 obtuvo el valor de 61.54% el cual nos indica que los casos negativos se han clasificado correctamente. La métrica 4 de precisión alcanzó el valor de 87.18% el cual muestra las predicciones con valor verdadero. La última métrica F1-score consiguió el resultado de 89.47% siendo la media armónica entre la precisión y el recall.

A continuación, en la Tabla 7 se muestra la matriz de confusión del sistema en tiempo real, para luz nocturna.

Tabla 7

Matriz de confusión dataset tiempo real (luz nocturna)

	Predicción	
Observación	Positivos	Negativos
Positivos	10	15
Negativos	12	13

1. Métrica 1: Nivel de exactitud (Accuracy)

$$Exactitud = \frac{\text{VP}+\text{VN}}{\text{Total}} = \frac{10+13}{(10+12+15+13)} * 100\% = 46.00\%$$

Ecuación 11. Métrica de exactitud

2. Métrica 2: Nivel de sensibilidad (Recall)

$$Sensibilidad = \frac{\text{VP}}{\text{Total de positivos}} = \frac{10}{10+12} * 100\% = 45.45\%$$

Ecuación 12. Métrica de sensibilidad

3. Métrica 3: Nivel de especificidad

$$Especificidad = \frac{\text{VN}}{\text{Total de negativos}} = \frac{13}{13+15} * 100\% = 46.43\%$$

Ecuación 13. Métrica de especificidad

4. Métrica 4: Nivel de precisión

$$Precision = \frac{\text{VP}}{\text{Total clasificados positivos}} = \frac{10}{10+15} * 100\% = 40.00\%$$

Ecuación 14. Métrica de precisión

5. Métrica 5: F1-score

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 2 * \frac{40.00\% * 45.45\%}{40.00\% + 45.45\%} = 42.55\%$$

Ecuación 15. Métrica de F1-score

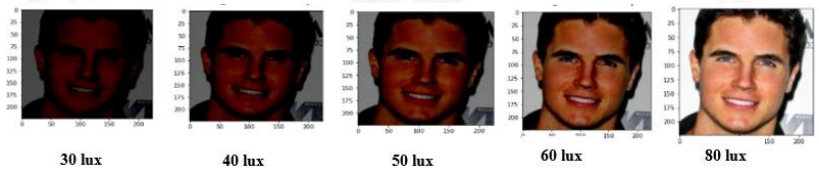
De acuerdo con las pruebas realizadas en tiempo real con luz nocturna con la base de datos propia se obtuvo un nivel de exactitud de 46.00%. Respecto al nivel de sensibilidad, siendo la métrica 2, se tuvo como resultado 45.45% donde esta métrica indica la capacidad del estimador para discernir los casos positivos. La métrica 3 obtuvo el valor de 46.43% el cual nos indica que los casos negativos se han clasificado correctamente. La métrica 4 de precisión alcanzó el valor de 40.00% el cual muestra las predicciones con valor verdadero. La última métrica F1-score consiguió el resultado de 42.55% siendo la media armónica entre la precisión y el recall.

B. Condiciones de Iluminación

En la primera fase de la investigación, el sistema propuesto utilizó la base de datos Pubfig83, con imágenes en ambientes cerrados, luego utilizando la herramienta FastStone Image Viewer, se modificó las condiciones de iluminación de las imágenes, simulando diferentes entornos.

Figura 20

Variación de cantidad de iluminación de imágenes del dataset Pubfig83 (lux)



El porcentaje de acierto del modelo para cada ambiente se muestra en la Tabla 8, los resultados fueron favorables para el rango entre [60,100] lux pero con nulo o bajo nivel de acierto para condiciones de iluminación menores a 60 lux. En un entorno controlado, siempre que se mantenga dentro del rango adecuado, la iluminación sigue siendo un factor clave para el rendimiento y la precisión de los sistemas de reconocimiento facial. Contar con una iluminación apropiada es esencial para garantizar que el sistema pueda identificar y distinguir a las personas de manera precisa.

