# Acta de Correcciones al Proyecto de Grado
# Ingeniería Electrónica

**Fecha:** sábado, 13 de mayo de 2023

**Autores:** Juan Camilo Cordoba Bravo, Jose Ivan Torres Ordoñez

**Nombre del Proyecto de Grado:** DEVELOP OF PROTOTYPE SYSTEM FOR PEOPLE RECOGNITION BASED ON EAR BIOMETRICS

**Director:** Diego Luis Linares Ospina

Como indica el artículo 2.27 de las Directrices de Trabajo de Grado, he verificado que los estudiantes indicados arriba han implementado todas las correcciones que los Jurados del Proyecto de Grado definieron que se efectuaran, como consta en el Acta de Calificación correspondiente.

________________________________________

Firma de Director del Proyecto de Grado

Nota de Aceptación

Aprobado por el Comité de Trabajo de Grado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana para optar el título de Ingeniero Electrónico.

Camilo Rocha

**Dr. CAMILO ROCHA**
Decano de la Facultad de Ingeniería

**DR. LUIS EDUARDO TOBON LLANO**
Director Carrera Ingeniería Electrónica.

**DR. DIEGO LUIS LINARES OSPINA**
Director Trabajo

**DR. HERNAN DARIO VARGAS CARDONA**
Jurado 1

Hernan Dario Benitez Restrepo

**DR. HERNAN DARIO BENITEZ RESTREPO**
Jurado 2

Santiago de Cali, 13 de mayo, 2023.

Señores
Pontificia Universidad Javeriana Cali.
Dr. Luis Eduardo Tobón Llano
Director Carrera de Ingeniería Electrónica.
Cali, Valle del Cauca.

Cordial Saludo.

Por medio del presente me permito informarle que los estudiantes José Iván Torres Ordoñez con código de identificación No. 8934715 y Juan Camilo Córdoba Bravo con código de identificación No. 8935544 trabajaron bajo mi dirección en el proyecto de grado titulado "Develop of prototype system for people recognition based on ear biometrics" el cual se encuentra corregido y listo para ser presentado.

Atentamente,

Dr. Diego Luis Linares Ospina

# DEVELOP OF PROTOTYPE SYSTEM FOR PEOPLE RECOGNITION BASED ON EAR BIOMETRICS

Degree work to opt for the title of:

ELECTRONIC ENGINEER


Juan Camilo Córdoba Bravo

Jose Ivan Torres Ordoñez


Supervised by: Diego Luis Linares Ospina PhD.


Pontificia Universidad Javeriana

School of Engineering and Computer Sciences

Career of Electronic Engineering


Santiago de Cali, May 12, 2023.

# Content

# 1 Abstract

In this work, a prototype of an ear biometric system based on Convolutional Neural Networks is designed and evaluated, thinking about the problems faced by people who cannot use the conventional fingerprint biometric system in Colombia. First, with the OpenCV Haar Cascade tool, the user's ear is extracted and a database of ninety-two users is generated, using data augmentation for later training. The characteristic of CNNs to extract features in their convolutional layers are used and transfer learning is performed with a Support Vector Machine as classifier that has the extracted CNN features as input. The CNN models used were VGG16 and FaceNet. A retraining of the VGG16 model available in Keras library was made, this model was retrained with images of ears so that it learns to extract its features. The FaceNet model developed by Google is used on its base form to get the features. These features are input to a C-SVM classifier, the SVM hyperparameters are adjusted with Sklearn Grid-Search technique, the CNN models use different SVM hyperparameters. Python scripts are developed to implement the proposed models, such as user enrollment, classifier training and the use of the proposed system. After having the algorithms ready, tests were made to evaluate their performance with different techniques such as Sklearn cross-validation to figure out the accuracy of the models, the False accept rate and False reject rate metrics, and finally the ROC curve for biometric systems to get the performance of this prototype system.

# 2 Introduction

Nowadays the identification and authentication of people have become a very important aspect for society, being the area of biometrics applied to the unequivocal recognition of people the field that has contributed the most to this objective since it uses unique and non-transferable features of individuals; This area has acquired great importance in aspects such as security when accessing facilities, using financial services, carrying out legal procedures or identifying yourself with public or private entities [1]. Currently, the best known, and most used biometric identification feature is the fingerprint due to its high reliability, speed, and practicality in identifying people [2].

In Colombia, the fingerprint identification method has been adopted in both state and private entities, this being the only mechanism available to perform authentication and access services or perform procedures that require it. In this sense, when there are failures in the reading of the fingerprint, there can be inconveniences for people that can range from having to do more procedures to validate their identity to the case of not being able to access the services that require biometric authentication; these consequences are much worse for those who have permanently lost their mark [3].

