



DETECCIÓN AUTOMÁTICA DE ROSTRO CON MASCARILLA PARA EVITAR CONTAGIO POR COVID, MEDIANTE ALGORITMO LBP EMBEBIDO EN RASP-BERRY PI Y USO DE ALERTA AUDIBLE

AUTOMATIC DETECTION OF FACE WITH MASK TO PREVENT COVID CONTAGION, USING LBP ALGORITHM EMBEDDED IN RASP-BERRY PI AND USE OF AUDIBLE ALERT

Juan Alberto Antonio Velázquez ^{1,*}, Leopoldo Gil Antonio ¹, Héctor Caballero Hernández ¹, Daniel Cervantes Ambriz ²

¹Tecnológico Nacional de México (TecNM): Tecnológico de Estudios Superiores de Jocotitlán, Carretera Toluca-Atlacomulco KM 44.8 Ejido de San Juan y, San Agustín, 50700 Cdad. de Jocotitlán, Méx.

²Tecnológico Nacional de México (TecNM): Instituto Tecnológico de Toluca, Av Tecnológico 100-s/n, Agrícola, 52149 Metepec, Méx

(recibido/received: 04-09-2024; aceptado/accepted: 12-12-2024)

RESUMEN

La detección de rostros de personas en sistemas de autoacceso, es una tarea retadora ya que la pandemia del coronavirus (SARS-CoV2) ha cambiado nuestra forma de vivir, sobre todo porque las empresas y algunas entidades gubernamentales como los hospitales necesitan otorgar permiso para acceder a sus instalaciones sobre todo si la persona no trae el cubrebocas, y aunque muchas instalaciones ya no permiten su uso, ya que las entidades antes mencionadas si justifican su uso sobre todo si en sus instalaciones hay personas con enfermedades crónicas que necesitan y justifican el uso del cubrebocas. En este trabajo, mencionamos un sistema de acceso que consta de un software embebido en un dispositivo *Rasp-berry PI*, el cual contiene una aplicación formada por el uso de una red neuronal *Multi-task cascade convolutional networks (MTCNN)*, con la cual se hace el entrenamiento de las imágenes de rostros que traen y no traen cubrebocas, posteriormente y para una mejor identificación de los rostros se utiliza un algoritmo basado en *Histograma de Patrones Binarios Locales (LBPH)*, con el cual se obtienen las características del rostro y posteriormente se puede clasificar si la persona a identificar tiene o no cubrebocas y en su caso enviar una alerta audible. Al final, nuestro sistema propuesto tiene una precisión promedio del 93% en la detección de rostros con mascarillas.

Palabras clave: *LBPH*, *MTCNN*, cubreboca, componente de detección de rostros

ABSTRACT

The detection of the faces of people in self-access systems, is a challenging task since the pandemic of the coronavirus (SARS-CoV2) has changed our way of living, especially because companies and some government entities such as hospitals need to grant permission to access their facilities especially if the

* Autor de correspondencia.
Correo: juan.antonio@tesjo.edu.mx

person does not bring the mask, and although many facilities no longer allow its use, since the above mentioned entities if they justify its use especially if their facilities there are people with chronic diseases that need and justify the use of the mask. In this work, we mention an access system that consists of software embedded in a Rasp-berry PI device, which contains an application formed by the use of a neural network Multi-task cascade convolutional networks (MTCNN), with which the training of the images of faces that bring and do not bring mouthguards is done, Later, and for a better identification of the faces, an algorithm based on Local Binary Patterns Histogram (LBPH) is used, with which the characteristics of the face are obtained and later it is possible to classify if the person to be identified has or does not have a face mask. In the end, our proposed system has an average accuracy of 93% in the detection of faces with face masks.

Keywords: LBPH, MTCNN, Facemask, face detection component.

1. INTRODUCCIÓN

La propagación del *coronavirus COVID-19* a finales del 2019 ha generado una serie de desafíos y transformaciones que han afectado todos los rincones del planeta. Uno de estos retos es el uso de cubrebocas como medida de contención de transmisión del virus de tal manera que todas las personas tengan la necesidad de que al portarlo en ambientes cerrados y donde haya concentraciones de personas se pueda evitar altos contagios. Aunque en algunas organizaciones actualmente no es obligatorio el uso de cubrebocas, en otras es imprescindible su uso sobre la consideración que hay sobre todo en épocas del año que el *coronavirus* y sus variantes prolifera. Por este motivo, es necesario que las instituciones sigan manteniendo protocolos de acceso donde se considere el uso de cubrebocas. Por esta razón es necesario aplicar mecanismos que validen la portación del cubrebocas al momento de acceder a las instalaciones de las instituciones que tienen normas de acceso. Este trabajo de investigación explica la implementación de un sistema automático para la detección de rostros que porten cubrebocas para el control de acceso de personal a cualquier institución de tenga como norma el uso de cubrebocas, además se considera como aporte un sistema de alerta de señalización visual y auditiva que funcione automáticamente al detectar un rostro que porte o no el cubrebocas de una persona. El sistema desarrollado tiene la finalidad asistir y/o remplazar al personal que se encarga de validar la portación de cubrebocas, para ello tiene una eficiencia de detección que supera el 80%. De igual manera tiene la capacidad de detectar rostros cuando se usan prendas en la cabeza como bufandas, gorros, sombreros. Así como, la portación del cubrebocas con múltiples colores y formas. Una parte relevante de la investigación es el entrenamiento del algoritmo *LBPH*, el cual se implementó para la clasificación de imágenes de rostros de personas. Para el entrenamiento, se utilizaron 2000 imágenes con una baja resolución y con diferentes luminosidades y ángulos. Para su funcionamiento el sistema se implanta en un dispositivo *Raspberry Pi 3 B*.

2. ESTADO DEL ARTE

La detección de personas que portan cubrebocas utilizando visión artificial es un campo de investigación en crecimiento donde se combina la inteligencia artificial con la tecnología reconocer imágenes en de personas que utilizan cubrebocas. A continuación, se describen trabajos relacionados con la problemática antes mencionada. En el trabajo de (Badoni y otros, 2024), se implementó un sistema con visión artificial utilizando un dispositivo embebido para un control de acceso a una vivienda, en el cual se estableció en sus requerimientos que se ve directamente afectado el rendimiento del *CPU* y procesador gráfico ya que se procesan imágenes y o gráficos. Utilizan un algoritmo basado en *PCA* con el cual extrae características de rostros de personas que acceden a una casa o residencia. Por otra parte, en (Kalkar y otros, 2020), se presenta una propuesta de sistema que destaca por su rentabilidad y mayor resolución, además de la capacidad de detectar intrusos sin ayuda externa. Se emplea el sensor de movimiento PIR, en conjunto con *Raspberry Pi*, para detectar movimientos y, en consecuencia, activar un video a través del módulo de cámara Pi. Las aplicaciones basadas en *IOT* permiten el acceso remoto para visualizar imágenes y recibir

notificaciones por correo electrónico en caso de detección de intrusos. En el trabajo de (Yuzchen y otros, 2021), se propone un sistema de detección con base en el teléfono móvil. En primera instancia, se extraen cuatro características de los *GLCM* de microfotografías de la mascarilla facial. Al final, se logra un sistema de detección con tres resultados, utilizando el algoritmo *KNN*. Así mismo en (Loey y otros, 2021), se presenta un modelo híbrido que utiliza el aprendizaje automático profundo y el clásico para identificar mascarillas faciales. El modelo propuesto consta de dos componentes: uno para la extracción de funciones mediante *Resnet50* y otro para el proceso de clasificación mediante árboles de decisión, además de una máquina vectorial de soporte (*SVM*) y el algoritmo de conjunto. Además, seleccionaron tres conjuntos de datos con personas que llevaban mascarillas: el conjunto de datos de rostros enmascarados del mundo real (*RMFD*), el conjunto de datos de rostros enmascarados simulados (*SMFD*) y *Labeled Faces in Nature (LFW)*. El clasificador *SVM* demostró una tasa de precisión de prueba del 99,64% en *RMFD*, del 99,49% en *SMFD* y del 100% en *LFW*.