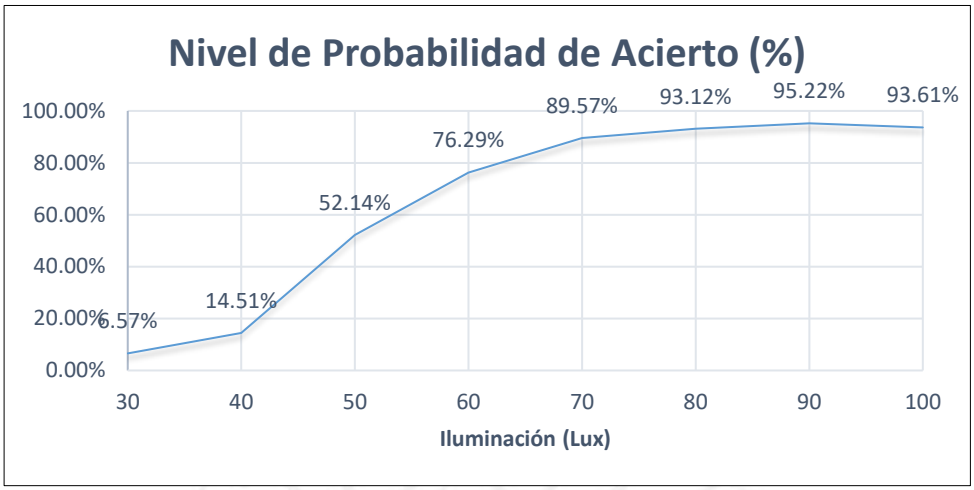
Tabla 8

Promedio de acierto del sistema por nivel de iluminación para un ambiente controlado con iluminación artificial

Nivel de iluminación	Probabilidad de acierto(%)
30 lux	6.57%
40 lux	14.51%
50 lux	52.14%
60 lux	76.29%
70 lux	89.57%
80 lux	93.12%
90 lux	95.22%
100 lux	93.61%

Figura 21

Probabilidad de acierto vs Iluminación (Lux)



En base a la experimentación realizada se puede afirmar que el nivel de iluminación óptimo se encuentra entre el rango de [50;90] luxes donde se han obtenido valores de precisión crecientes, lo que indica que se puede reconocer de manera adecuada la imagen de entrada sin dificultad pero en el caso del rango [0;40] luxes la imagen sufre variaciones de iluminación que dificultan al modelo identificar correctamente a la persona. Sin embargo, tal como se observa en la figura 21, cuando la cantidad de iluminación es excesiva (mayor a 90 lux), la precisión del modelo nuevamente decae.

Es preciso indicar que el uso de una aplicación en un dispositivo móvil para medir la cantidad de lúmenes es una solución pragmática y adecuada para el contexto de esta investigación, cuya objetivo no es obtener mediciones lumínicas de precisión absoluta sino evaluar el rendimiento del sistema de reconocimiento facial en entornos residenciales. La aplicación proporciona mediciones suficientemente aproximadas para determinar si las condiciones de iluminación son funcionales para el sistema en vez de requerir la exactitud científica de un luxómetro profesional. También, se garantiza la consistencia de los resultados al utilizar el mismo dispositivo y la misma aplicación para todas las mediciones, lo cual es esencial para realizar comparaciones válidas entre diferentes niveles de iluminación. Los resultados de la investigación demostraron que el rendimiento del sistema cambió de manera predecible bajo diferentes niveles de luz, lo que confirma que la aplicación fue adecuada para detectar cambios significativos en las condiciones lumínicas. A pesar de que un luxómetro puede mejorar la precisión, los resultados obtenidos son representativos y cumplen con los objetivos del estudio siendo suficientes para evaluar la efectividad del sistema en condiciones realistas. Para futuras investigaciones, se podría considerar el uso de un luxómetro en entornos más críticos, pero las mediciones realizadas en este caso fueron válidas y consistentes.

C. Distancia Máxima

Experimentalmente se pudo establecer la distancia máxima que la persona puede estar frente a la cámara para que pueda ser reconocido con éxito ya que la cantidad de iluminación que recibe el objeto está relacionada con la distancia de la fuente, el ángulo de visión y la distancia focal. En el entorno real de operación del sistema el rango de cantidad de iluminación está dentro del rango de [500 lux para luz diurna, 300 lux para luz nocturna] medido con la App Lighmeter. Por este motivo, el sistema es altamente sensible a cambios de iluminación. Así se determinó que la distancia mínima y máxima de la cámara hacia la persona para que ésta pueda ser identificada correctamente está en el rango [1.5mt, 2.5mt] tal como se muestra en la Tabla 9. Acá se muestra que para una distancia de 4 metros la cámara tiene una precisión de 60%, que disminuye a medida que aumenta la distancia. Es crucial destacar que se llevaron a cabo 10 pruebas para cada persona, resultando en un total de 50 pruebas.