These failures in recognition by these systems can be caused by different factors that can be genetic diseases such as dermatitis, aging, and heavy work [4]. According to statistical studies conducted by the Universidad San Sebastián, it was shown that age is a variable that affects fingerprints, and the minimum age at which the fingerprint fades is 63 years for women and 66 years for men. [4]. In [3] a national survey was conducted in the Country of Lebanon in 2013 where the index of the absence of fingerprints was 0.18%. To solve this problem, a backup system for biometric identification must be devised, designed, and implemented to replace the fingerprint system.

Among the existing types of biometrics, there are several options to support the fingerprint biometrics system, such as scanning the face, hand, iris, ear, and voice [5]. Taking into account the current context and the restrictions that are presented such as the health situation and aspects of privacy in the collection of personal data, ease of use and convenience for people, we found that the best alternative for the identification of people is the use of biometrics of the ear since it has characteristics such as not being significantly affected by age, does not require contact with any surface for the respective identification and is not affected by the use of masks, glasses, beard, among others [6].

This problem and its context lead to important questions for the development of this work such as: How to automatically identify biometric characteristics of the ear? How to use these characteristics for the identification of people? and how to develop a biometric alternative system of authentication using ear biometrics? Answering these questions will guide the development of this work.

# 3 Objectives

## 3.1 General Objective

Develop and evaluate a prototype ear-based biometric authentication system using Convolutional Neural Networks that functions as an alternative to a fingerprint biometric system.

## 3.2 Specific Objectives

1. Correctly obtain 2d images of the ear to create a data set of 100±10 individuals.
2. Process images to improve the exposure of patterns necessary for the identification of persons.
3. Design algorithms for biometric ear recognition.
4. Analyze the performance of the models.

# 4 Scope

- Functional system prototype developed in python for ear biometric recognition based on Convolutional Neural Networks.

- Model accuracy rate must exceed 80%.

- Testing Database of 100±10 people data.

# 5 Justification

The ability to identify and differentiate one person from another is fundamental for modern society; for this purpose, there exist many techniques such as biometrics that use unique physical characteristics such as fingerprint, iris, face, ear, etcetera.[5]. In Colombia, the most widely used authentication method is the fingerprint. Although this method is widely accepted and easy to use [2], there are drawbacks for people who have anomalies in their fingerprints; these can be caused by diseases, hard work, or aging [3] causing that these people cannot access easily to different services. In this sense, the implementation of a multimodal identification system in offices for procedures or money transfers will provide inclusion and ease of access to services for people who cannot authenticate with their fingerprints.

Considering that in Colombia the use of the fingerprint for authentication is the only one that has a database regulated by the National Civil Registry with the Automatic Fingerprint Identification System (AFIS) [1]. It is increasingly common to use this method for people to verify their identity at public or private entities such as banks, money transfer offices, among others by using a fingerprint reader and accessing services in those entities. Therefore, when a person cannot be identified by the fingerprint, access is denied or more procedures must be performed to access services such as sending and receiving money, opening bank accounts, subsidies, passport procedures, among others [7].

On the other hand, a survey in a Lebanese population with a sample of 145,600 people shown that 0.18% had no fingerprints, with a predominance in the population over 65 years of age [3]. Also, a survey in Chile of 487 people between 55 and 107 years of age found that 48.25% had no fingerprints [4]. These surveys show that there are a considerable number of people who cannot use the fingerprint biometrics method and that older adults are the most likely to have these difficulties. This is a serious drawback in Colombia given that most of these older people are beneficiaries of government subsidies and because of their age it is more difficult for them to do these procedures to access them.

To sum up, it becomes clear the need to implement a supporting alternative to the fingerprint biometrics system with a focus on processing offices or money transfer offices, since people who have problems with their fingerprint also need to access many of its services and their condition makes it more complicated or impedes it. In addition, a multimodal system will reinforce security, accelerate, and ease processes both for customers as well as for the entities in question. Considering the scope and actors involved, the implementation of this project will have an important impact on the social context, since it will provide adequate conditions to carry out procedures for many people, most of them elderly, as well as improve the efficiency of the procedures at the targeted establishments when these cases occur.

# 6 Theoretical Framework

In the area of biometric recognition, given the characteristics of the problem posed such as the focus on identification of older adults, as well as the restrictions caused by the Covid-19 and privacy aspects in the collection of personal data, we found that the biometric recognition of the ear is the main candidate to support the fingerprint system in case of failure [6]. A review of the different theoretical bases and previous works related to ear biometry and techniques used for the identification of people that are relevant to the development of this project.