En (Nagrath y otros, 2021), propone un enfoque novedoso que utiliza técnicas avanzadas, como el aprendizaje profundo, *TensorFlow*, *Keras* y *OpenCV*, para detectar eficazmente las máscaras faciales y utiliza el algoritmo *Single Shot Multibox Detector* para la detección de rostros y la arquitectura *MobileNetV2* para el clasificador, que se caracteriza por su diseño ligero y su capacidad de funcionar en tiempo real en dispositivos integrados, como *NVIDIA Jetson Nano* y *Raspberry Pi*. La técnica propuesta presenta una puntuación de alta precisión de 0,9264 y una puntuación de F1 de 0,93 y es útil para el reconocimiento facial, incluidos los que se centran en los puntos de referencia faciales y la detección de rasgos faciales. En (Perez Cayetano y otros, 2021), desarrollaron un sistema capaz de detectar rostros con y sin cubrebocas para la gestión de acceso a una institución educativa, el cual tiene como propósito asistir a las personas encargadas de verificar la portación de cubrebocas o en su defecto sustituirlas por un sistema capaz de realizar su labor, el cual está basado en *Deep Learning*. Su algoritmo tiene un grado de exactitud de un 92% en sus pruebas. Cabe destacar que en este artículo no se menciona que este embebido en un dispositivo o cuenta con medidas de señalamiento.

En (Guo y otros, 2022), proponen un sistema de seguridad basado en el reconocimiento facial para el tráfico ferroviario urbano, utilizan un algoritmo de reconocimiento facial del sistema de seguridad inteligente empleado en el tráfico, introducen la tecnología de reconocimiento facial y el análisis del modo principal de reconocimiento facial, proporcionan ideas prácticas para el diseño de aplicaciones. Al final cuando se utiliza el software *Mugshot* para el reconocimiento mediante una cámara web, muestra un resultado del 5,21%.

Por su parte (Cardenas y otros, 2017), compara mediante un análisis varias redes neuronales para el reconocimiento facial. En dicha comparación se mencionan las redes neuronales; *Hopfiel*, *Backpropagation*, *Algoritmo PCA*. Donde *Hopfiel* es el mejor con un 80% de eficiencia en el reconocimiento fácil. Cabe destacar que las pruebas se realizaron en igualdad de condiciones.

3. SISTEMA DE RECONOCIMIENTO FACIAL

El reconocimiento facial es una tecnología de vanguardia que recientemente ha incursionado en diferentes campos de la investigación incluyendo el análisis de imágenes y la extracción de características de archivos digitales. Este proceso tecnológico ha atraído la atención de la comunidad científica, ya que tiene la capacidad de emular la aptitud del cerebro humano para el reconocimiento de individuos basado en patrones. En concreto, este mecanismo está situado en el lóbulo temporal, que se encuentra debajo de las sienes y es responsable de reconocer rasgos faciales como el color de la piel, el tamaño de los ojos, así como la nariz y la boca. Estas regiones se denominan áreas faciales que son esenciales para determinar los atributos faciales cruciales. Por el contrario, un sistema de reconocimiento facial es una aplicación informática que sirve para identificar automáticamente a las personas. Esta tecnología se hace viable gracias a tres etapas fundamentales: detección de rostros, extracción de características y reconocimiento. La detección facial se refiere al proceso computacional mediante el cual se utiliza un software para determinar la presencia o ausencia de rostros humanos en fotografías o videos. Cabe señalar que la

detección facial no implica la identificación de individuos, sino que se centra únicamente en la detección de la presencia o ausencia de un rostro. Por lo tanto, es imperativo tener en cuenta que el reconocimiento facial no guarda y tampoco retiene ninguna información o detalle de la persona detectada, ya que es completamente anónimo. En consecuencia, si el software detecta un rostro en una determinada foto o vídeo, vuelve a detectar el mismo rostro más adelante, no reconocerá el rostro como el del mismo individuo, ya que solo captura el rostro humano en una sola imagen determinada. No obstante, es pertinente señalar que la detección facial puede contener información demográfica, como el sexo o la edad de la persona detectada, que puede ser útil para las estadísticas demográficas. En esencia, la detección facial por sí sola no puede utilizarse para el reconocimiento de un individuo.

4. REDES CONVOLUCIONALES EN CASCADA MULTITAREA (MTCNN) PARA LA DETECCIÓN DE ROSTROS Y LA ALINEACIÓN DE LOS PUNTOS DE REFERENCIA FACIALES

Para la detección de rostros en sistemas de reconocimiento automático existen diferentes clasificadores para reconocer un rostro específico. El funcionamiento de *MTCNN* de acuerdo con el trabajo realizado por (Zhang y otros, 2023) es útil para el reconocimiento de rostros en diferentes poses, con diferente iluminación, etc. El desarrollo de la aplicación es útil para la de detección de rostros y, en última instancia, sobre el funcionamiento del entrenamiento de un clasificador para verificar de quien es el rostro y difuminarlo en un vídeo en las próximas detecciones. Las redes en cascada multitarea convolucionales (*MTCNN*), es un marco formulado como una solución para la detección de rostros, así como para la alineación facial. El procedimiento consta de tres etapas de redes convolucionales que son capaces de reconocer rostros y la posición de los puntos de referencia, como los ojos, la nariz y la boca. El trabajo explica como *MTCNN* funciona como un medio para integrar ambas tareas (reconocimiento y alineación de rostros) mediante el aprendizaje multitarea. En la primera etapa, emplea una *CNN* superficial para generar rápidamente ventanas de candidatos. Posteriormente, en la segunda etapa, mejora las ventanas de candidatos sugeridas a través de una *CNN* más intrincada. Finalmente, en la tercera etapa, utiliza una tercera *CNN*, que es más compleja que las anteriores, para refinar aún más el resultado y producir posiciones de puntos de referencia faciales. El algoritmo *MTCNN* se compone de tres etapas. El paso inicial consiste en cambiar el tamaño de la imagen a diferentes escalas para construir una pirámide de imágenes.

Esta pirámide de imágenes sirve como entrada a la siguiente red en cascada que consta de tres etapas. La imagen se redimensiona a diferentes escalas para crear la pirámide antes mencionada. La primera etapa, conocida como Red de propuestas (*P-Net*), es una red totalmente convolucional (*FCN*). A diferencia de una *CNN*, una *FCN* no incluye una capa densa como parte de su arquitectura. La red de propuestas obtenida en esta etapa se utiliza para adquirir las ventanas candidatas junto con sus correspondientes vectores de regresión de cuadros delimitadores. La regresión de límites es una técnica muy utilizada para predecir la ubicación de los fotogramas cuando el objetivo es detectar un objeto de una clase predefinida, como las caras en este caso. Tras obtener los vectores del cuadro límite, se lleva a cabo el refinamiento para integrar las áreas superpuestas. El resultado final de esta etapa es el perfeccionamiento del conjunto de ventanas de candidatos, lo que se traduce en una reducción del volumen de candidatos.

En la etapa 1, la red de propuestas (*P-Net*), se utiliza una red totalmente convolucional (*FCN*). A diferencia de una *CNN*, una *FCN* no incorpora una capa densa dentro de su arquitectura. La función principal de esta red de propuestas es obtener ventanas candidatas, junto con sus respectivos vectores de regresión de cuadros delimitadores. La regresión de límites es una técnica muy empleada para predecir la ubicación del marco, especialmente en los casos en que el objetivo es detectar una clase predefinida de objetos, como rostros. Tras la adquisición de los vectores del cuadro delimitador, se ejecuta el refinamiento para fusionar las regiones superpuestas. El resultado final de esta etapa es un conjunto de ventanas candidatas, después del refinamiento, cuyo volumen se ha reducido.