Tabla 9

Distancia máxima de la cámara

Distancia/Personas	Persona 1	Persona 2	Persona 3	Persona 4	Persona 5	Precisión
1 m	✔	✔	✘	✔	✔	80%
2 m	✔	✔	✔	✔	✔	100%
3 m	✔	✔	✔	✘	✔	80%
4 m	✔	✔	✔	✘	✘	60%
5 m	✘	✘	✘	✔	✔	40%
6 m	✘	✘	✘	✘	✘	0%

Leyenda: ✔ : Sí reconoce / ✘: No reconoce

D. Uso de Outfits

La calidad de la imagen de entrada y algunos elementos de oclusión ocasionan la reducción de la precisión del sistema de reconocimiento facial. Los elementos de oclusión más frecuentes son el uso de gafas, gorros, entre otros. Por ello, en la presente investigación se realizó 10 pruebas por cada persona para cada elemento de oclusión ya mencionados. Los resultados obtenidos se aprecian en la tabla 10 la cual presenta la precisión para elemento de oclusión para identificar una persona a la vez en condiciones de luz diurna.

Tabla 10

Probabilidad de acierto en la clasificación facial con oclusión parcial usando base de datos en tiempo real (%)

Personas	Lentes	Gorro
Sujeto 1_Hassler	74.79%	76.31%
Sujeto 2_Stephanie	72.80%	73.97%
Sujeto 3_Ruth	73.92%	75.47%

Como se observa en los resultados, los elementos como lentes y gorro no afectan completamente o solo reducen la probabilidad de acierto del reconocimiento. Sin embargo en casos donde el sujeto porta una mascarilla, experimentalmente se comprobó que el porcentaje de reconocimiento es nulo.

8. VALIDACIÓN

En este acápite se realizó la comprobación del sistema en tiempo real reentrenando, el cual se aplicó transfer-learning, con conjuntos de datos adicionales para probar si el rendimiento del mismo se mantiene a medida que aumenta el número de personas. El sistema en tiempo real fue puesto a prueba añadiéndole los conjuntos de datos: FEI Dataset (10 sujetos) y Georgia Dataset (15 sujetos), dando un total de 30 sujetos en total, incluyendo los 5 sujetos registrados inicialmente en la base de datos.

La Tabla 11 muestra los resultados obtenidos respecto a las pruebas realizadas donde en la prueba 1 se utilizó solo el dataset propio de 5 personas con 75 imágenes cada uno, 15 fueron para prueba y 60 para entrenamiento obteniendo como resultado promedio 87.18% de precisión. En la prueba 2, se añadió el FEI Dataset al primer conjunto de datos obteniendo 86.47% de resultado de precisión, en esta prueba se realizó con un total de 10 sujetos los cuales cada uno tenía de igual forma 75 imágenes por cada sujeto. Por último, la tercera prueba se realizó añadiendo el Georgia Dataset el cual contiene 15 sujetos y 75 imágenes por cada sujeto. Este

último dataset se agregó al primer y segundo conjunto de datos obteniendo en total 30 sujetos, se puso a prueba el sistema obteniendo 84.76% de precisión.

Figura 22

Modelo de validación del sistema en tiempo real (1) Dataset propio (2) Dataset propio + FEI Dataset (3) Dataset propio + FEI Dataset + Giorgia Dataset

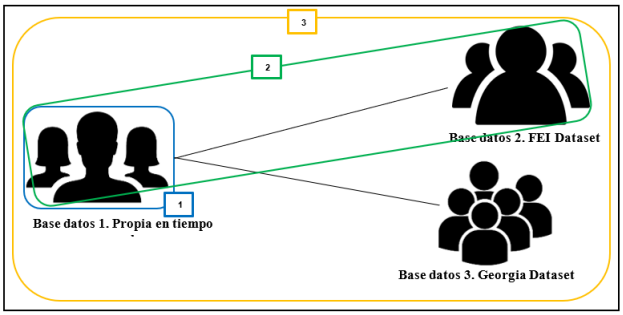


Tabla 11

Resultados obtenidos de validación del sistema en tiempo real

Dataset	Precisión
Dataset Propio	87.18%
Dataset Propio + FEI dataset	86.47%
Dataset Propio + FEI dataset + Georgia Dataset	84.76%