En la etapa 2, implica la utilización de la red *Refine (R-Net)*, en la que se procesan todos los candidatos a *P-Net*. Cabe destacar que esta red es una red neuronal convolucional (*CNN*) a diferencia de la anterior, que era una red totalmente convolucional (*FCN*), debido a la presencia de una capa densa en la etapa final de la arquitectura de la red. El *R-Net* reduce eficazmente el número de candidatos, realiza la calibración mediante regresión de recuadros delimitadores e integra la supresión no máxima (*NMS*) para fusionar los candidatos superpuestos. La *R-Net* produce un vector de 4 elementos que define el cuadro delimitador de la cara, un vector de 10 elementos para localizar los puntos de referencia faciales y si la entrada es una cara o no.

La etapa 3 involucra la red saliente (*O-Net*), que tiene cierto parecido con la *R-Net*. Sin embargo, el objetivo de *O-Net* es proporcionar una descripción más detallada del rostro, con el objetivo de generar las ubicaciones precisas de los cinco puntos de referencia faciales clave: los ojos, la nariz y la boca.

Las tres tareas fundamentales del *MTCNN* implican la generación de tres componentes cruciales, a saber, la clasificación facial y no facial, la regresión del marco límite y la localización de los puntos de referencia faciales. La tarea de clasificación facial binaria plantea un desafío importante que implica la utilización de la pérdida de entropía cruzada ver fórmula.

$$L_i^{\text{det}} = (y_i^{\text{det}} \log(P_i) + (1 - y_i^{\text{det}})(1 - \log(P_i))) \quad (1)$$

Donde:

$$y_i^{\text{det}} \text{ etiqueta de ground truth donde } y_i^{\text{det}} \in \{0,1\}$$

El objetivo de aprendizaje de la regresión de bounding box se enmarca como un problema de regresión. Implica calcular el desfase entre las ventanas candidatas y la realidad del terreno más cercana. Esto se logra mediante la aplicación de la pérdida euclidiana como función de pérdida ver Ecuación (2).

$$L_i^{\text{box}} = \|\hat{y}_i^{\text{box}} - y_i^{\text{box}}\| \quad (2)$$

Donde:

$\hat{y}_i^{\text{box}}$ es el objetivo extraído de la red neuronal.

y_i^{box} es el objetivo extraído de la red neuronal

El proceso de localización de puntos de referencia faciales implica formular sus respectivas ubicaciones como un problema de regresión. La función de pérdida utilizada en este proceso es la distancia euclidiana ver Ecuación (3).

$$L_i^{\text{landmark}} = \|\hat{y}_i^{\text{landmark}} - y_i^{\text{landmark}}\|_2^2 \quad (3)$$

5. ALGORITMO LBPH

La utilización de un algoritmo de reconocimiento facial es necesario para identificar el rostro de un individuo. El rendimiento de este algoritmo es conocido por su capacidad para reconocer los rasgos faciales desde la perspectiva frontal y lateral. *Open CV* también incorpora dos modelos adicionales, *EigenFace* y *FisherFace*, que adoptan un enfoque holístico para facilitar la identificación. La información se ve como un vector dentro de un espacio de alta dimensión, el *LBPH* evita esto considerando únicamente las características locales de un objeto, en lugar de examinar la imagen en su totalidad como un vector de gran dimensión. Esta forma de extracción de las características produce implícitamente una dimensionalidad reducida. La base de este enfoque se basa en el operador de patrones binarios locales (*LBP*), que se basa en el análisis de texturas *2D*.

Pseudocódigo para detección de rostros con mascarilla

Data: Imágenes de entrenamiento con y sin mascarilla

Result: Clasificación de personas con o sin mascarilla y alerta en caso necesario

Inicio del sistema:

1. **Cargar entorno del dispositivo:** Configurar el sistema embebido en Raspberry Pi e inicializar los módulos necesarios (red neuronal y algoritmo *LBPH*).
2. **Entrenamiento de la red neuronal:**
 1. Obtener imágenes de rostros con y sin mascarilla.
 2. Preprocesar las imágenes (escalado, normalización, etc.).
 3. Entrenar la red neuronal *Multi-task cascade convolutional networks (MTCNN)* con las imágenes etiquetadas.
3. **Detección de rostro:**
 1. Capturar imágenes de video en tiempo real.
 2. Usar *MTCNN* para detectar la región del rostro en las imágenes.
4. **Extracción de características:**
 1. Aplicar el algoritmo *Local Binary Patterns Histogram (LBPH)* en las regiones de interés (rostros detectados).
 2. Obtener un vector de características representativo para cada rostro.
5. **Clasificación y alerta:**
 1. Clasificar si el rostro tiene o no mascarilla utilizando el modelo entrenado.
 2. Si no tiene mascarilla:
 - Generar una alerta audible.
 - Denegar el acceso al sistema.
 3. Si tiene mascarilla:
 - Otorgar acceso.
6. **Evaluación del sistema:**
 1. Medir precisión del sistema utilizando un conjunto de validación.
 2. Reportar precisión promedio (93% según los resultados obtenidos).

Fin del algoritmo

5.1 Operador *LBPH*

La técnica del patrón binario local (*LBP*) se emplea con el propósito de representar brevemente la estructura local de una imagen mediante una comparación de cada píxel con sus vecinos inmediatos. Específicamente, un píxel elegido sirve como punto focal, y los píxeles que lo rodean están sujetos a una restricción en sus valores. Para obtener más información, si la intensidad del píxel central supera a la de sus vecinos, se le asigna un valor de cero. Por el contrario, si la intensidad del píxel vecino es mayor o igual que la del píxel central, el valor asignado es uno. Vale la pena señalar que el patrón binario representado en la Figura 1, sirve como ejemplo, en el que un área localizada de la imagen está representada por la secuencia 00010011.

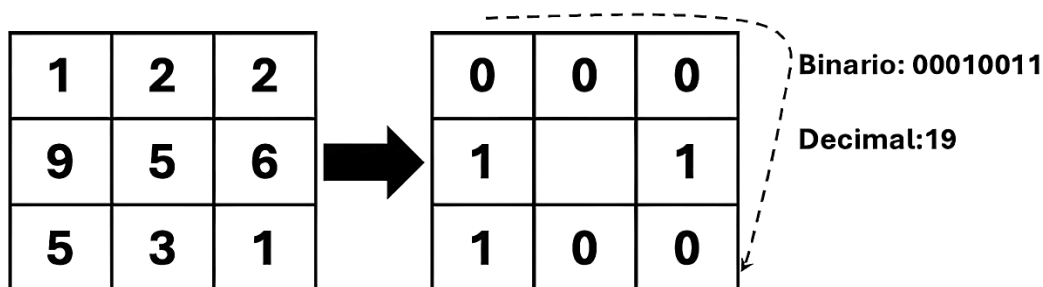


Figura 1. Representación gráfica del funcionamiento del operador *LBPH* donde se lee una matriz de valores

La Ecuación (4) presenta la fórmula del operador *LBP*, que recorre secuencialmente la matriz de 3 x 3 seleccionando un píxel central y sus vecinos circundantes. El píxel central se compara sistemáticamente con cada uno de sus vecinos. En el caso de que el píxel central sea más pequeño que el píxel comparado, se asigna un valor de 1, mientras que, en el caso contrario, se designa un valor de 0. Es importante señalar que *P* indica el número de vecinos para tener en cuenta, mientras que *R* representa el tamaño del vecindario. Además, g_c y g_p se refieren respectivamente a los valores de gris del píxel central y de cada uno de los *p* píxeles de la vecindad.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (4)$$

En la Figura 1 presentada, la derivación de la Ecuación (5) se demuestra utilizando el ejemplo mencionado anteriormente. Cabe destacar que la función $s(x)$ se utiliza para comparar la diferencia entre el valor del píxel actualmente evaluado y el píxel central. Si esta diferencia es mayor o igual a cero, la función devuelve un valor de 1. Sin embargo, si la diferencia es negativa o, en otras palabras, inferior a cero, la función devuelve el valor 0.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (5)$$

Continuando con el análisis de los histogramas de patrones binarios locales (LBPH), el primer paso consiste en obtener un inventario completo de los patrones binarios locales. Posteriormente, cada patrón se convierte metódicamente en un número decimal mediante la transición de una representación binaria a una decimal. Finalmente, se genera un histograma a partir de todos los valores decimales obtenidos. Para obtener una representación visual, consulte la Figura 2.