Figura 23

Reconocimiento de sujetos en tiempo real del dataset propio

Los resultados obtenidos muestran que como era de esperarse a medida que aumenta la cantidad de clases posibles, el sistema disminuye la precisión. Sin embargo, como se observa en el último caso habiendo aumentado la cantidad de personas de 5 a 30, el sistema solo disminuyó su precisión en un 4%. Esto se podría mejorar si de cada persona se toman más imágenes durante la etapa de entrenamiento del modelo, con lo cual se podría preservar la calidad del reconocimiento.

9. CONCLUSIONES

El sistema de reconocimiento facial basado en redes neuronales convolucionales (CNN) consiguió una precisión del 87.71% para imágenes fijas en el Dataset PubFig83 y del 84.00% en el sistema en tiempo real con iluminación diurna. Estos resultados indican que el modelo tiene un rendimiento adecuado en entornos controlados, aunque la precisión disminuye en condiciones más desafiantes, como la iluminación nocturna donde la precisión cae al 46.00%.

La capacidad del sistema se vio afectada por la iluminación y la distancia de la cámara. El sistema obtuvo una probabilidad de acierto del 93.61% en condiciones de buena iluminación

(500 lux). Sin embargo, la precisión disminuyó al 45.35% bajo luz artificial 10 lux (nocturna). Además, la distancia máxima efectiva fue de 2.5 metros, lo que fue ideal para la detección y reconocimiento sin comprometer la calidad de la imagen.

La precisión del sistema se mantuvo superior al 72% en pruebas con oclusiones parciales tales como el uso de lentes y gorros evidenciando cierta robustez ante estas variaciones. No obstante, el uso de mascarillas disminuyó la capacidad del sistema para reconocer correctamente a las personas, lo que representa un desafío para futuros avances.

Para mejorar la precisión del sistema y evitar el sobreajuste, especialmente con un conjunto de datos limitados, se eligió la arquitectura VGG-Face y se emplearon técnicas como el aumento de datos. La detección, normalización y extracción de características, el flujo de trabajo implementado, permitieron optimizar el reconocimiento facial en tiempo real y adaptarse a los requisitos del entorno residencial.

La capacidad de extraer características profundas y discriminativas, adecuadas para la identificación de rostros en imágenes y videos en tiempo real, motivó la elección de redes neuronales convolucionales (CNN) y, en particular, la arquitectura VGG-Face. Mientras que VGG-Face, previamente entrenado en grandes bases de datos, reduce el tiempo y los costos computacionales del entrenamiento y es ideal para uso en entornos residenciales con recursos limitados, CNN permite detectar patrones locales en las imágenes, lo cual es crucial para el reconocimiento facial.

Las etapas de detección, normalización, extracción de características y reconocimiento en tiempo real fueron adaptadas para funcionar en un entorno no controlado en el flujo de trabajo diseñado. La variabilidad en las condiciones de iluminación y la presencia de elementos de oclusión obligaron al sistema a utilizar técnicas como el aumento de datos para mejorar la robustez del modelo y mejorar su capacidad de generalización con datos nuevos.

El sistema tenía limitaciones significativas en condiciones de iluminación nocturna y oclusión completa (como el uso de mascarillas), lo que redujo significativamente su precisión. Estas restricciones resaltan la importancia de implementar mejoras futuras, como el uso de redes Transformers que puedan gestionar grandes volúmenes de datos y mejorar la precisión bajo diferentes condiciones de luz y oclusión, así como la exploración de técnicas avanzadas de reconocimiento facial con mascarilla, que se han vuelto más relevantes en contextos recientes.

10. TRABAJOS FUTUROS

Como trabajo futuro se proyecta incluir el reconocimiento facial con oclusión por mascarilla ya que la mascarilla representa la oclusión del 50% del rostro. Así mismo se propone la implementación del mismo sistema con un conjunto de datos mayor utilizando Redes Transformers puesto que esta arquitectura está diseñada para trabajar con grandes cantidades de datos. De este modo se podrá comparar el resultado obtenido en lo referente a exactitud y precisión con la técnica de Redes Neuronales Convolucionales (CNN) utilizada en este trabajo.

REFERENCIAS

Axis Communication. (2020). Identificación y reconocimiento. Obtenido de https://www.axis.com/es-pe/learning/web-articles/identification-and-recognition/resolution

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. Proceedings - 13th IEEE International Conference

on Automatic Face and Gesture Recognition, FG 2018, 67-74. https://doi.org/10.1109/FG.2018.00020

Castaño, D., & Alonso, J. (2019). Sistema de reconocimiento facial para control de acceso a viviendas.

Chaturvedi, K., & Vishwakarma, D. K. (2020). Face Recognition in an Unconstrained Environment using ConvNet. ACM International Conference Proceeding Series, 67–71. https://doi.org/10.1145/3378904.3378905

Christanti, V., Susanto, N., & Santun Naga, D. (2018). Spelling Correction for Text Documents in Bahasa Indonesia Using Finite State Automata and Levinshtein Distance Method. MATEC Web of Conferences, 164, 1–15. https://doi.org/10.1051/matecconf/201816401047

Cohen, N., Gattuso, J., & MacLennan-Brown, K. (2009). CCTV Operational Requirements Manual. En Home Office Scientific … (Número 28). http://www.focus-security.com/_includes/docs/pdf/cctv/cctv-fit-for-purpose.pdf

Cook, C. M., Howard, J. J., Sirotin, Y. B., Tipton, J. L., & Vemury, A. R. (2019). Demographic Effects in Facial Recognition and Their Dependence on Image Acquisition: An Evaluation of Eleven Commercial Systems. IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(1), 32-41. https://doi.org/10.1109/tbiom.2019.2897801

Coskun, M., Ucar, A., Yildirim, O., & Demir, Y. (2017). Face recognition based on convolutional neural network. Proceedings of the International Conference on Modern Electrical and Energy Systems, MEES 2017, 2018-Janua(November), 376-379. https://doi.org/10.1109/MEES.2017.8248937

El Khiyari, H., & Wechsler, H. (2016). Face Recognition across Time Lapse Using Convolutional Neural Networks. Journal of Information Security, 07(03), 141-151. https://doi.org/10.4236/jis.2016.73010

Fuentes, H. A. (2011). Sistemas de reconocimiento basados en la imagen facial. 8, 7–15.

Gao, Y., & Mosalam, K. M. (2018). Deep Transfer Learning for Image-Based Structural Damage Recognition. Computer-Aided Civil and Infrastructure Engineering, 33(9), 748-768. https://doi.org/10.1111/mice.12363

García, M. (2009). Diseño e implementación de una herramienta de detección facial. [Fecha de consulta: 25 de mayo 2020]. Disponible en: http://tesis.ipn.mx/bitstream/handle/123456789/6111/DISENOIMPLEMFACIAL.pdf

Grágeda, N. (2023). Reconocimiento de emociones utilizando la voz en ambiente dinámicos de interacción Humano-Robot. https://repositorio.uchile.cl/handle/2250/194750

Grother, P., Ngan, M., & Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) Part 3 : Demographic Effects. Nistir 8280, December, https://doi.org/10.6028/NIST.IR.8280.

Ho Chan, C., Zou, X., Poh, N., & Kittler, J. (2018). Illumination invariant face recognition: A survey. Computer Vision: Concepts, Methodologies, Tools, and Applications, 58-79. https://doi.org/10.4018/978-1-5225-5204-8.ch003

International, A. (2020). Airfal Soluciones con pasión para mejorar el mundo. Obtenido de https://www.airfal.com/sin-categorizar/calculadora-de-lumenes-lumen-a-lux-a-candela-15835/

Jeon, J., Park, J. C., Jo, Y. J., Nam, C. M., Bae, K. H., Hwang, Y., & Kim, D. S. (2016). A real-time facial expression recognizer using deep neural network. ACM IMCOM 2016: Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication. https://doi.org/10.1145/2857546.2857642

Karahan, Ş., Yildirim, M. K., Kirtaç, K., Rende, F. Ş., Bütün, G., & Ekenel, H. K. (2016). How image degradations affect deep CNN-based Face recognition? Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI), P-260. https://doi.org/10.1109/BIOSIG.2016.7736924

Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009). Attribute and simile classifiers for face verification. Proceedings of the IEEE International Conference on Computer Vision, 365–372. https://doi.org/10.1109/ICCV.2009.5459250

Lan, J., Xie, Y., Wang, H., & Liu, G. (2019). A face recognition system based on improved convolutional neural network. ACM International Conference Proceeding Series, 230-235. https://doi.org/10.1145/3377713.3377765

Lu, X., Lu, X., & Lansing, E. (2003). Image analysis for face recognition. Personal Notes, May, 5, 1-37. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.4026&rep=rep1&type=pdf

Lumini, A., Nanni, L., & Brahnam, S. (2017). Ensemble of texture descriptors and classifiers for face recognition. Applied Computing and Informatics, 13(1), 79-91. https://doi.org/10.1016/j.aci.2016.04.001

Masi, I., Tuấn, A., Hassner, T., Sahin, G., & Medioni, G. (2019). Face-Specific Data Augmentation for Unconstrained Face Recognition. International Journal of Computer Vision. https://doi.org/10.1007/s11263-019-01178-0

Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks, 16(5-6), 555-559. https://doi.org/10.1016/S0893-6080(03)00115-1

Mehdipour, M., & Kemal, H. (2016). A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 102–109. https://doi.org/10.1109/CVPRW.2016.20

Moses, Y., Adini, Y., & Ullman, S. (1994). Face recognition: The problem of compensating for changes in illumination direction. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 800 LNCS(7), 286-296. https://doi.org/10.1007/3-540-57956-7_33

Nefian, A. V., & Hayes, M. H. (2000). Maximum Likelihood Training of the Embedded HMM for face detection and recognition.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. Section 3, 41.1-41.12. https://doi.org/10.5244/c.29.41

Pei, Z., Xu, H., Zhang, Y., Guo, M., & Yee-Hong, Y. (2019). Face recognition via deep learning using data augmentation based on orthogonal experiments. Electronics (Switzerland), 8(10), 1-16. https://doi.org/10.3390/electronics8101088

Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model. 1. http://arxiv.org/abs/1709.01664

Ríos, J. E. (2013). Reconocimiento facial en tiempo real.

Shlens, J. (2003). A Tutorial on Principal Component Analysis Derivation, Discussion and Singular Value Decomposition.

Simancas, B. (2019). Desarrollo de un sistema de Identificación mediante técnicas de reconocimiento facial.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. 1-14. http://arxiv.org/abs/1409.1556

Sims, D. (1994). Biometric Recognition: Our Hands, Eyes, and Faces Give Us Away. IEEE Computer Graphics and Applications, 14(5), 14-15. https://doi.org/10.1109/38.310718

Tenorio, E. Z., & Thomaz, C. E. (2011). Análise Multilinear Discriminante de Formas Frontais de Imagens 2D de Face. X Sbai - Simpósio Brasileiro de Automação Inteligente, X, 1043–1048.

Universidad Veracruzana . (2010). Evaluación de los niveles de iluminación en las áreas de trabajo de LATEX. Obtenido de https://www.uv.mx/cosustenta/files/2014/06/ANALISIS-LATEX_VERSION-FINAL.pdf

Vegas, J., & Merayo, N. (2023). Análisis emocional en redes sociales basados en modelos de aprendizaje automático transformers BERT. https://uvadoc.uva.es/handle/10324/62911

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1(February). https://doi.org/10.1109/cvpr.2001.990517