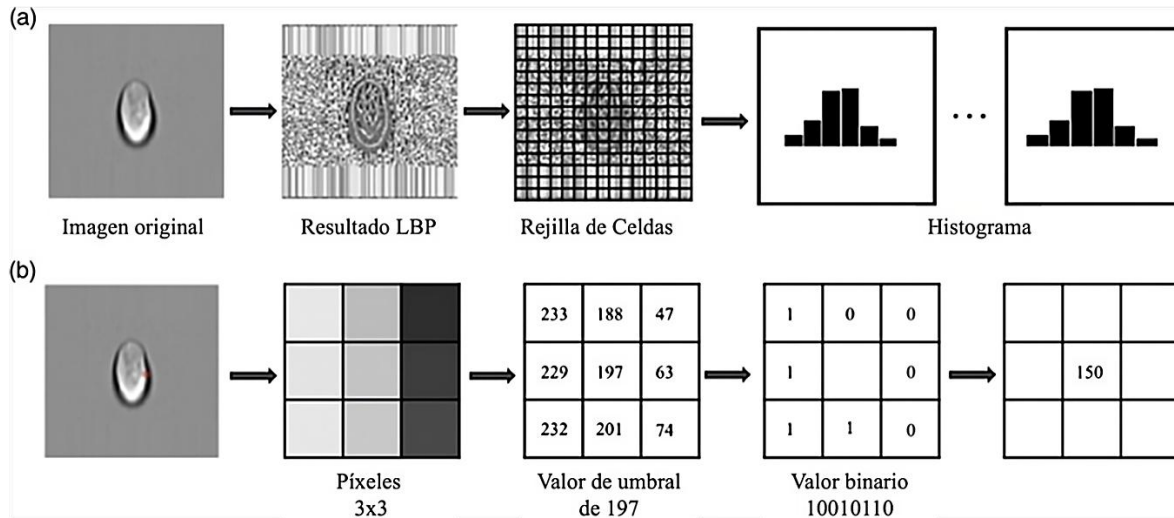


Figura 2. Etapas consecutivas para obtener los patrones binarios de una imagen y la obtención del histograma con el algoritmo *LBP*

El histograma final se crea para cada cara en el conjunto de datos de entrenamiento. Eso significa que, si había 100 imágenes en el conjunto de datos de entrenamiento, *LBPH* extraerá 100 histogramas después del entrenamiento y los almacenará para su posterior reconocimiento.

5.1.1. Reconocimiento

El reconocimiento facial que utiliza el método del histograma de patrón binario local (*LBPH*) es un proceso multifacético que comprende cuatro etapas distintas. La etapa inicial implica la presentación de una imagen novedosa al sistema de reconocimiento facial para su análisis. Posteriormente, el reconocedor genera un histograma para esta imagen en particular. El histograma generado se compara entonces con los histogramas existentes en el sistema. A continuación, se determina la coincidencia óptima y se devuelve la etiqueta de persona correspondiente. Para comparar los histogramas, se pueden emplear varios métodos, incluida la distancia euclidiana, el chi-cuadrado, entre otros. La fórmula de distancia euclidiana utilizada para comparar dos histogramas se representa en la Ecuación (6).

$$D = \sqrt{\sum_{i=1}^n (\text{hist1}_i - \text{hist2}_i)^2} \quad (6)$$

Para emplear la técnica del *chi-cuadrado* como medio de comparar dos histogramas, es imperativo utilizar la fórmula que se muestra en la Ecuación (7). En esta fórmula, n indica el número total de elementos presentes en el histograma, mientras que S y M representan histogramas diferentes.

$$\chi^2(S, M) = \sum_{i=1}^n \frac{(S_i - M_i)^2}{S_i + M_i} \quad (7)$$

6. CARACTERÍSTICAS DEL HARDWARE

A continuación, se describe la tarjeta electrónica utilizada en este proyecto, el cual es el *Rasp Berry Pi*, la cual fue útil para la configuración y del sistema de detección de rostros con cubre boca y sin cubre boca. Esta sección abarca la placa de desarrollo *Raspberry Pi* en todas sus versiones, reconocidas a nivel mundial por su calidad excepcional. Estas placas no están diseñadas explícitamente para acelerar las aplicaciones de aprendizaje profundo, sino que sirven como tableros de uso general para los estudiantes de informática. A lo largo de los años, se han lanzado numerosos modelos y revisiones de esta placa. Específicamente, uno de estos modelos se empleará en el proyecto en discusión. La *Raspberry Pi 4* es actualmente la versión más avanzada, mientras que el proyecto utilizará la *Raspberry Pi 3 B+*. En las fases de recolección y entrenamiento, el sistema se ejecutó en una PC de escritorio la cual cuenta con las siguientes características:

- CPU: Xeon E5-2665 a 2.40 GHz con 8 hilos y 16 núcleos.
- RAM: 8gb Quad Channel DDR3 a 1333Mhz.
- Almacenamiento: Raid 0 con 2 SSD SATA3.
- GPU: Radeon RX 560 4 Gb.
- OS: Windows 11.
- Cámara: 2 Mp a 720p.

En la fase de detección de imágenes se utilizó una placa Raspberry pi 3 Modelo B, la cual cuenta con las siguientes características:

- CPU: ARMv8 Cortex-A53 64-bit Quad-Core @ 1.2GHz.
- RAM: 1gb.
- Almacenamiento: SD de 16 GB clase 10.
- GPU: Broadcom VideoCore IV @ 400MHz
- OS: Raspberry OS (Legacy)

- Cámara: 2 Mp a 720p.

6.1 Instalación de sistema operativo y librerías

El sistema operativo que se instaló fue *Raspberry OS*, en su versión *LEGACY*, ya que esta versión es más compatible con las librerías. Las librerías se instalaron el gestor de paquetes integrado de *Python (PIP)*. Para la instalación de sistema operativo, se utiliza una herramienta proveída por el fabricante, es muy fácil de usar y mantiene versiones actualizadas de las distribuciones. Como se puede apreciar en Figura 3, sólo se elige la distribución a instalar, la tarjeta *SD* donde se desea instalar y finalmente se instala el sistema.

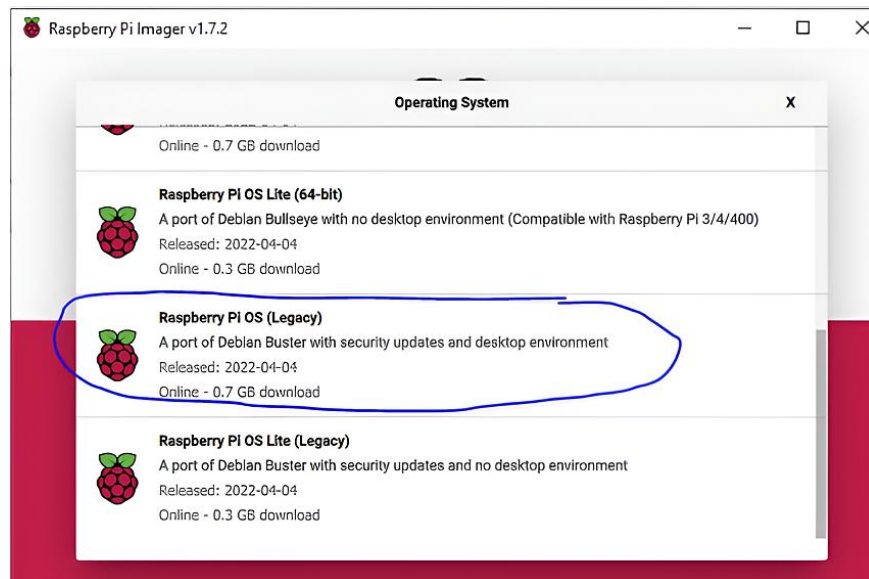


Figura 3. Interfaz gráfica sistema operativo Rasp berry versión Legacy

6.2 Proceso experimental en la detección de rostros con cubreboca

En esta sección se describe brevemente la propuesta para la detección de rostros con cubrebocas. Básicamente algoritmo de la Figura 4, consiste en las etapas siguientes:

- Creación de un dataset de imágenes de rostros con cubrebocas y sin cubrebocas
- Entrenamiento y detección de rostros con la red neuronal *MTCNN*
- Análisis y configuración del algoritmo *LBPH* para la detección y clasificación de rostros
- Configuración del sistema de detección en tarjeta *RASP BERRY Pi*
- Identificación de rostros que traen cubrebocas y que no traen cubrebocas

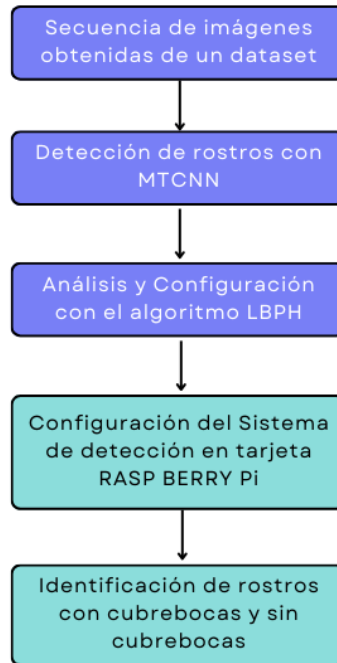


Figura 4. Metodología propuesta para la detección de rostros con cubrebocas y sin cubrebocas

7. DESARROLLO DEL SISTEMA DE DETECCIÓN

En este apartado se explican las etapas en las cuales se realizó el sistema de detección de rostros con cubrebocas o sin cubrebocas.

7.1 Recolección de datos.

Las imágenes se captan a 2 *MP*, pero en la recolección asistida de imágenes, el rostro se recorta con una resolución de 150 por 200. A si mismo tienen un formato de salida “*JPG*”. La construcción del dataset fue a partir de fotogramas captados por una webcam, conectada vía *USB*. Cada imagen es recortada automáticamente de tal forma que no exista una inferencia con el fondo. Esto es posible gracias a la biblioteca *MTCNN* que como ya se ha mencionado funge como detector facial previendo las coordenadas del rostro detectado. Finalmente se guardan como archivo utilizando *OpenCV*, en específico el método “*imwrite()*”.

La cantidad de imágenes tomadas en total son de 1000 imágenes por clase, siendo 2000 en total. Adicionalmente y como se puede observar en la Figura 5, las imágenes se toman con diferente luminosidad y así mismo desde diferentes puntos y ángulos para poder evitar lo más posible malas lecturas y/o falsos positivos. Esta variación de luminosidad se logra gracias a una fuente alternativa, ya sea girándola alrededor o moviendo de arriba hacia abajo. Nótese que la luminosidad va incrementándose de la sección A hasta la N, Figura 5.



Figura 5. Diferencial de iluminación en algunos fotogramas con rostros sin cubreboca

El *dataset* que se construyó para el entrenamiento del algoritmo *LBPH*, cuenta con dos clases; “Persona_Con_Cubrebocas” y “Persona_Sin_Cubrebocas”. En la Figura 6, se puede apreciar una muestra de cada una.



Figura 6. Muestra de 2 clases a representar, "persona con cubreboca" o "persona sin cubreboca"

Adicionalmente a lo anterior es altamente importante depurar el dataset, ya que la anteriormente mencionada red neuronal convolucional puede tener fallos a la hora de detectar el rostro a recortar. Estas muestras defectuosas, Figura 7, pueden producir falsos positivos a la hora de la detección.

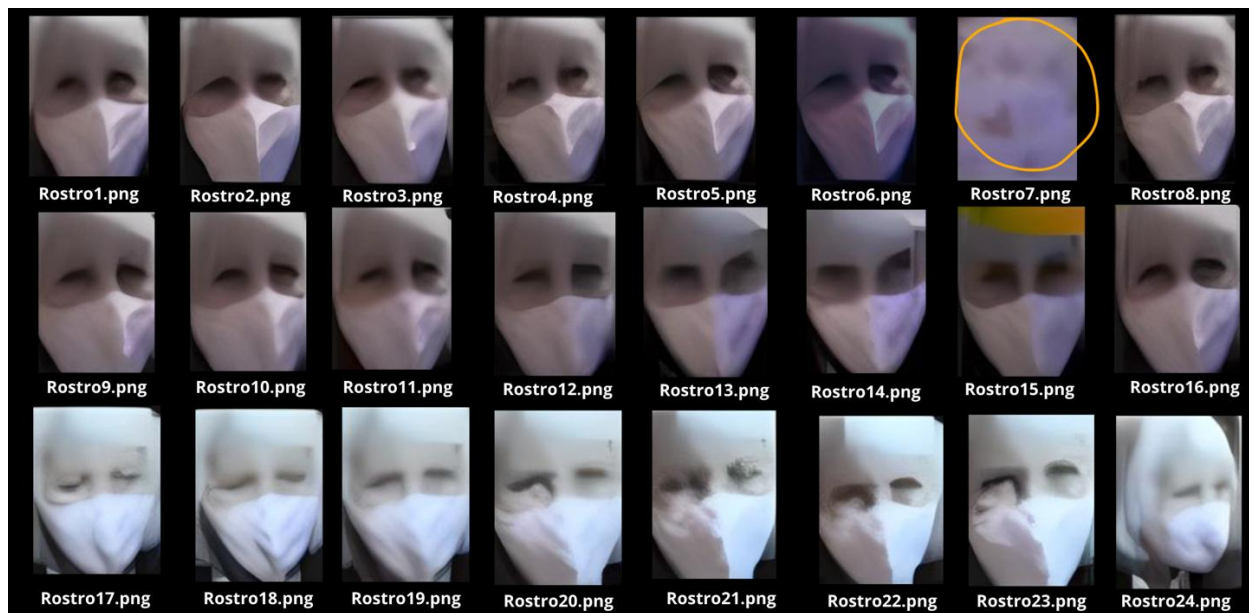


Figura 7. Interferencia en la detección facial

7.2 Conformación y depuración del dataset

Adicionalmente a lo anterior es altamente importante depurar el dataset, ya que la anteriormente mencionada red neuronal convolucional puede tener fallos a la hora de detectar el rostro a recortar. Estas muestras defectuosas se vieron en la Figura 7, pueden producir falsos positivos a la hora de la detección.

Finalmente, se implementa el algoritmo *LBPH*, provisto por la librería de *OpenCV*, pasándole como atributo las características obtenidas al leer las imágenes del dataset. Para que por último se escriba el modelado en formato *XML*, véase la Figura 8.

7.2.1 Detección

La capa de detección ocupa la librería *MTCNN*, la cual es la red neuronal convolucional que se ocupa para detectar rostros de manera genérica, es decir; detecta rostros ya sea con cubrebocas y sin cubrebocas. Posteriormente devuelve la afinidad y las coordenadas del rostro en cuestión para después evaluar si cumple con un mínimo de 85% de afinidad. Ya obtenidas las coordenadas se recorta a 150 por 200 píxeles el rostro y se convierte en escalas de grises, ya que es una condición que se necesita para la predicción implementando el algoritmo *LBPH*, ya que estas dos condiciones optimizan su funcionamiento. Junto con el sistema de detección.

7.2.2. Control de acceso

Para finalizar, dependiendo de la clase detectada, se da o no acceso. Nótese que en la Figura 4, cuando el resultado de la predicción del algoritmo *LBPH*; es 1, simplemente no hace nada. Análogamente cuando es 0 en 8 fotogramas seguidos, concede el acceso; encendiendo el led verde y reproduciendo audio.

7.2.3. Evaluación de resultados

En esta etapa de pruebas se implementaron pruebas unitarias, se realizaron con el módulo de test, que es capaz de determinar de forma casi automática (requiere supervisión) el tipo de imagen de entrada.

Con este módulo se corrió una prueba con 50 imágenes de forma unitaria. Para esto se crea de forma automatizada la carpeta y a partir de la cámara se obtienen imágenes que se evalúa la portación de cubrebocas, de ser correcta, se agrega a la carpeta. Así mismo, cabe resaltar que se tiene que repetir la prueba ya que son dos clases en el clasificador, una con portación de cubrebocas y otra sin cubrebocas.

Finalmente se promedia el resultado de ambas pruebas. Ahora bien, como tal no es tan automatizado o mejor dicho sin libre supervisión ya que hay que evaluar manualmente el número de falsos positivos y falsos negativos. Los cuales son evaluados de la siguiente manera:

TP: Cuando se detecta cubrebocas correctamente.

TN: Cuando no se detecta cubrebocas correctamente.

FP: Cuando no se espera con cubrebocas, pero detecta la portación de este.

FN: Cuando no se detecta cubrebocas, pero se espera que lo detecte.

Las características que evaluar son:

- a) Precisión, representa el porcentaje de predicciones correctas con respecto al total, Ecuación (8).

$$\text{precisión} = \frac{TP}{TP + FP} \quad (8)$$

- b) Sensibilidad y exhaustividad, representa la tasa positiva verdadera o TP. Es la relación de casos positivos bien clasificados por el modelo sobre el número total de casos positivos, Ecuación (9).

$$\text{Exhaustividad} = \frac{TP}{TP + FN} \quad (9)$$

- c) La exactitud Indica qué tan cerca está la medición del valor real, Ecuación (10).

$$\text{Exactitud} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

En la Figura 8, se aprecia el sistema semi automático, etiqueta cada una de las fotos, tal que en este caso la prueba consistía en detectar rostros sin cubrebocas, le asigno la etiqueta “correcto” a la imagen ya que no se porta cubrebocas.

La precisión representa el porcentaje de predicciones correctas con respecto al total. Es por tanto el cociente entre los casos bien clasificados por el modelo (verdaderos positivos y negativos, es decir, los valores de la diagonal de la matriz de confusión) y la suma de todos los casos.

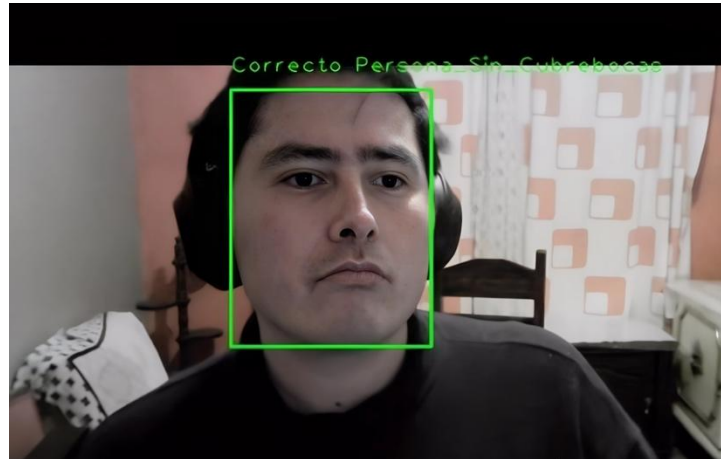


Figura 8. Resultado correcto al momento de detectar un rostro sin cubrebocas.

7.2.4. Evaluación de la detección de los rostros con cubrebocas y sin cubrebocas

La prueba consiste en 50 muestreos sin cubrebocas en la que 48 imágenes se guardaron en disco automáticamente, esto quiere decir que la primera prueba arrojó:

- TN: 48
- FP:2

En la Figura 9, se pueden observar las evidencias de dicha prueba. En dicha prueba hubo 2 imágenes en las que ni siquiera se pudo detectar como rostro, por eso se designa como un falso positivo.

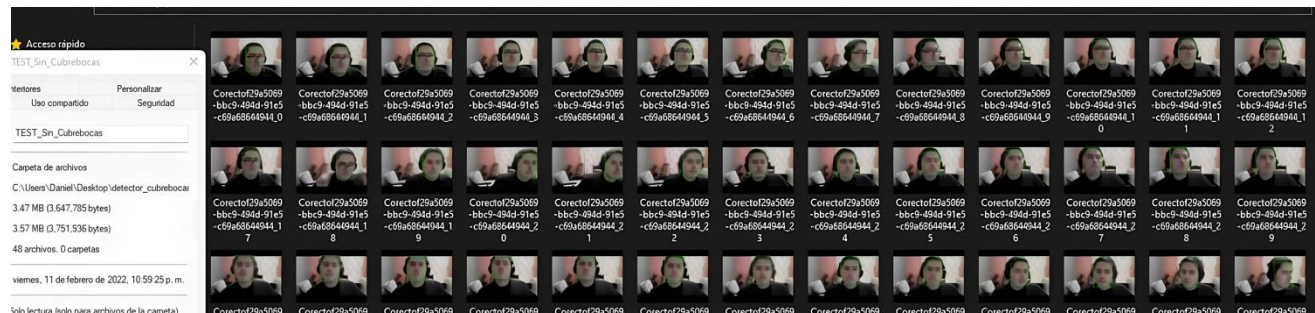


Figura 9. Conjunto de imágenes correspondientes al test de detección de personas sin cubrebocas

En la segunda prueba, en el cual se probó la detección de cubrebocas se obtuvieron los siguientes resultados.

- TP:44
- FN:6

En la Figura 10, se puede apreciar de nuevo el etiquetado automático del segunda prueba. En este caso se señala que se porta cubrebocas.



Figura 10. Resultado correcto al momento de detectar un rostro con cubrebocas

Así mismo en la Figura 11, se pueden apreciar las evidencias de la prueba, recordando que se realizan la toma de 50 fotogramas. A sí mismo se puede notar que se le asigna un nombre único para no crear conflicto de escritura o que rescriba un archivo ya existente causando un sesgo en el cálculo.

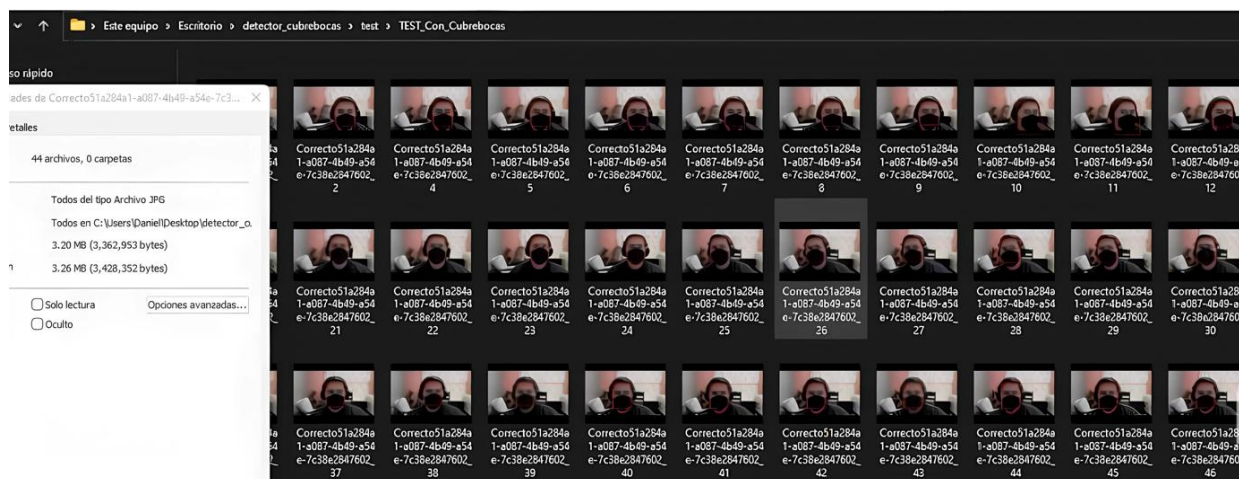


Figura 11. Conjunto de imágenes correspondientes al test de detección de personas con cubrebocas

Después de haber realizado las pruebas con un solo rostro, se procedió a realizar pruebas con distintas personas. En este caso se requirió la colaboración de 10 personas anónimas, cuyas escenas donde se localizaron estas personas se videograbaron bajo condiciones de control de luz con lo cual se obtuvieron imágenes de los rostros que sirvieron para realizar tanto el entrenamiento como las pruebas y con ello se utiliza la técnica de la sección anterior. En la Tabla 1 se muestran los resultados de la prueba mencionada, cabe mencionar que en el caso del test efectuado en las personas 3, 4 y 9 utilizaban prendas tales como; gafas, gorro y bufanda. Por otra parte, el promedio de eficiencia o mejor dicho la exactitud de reconocimiento es del 92.1%. La primera columna de dicha tabla representa el número de 44. persona a la que se está aplicando el Test. La segunda columna es el número de fotogramas tomados.