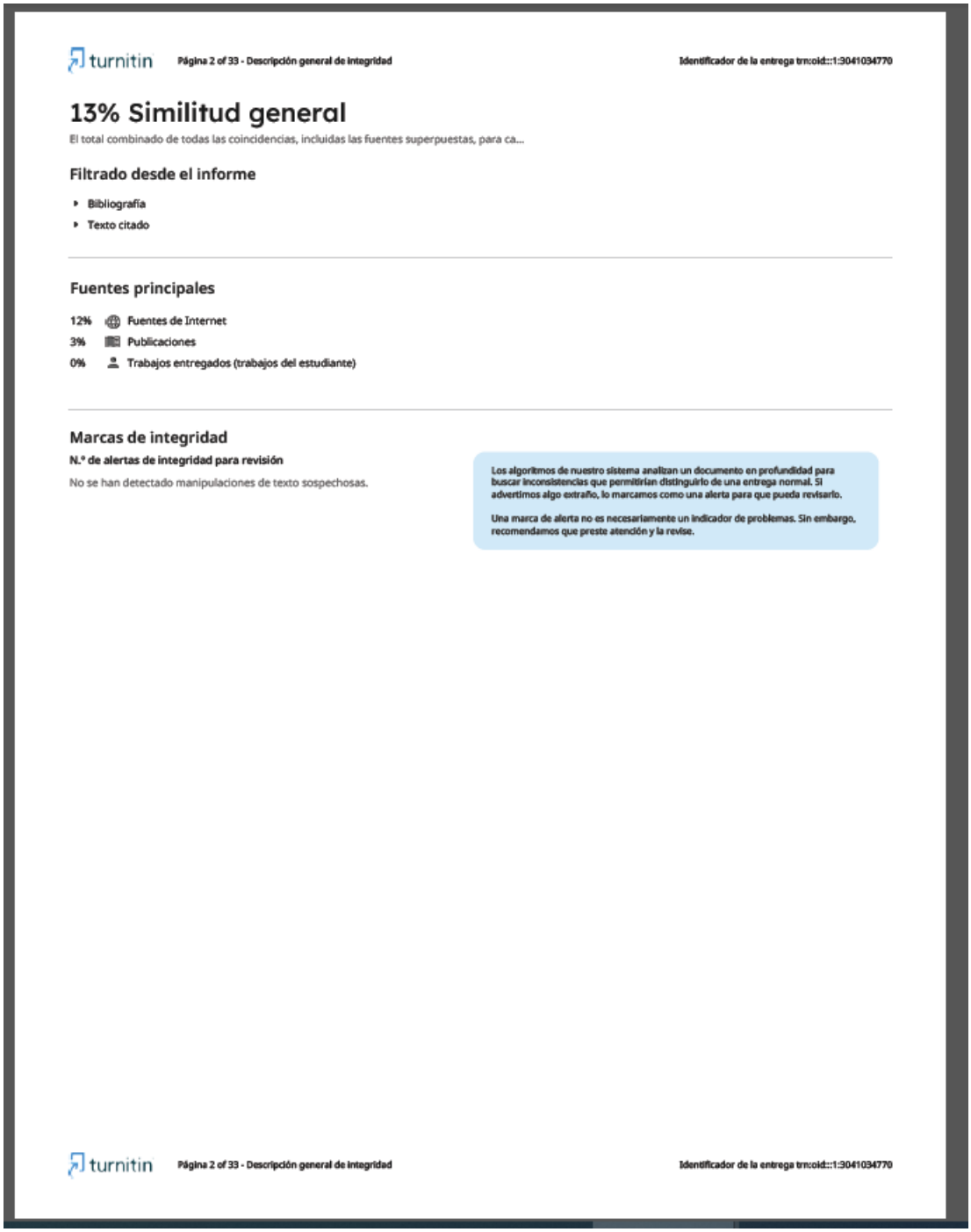
Vu, T. H., Osokin, A., & Laptev, I. (2015). Context-aware CNNs for person head detection. Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, 2893-2901. https://doi.org/10.1109/ICCV.2015.331

Xia, X.-L., Xu, C., & Nan, B. (2017). Facial Expression Recognition Based on TensorFlow Platform. ITM Web of Conferences, 12, 01005. https://doi.org/10.1051/itmconf/20171201005

Xu, L., Xiang, J., & Yuan, X. (2018). Transferring Rich Deep Features for Facial Beauty Prediction. http://arxiv.org/abs/1803.07253

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider Face: A face detection benchmark. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem, 5525–5533. https://doi.org/10.1109/CVPR.2016.596

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face Recognition: A Literature Survey. IEEE Transactions on Image Processing, 35(4), 399–458. https://doi.org/10.1109/TIP.2009.2021737



turnitin Página 2 of 33 - Descripción general de integridad Identificador de la entrega trn:oid:::1:3041034770

13% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado

Fuentes principales

12% Fuentes de Internet
3% Publicaciones
0% Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

turnitin Página 2 of 33 - Descripción general de integridad Identificador de la entrega trn:oid:::1:3041034770