En la tercera columna se encuentra el número de fotogramas que no se detectó el cubrebocas de forma correcta, es decir; no se espera detectar cubrebocas y no se detecta cubrebocas. La columna FP es lo contrario a la anterior; son los fotogramas en los que no se espera la portación de cubrebocas, pero se detecta este mismo. En la quinta columna se encuentra el número de imágenes en la que se detecta correctamente el cubrebocas. La siguiente columna son los fotogramas que no se detectó la presencia de cubrebocas, pero se esperaba captar la portación de este. Finalmente se tienen los resultados de las columnas de la precisión, exhaustividad y exactitud que así mismo so el resultado de aplicar las fórmulas de la sección anterior sobre las columnas *TN*, *FP*, *TP* y *FN*.

Visto de forma gráfica, tal cual como se ve en la Figura 12, en el eje vertical, es el valor del resultado de las últimas tres columnas de la tabla anterior, representado en función al número de persona evaluada que así mismo viene siendo definido por el eje horizontal. Donde el color verde representa la precisión, el color azul oscuro describe la exhaustividad y por último el color rojo define la exactitud. Para transponder el valor de estas columnas, solo se requiere multiplicarlos por 100.

Nótese que en el caso aislado de las personas; 3, 4 y 9. El reconocimiento tiende a bajar la eficiencia en las pruebas donde se portaban prendas en la cabeza ya que estas mismas tienden a causar poca inferencia en la detección facial. En la Ecuación (10) se puede observar el calculo de la exactitud donde se verifica la cantidad de $TP=45$, $TN=47$, $FP=3$ y $FN=5$ y el resultado es de 92%

$$Exactitud = \frac{(45 + 47)}{45 + 47 + 3 + 5} = \frac{92}{100} = 0.92 \quad (10)$$

Tabla 1 Resultados obtenidos en una muestra de 10 fotogramas donde se observa los falsos positivos, falsos negativos, verdaderos negativos y verdaderos positivos.

Test								
# de personas	Fotogramas evaluados	TN	FP	TP	FN	Precisión	exhaustividad	exactitud
1	50	47	3	45	5	0.9375	0.9	0.92
2	50	48	2	49	1	0.960784314	0.98	0.97
3	50	43	7	48	2	0.872727273	0.96	0.91
4	50	42	8	44	6	0.846153846	0.88	0.86
5	50	48	2	45	5	0.957446809	0.9	0.93
6	50	49	1	46	4	0.978723404	0.92	0.95
7	50	48	2	42	8	0.954545455	0.84	0.9
8	50	49	1	47	3	0.979166667	0.94	0.96
9	50	48	2	40	10	0.952380952	0.8	0.88
10	50	45	5	48	2	0.905660377	0.96	0.93
Promedio		46.7	3.3	45.4	4.6	0.993450891	0.908	0.921

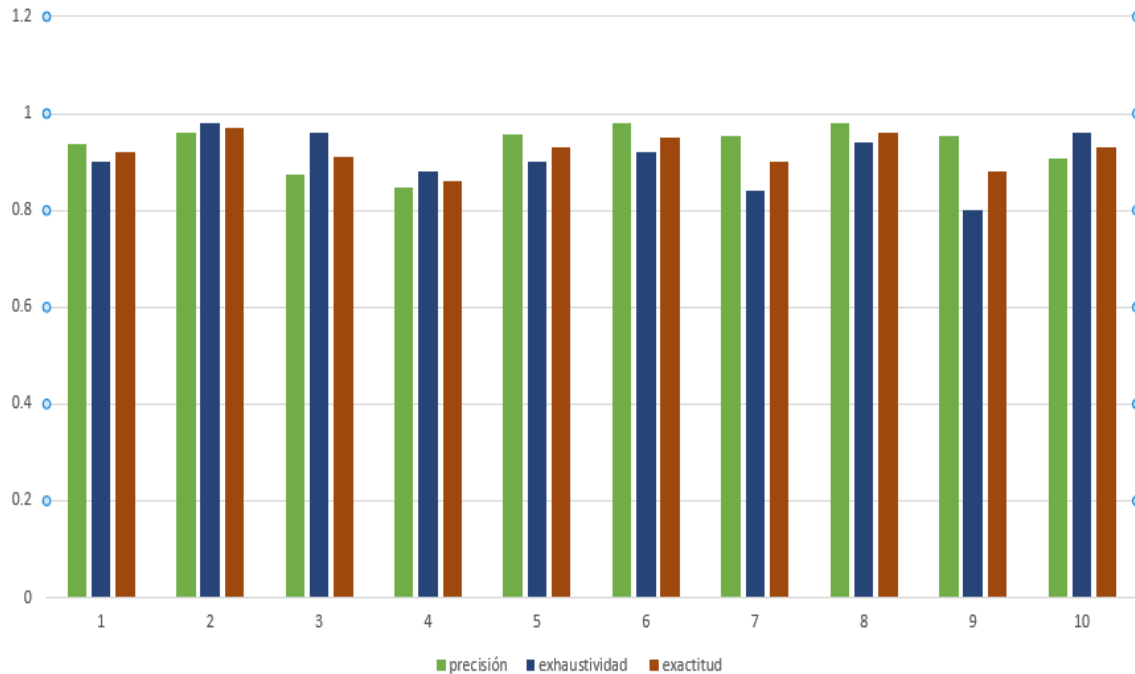


Figura 1. Gráfica de barras que muestra los resultados de precisión, exhaustividad y exactitud en la muestra de 10 personas

8. CONCLUSIONES

Durante el desarrollo del presente trabajo de investigación, se presentaron algunas problemáticas como pueden ser la luminosidad de las imágenes, problemas de compatibilidad de librerías, problemas de rendimiento. El uso de un sistema de acceso que consta de un software embebido en un dispositivo *Raspberry PI*, el cual contiene una aplicación formada por el uso de una red neuronal *Multi-task cascade convolutional networks (MTCNN)*, con la cual se hace el entrenamiento de las imágenes de rostros que traen y no traen cubrebocas, posteriormente y para una mejor identificación de los rostros se utiliza un algoritmo basado en *Histograma de Patrones Binarios Locales (LBPH)*, y que el dispositivo al final es capaz de emitir un pitido de alerta en caso de encontrar una persona que no contenga cubrebocas y esto es útil ya que lo que se espera es disminuir los casos de *COVID 19* ya que la estadísticas revelan que esta enfermedad respiratoria aumentan en épocas de frío. Por otra parte, usar un dispositivo embebido en una tarjeta como *Rasp Berry Pi* hace que el prototipo sea portable y capaz de detectar rostros en tiempo real y mostrar los resultados como lo es un sonido de alerta.

Por último, se puede concluir en base a las pruebas que se supera la meta del 80% de exactitud en el reconocimiento facial de forma propuesta desde un inicio. Aun con los obstáculos de prendas en la cara se alcanza un 92.1% de exactitud. De forma similar hay una inferencia entre la *MTCNN* y el algoritmo *LBPH*, tal que por un lado se detecte un rostro, pero no se haya clasificado bien el tipo de imagen y análogamente no se haya detectado el rostro y posiblemente si se haya podido clasificar bien, la solución propuesta e implementada fue encontrar un balance entre el mínimo porcentaje de exactitud aceptable de la *MTCCN* tal que se detecte lo más posible rostros a pesar de que tenga un cubreboca.

CONTRIBUCIÓN DE AUTORÍA CRediT

1. Juan Alberto Antonio Velázquez, autor principal
2. Leopoldo Gil Antonio , co-autor
3. Héctor Caballero Hernández, co-autor
4. Daniel Cervantes Ambríz, co-autor

Conceptualización: La idea principal de la investigación resultó de como crear un sistema de detección de personas que portan cubrebocas y las que no portan cubrebocas, para que en caso de que no traigan cubrebocas no dejarlas pasar a la institución, la idea y la formulación de los objetivos y metas generales de la investigación, consisten en usar un algoritmo basado en localización binario de patrones

La metodología del trabajo incluye:

- Desarrollo del modelo de detección: Utilización de redes neuronales Multi-task Cascade Convolutional Networks (MTCNN) para entrenar un modelo con imágenes de personas con y sin mascarillas.
- Extracción de características: Implementación del algoritmo de Histograma de Patrones Binarios Locales (LBPH) para obtener rasgos faciales significativos que permitan una clasificación efectiva.
- Clasificación y alertas: Diseño de un sistema que emita una alerta audible si se detecta que una persona no porta cubrebocas.

Software: El desarrollo de software incluyó:

- Programación e implementación: Creación de un programa embebido en un dispositivo Raspberry Pi.
- Integración de algoritmos: Codificación del modelo MTCNN y del algoritmo LBPH para entrenar y procesar datos de imágenes.
- Pruebas de funcionamiento: Evaluación del software en distintas condiciones para comprobar la precisión del sistema en la detección de mascarillas.

Validación: El sistema fue validado mediante pruebas de replicabilidad y precisión, alcanzando una efectividad promedio del 93% en la detección de rostros con mascarillas. Esto involucró comparar los resultados de las detecciones contra datos etiquetados manualmente para asegurar la confiabilidad del modelo.

Análisis formal: Se emplearon técnicas computacionales y matemáticas para:

- Entrenar la red neuronal MTCNN.
- Analizar las características obtenidas mediante LBPH.
- Sintetizar resultados para evaluar la efectividad del sistema propuesto en términos de precisión y confiabilidad.

Investigación: La investigación incluyó:

- Recolección de imágenes de rostros con y sin cubrebocas para el entrenamiento del modelo.
- Experimentación con diferentes configuraciones de los algoritmos y ajustes de parámetros para mejorar los resultados.

Recursos: El proyecto requirió:

- Un dispositivo Raspberry Pi para el sistema embebido.
- Herramientas de software para el entrenamiento y evaluación del modelo (por ejemplo, bibliotecas como OpenCV y TensorFlow).
- Imágenes de personas con y sin cubrebocas para el entrenamiento del sistema.

Curación de datos: Los datos utilizados para el entrenamiento y validación fueron cuidadosamente anotados y organizados para garantizar su calidad. Se creó un conjunto de datos etiquetados que incluyó tanto rostros con mascarillas como sin ellas.

Redacción - El trabajo fue revisado críticamente por miembros del equipo, quienes refinaron el contenido para mejorar su claridad y precisión antes de la publicación.

Redacción - Revisión y Edición: Preparación, creación o presentación del trabajo publicado por los miembros del grupo de investigación original, específicamente revisión crítica, comentario o revisión – incluyendo las etapas previas o posteriores a la publicación.

Supervisión: La supervisión del proyecto incluyó la planificación de las actividades de investigación, la guía técnica para el desarrollo del software y la evaluación de resultados.

Administración del proyecto: Se gestionó el cronograma, los recursos y las actividades necesarias para garantizar el cumplimiento de los objetivos del proyecto

Adquisición de fondos: El proyecto no contó con apoyo financiero sino que fueron con recursos propios con lo cual se permitió adquirir el dispositivo Raspberry Pi, recursos computacionales y software necesario para la investigación.

DECLARACIÓN DE INTERESES CONTRAPUESTOS

Los autores declaran que no tienen intereses financieros en conflicto ni relaciones personales conocidas que pudieran haber influido en el trabajo presentado en este artículo y declaramos que el artículo no ha sido publicado Dado que el sistema propuesto tiene aplicaciones en control de acceso, su implementación podría beneficiar a ciertas empresas, hospitales o entidades gubernamentales, sin embargo la aplicación es de fines educativos y su posible uso en la institución denominada Tecnológico de Estudios Superiores de Jocotitlán, Si alguno de los autores del trabajo tiene vínculos (financieros o laborales) con estas entidades, esto debe declararse, ya que podría interpretarse como un interés que influye en el desarrollo del proyecto. Dado que el sistema implica la detección de rostros, podría haber preocupaciones relacionadas con la privacidad de los datos recopilados y procesados. Si alguna entidad involucrada tiene intereses en el uso de estos datos con fines distintos a los declarados en el proyecto, esto también sería un conflicto de interés.

DISPONIBILIDAD DE DATOS

Los datos utilizados en este trabajo incluyen imágenes de rostros con y sin cubrebocas que han sido obtenidas y etiquetadas específicamente para este proyecto. Los datos han sido anonimizados y, debido a restricciones de privacidad, no están disponibles públicamente. Sin embargo, el modelo entrenado y el código fuente del software implementado en el dispositivo Raspberry Pi se encuentran disponibles en un repositorio público bajo una licencia de código abierto, accesible para su reproducción y validación. Los resultados de validación y las métricas del modelo pueden consultarse en el apéndice del presente trabajo. Esto permite un equilibrio entre la transparencia científica y el respeto a las restricciones legales y éticas.

AGRADECIMIENTOS

Este estudio no recibió ninguna subvención específica de ninguna agencia de financiación del sector público, comercial o sin fines de lucro.

REFERENCIAS

- Cardenas , F., V.T.T., M., Hernandez, J., & Hernandez, E. (2017). Reconocimiento facial mediante redes neuronales “hopfield”, “backpropagation” y algoritmo pca: un enfoque comparativo. *Ciencia Huasteca Boletín Científico de la Escuela Superior de Huejutla*.
- Loey, M., Manogaran, G., Hamed N. Taha, M., & M. Khalifa, N. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*.
- Badoni, P., Wadhwa, M., & Walia, R. (2024). System of IntelliGuard Access Using IoT. *Conference: 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*.
- Guo, Z., Xiao, Z., Alroobaea, R., M. Baqasah, A., Althobaiti, A., & Harsimranjit, S. (2022). Design and Study of Urban Rail Transit Security System Based on Face Recognition Technology. *Informatica*.
- Kalkar, S., Shubhankar, K., Kajale, R., & M. Kulkarni, S. (2020). Face recognition in embedded systems for security surveillance. *International Research Journal of Engineering and Technology (IRJET)*, 3655-3662.
- Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., & Hemanth, J. (2021). SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustainable Cities and Society*.
- Perez Cayetano, I., Jimenez Preciado, J., Antonio Velazquez, J., & Lopez Gonzalez, E. (2021). Detección automática de rostros con cubreboca o sin cubreboca para restringir el acceso a institución educativa. *CI2T 2021 Congreso Internacional de Investigación Tijuana 2021*, 2-7.
- Yuzchen, C., Hu, M., Hua, C., Guangtao, Z., Jian, Z., & Quinli, L. (2021). Face Mask Assistant: Detection of Face Mask Service Stage Based on Mobile Phone. *IEEE Sensors Journal*, 11084-11093.
- Zhang, K., Zhang, Z., Li, Z., & Qiua, Y. (2023). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*.