## 6.1 Digital Image Processing

Among the most important theoretical aspects and techniques we have the digital image processing which consists of treating an image in such a way that allows obtaining relevant information; in the case of ear recognition, we find that this is a very important stage to improve the performance of the identification system, since it allows adjusting the images obtained, making the conversion to grayscale and then obtaining the edges of the ear that are important for further analysis. In [6], the authors propose the use of these techniques for image preprocessing to obtain the most relevant features as a first phase for ear recognition and identification. Additionally, in [8] the author uses these techniques to determine relevant descriptors for human hands from the image capture to the characteristics identification using digital image processing.

## 6.2 Information Security

In the degree work of Sanchez J [9] of the Universidad Nacional Abierta y a Distancia (UNAD) studies and analyzes the computer security of several types of biometrics, and the reliability of each of them, something to highlight is the emphasis on the fact that fingerprint biometrics can be easily forged by capturing the image of the finger impregnated on a surface, also the focus on the biometrics of the ear which is easy to capture and its characteristics remain until old age. The computer security analysis presented in this degree work supports the development of this work and the security of the biometric method is an important aspect to consider.

## 6.3 Voronoi Diagrams and Canny Edge Detector

A tool that can be useful for the automation of the ear biometric system is the Voronoi Diagrams in conjunction with the Canny Edge Detector to extract the ear features [10]. Voronoi diagrams are the decomposition of a metric space into regions, associated to the presence of objects, each object has assigned a region of the metric space, formed by the points that are closer to it than to any of the other objects. These diagrams fulfill the following properties: The set of points equidistant from two given points is a line, The set of points equidistant from three or more given points is a point (if it exists). This tool helps to divide the ear image into sections allowing the extraction of image key features.

Accordingly, another useful tool for ear characterization is the Canny Algorithm, which acts as an edge detector on an image; it is one of the best contour detection methods. The basic operation of this algorithm is based on the first derivative, and it is used to identify the regions of the image where the intensity of the grayscale changes, where there is no change of intensity the first derivative is zero and where the intensity changes the first derivative acquires a constant value, Therefore a change of intensity manifests itself as a sharp change in the first derivative, a feature that is used to detect an edge, and on which the Canny algorithm is based [11].

## 6.4 Convolutional Neural Networks

Another important aspect in ear recognition is the use of convolutional neural networks (CNNs) using machine learning algorithms that allow training a recognition model from sets of previously labeled images and then performing the respective testing of the model to determine its effectiveness in the assigned task. In the case of ear recognition in [6] we have that CNNs aim to obtain key features such as ear edges; as well as, by applying multiple convolution layers, it allows to obtain useful features such as color, gradient, among others.

Additionally, in [12] a review of different CNNs models such as AlexNet, VGGNet, Inception and ResNet is carried out and the use of transfer learning is proposed, which consists of using the knowledge acquired by a deep CNN in a specific task with the objective of feeding the knowledge of another Network that has a similar related task. In this sense, the authors propose this method to scale ear recognition to images obtained under unconstrained conditions; the efficiency of recognition in this environment is crucial to achieve favorable results.

## 6.5 Ear Images Datasets

A key factor for this project's development is the available databases, which consist of ear images. In this sense, we find different databases that are suitable for this research and that allow both the respective training of algorithms in the case of CNNs, and the verification of the results obtained by the generated models. Here, we find databases such as Annotated Web Ears (AWE) Dataset which consists of images of celebrities obtained on the web which were cropped around the ear; additionally, the Ear Dataset from Kaggle [13] The Unconstrained Ear Recognition Challenge (UERC) dataset which was generated for a competition in ear biometrics with 2017 and 2019 versions and the EarVN1.0 dataset is one of the largest datasets recently released and has ear images collected under unconstrained conditions such as UERC [12], [14]. Finally, the AMI Ear database created by Esther Gonzales has ear images from one hundred subjects on a uniform and controlled environment [15].

## 6.6 OpenCV

OpenCV (Open-Source Computer Vision Library) is a machine vision and machine learning library. This library has more than 2500 algorithms optimized for computer vision and artificial intelligence applications [16].

## 6.7 Haar Cascade Detector

The Haar Cascade Detector is an object detection algorithm that uses Haar features to identify objects in images. These features are based on the idea that objects in an image have different levels of light intensity and shadows, and that these differences can be used to identify objects. It is commonly used to detect people's faces, but it also finds applications in ear detection [17]

Haar features can be displayed as black and white rectangles superimposed on an image. Each Haar feature is composed of two or more rectangles with different intensities of light and shadow, which overlap to create a pattern. There are different types of Haar features, including simple Haar features, such as horizontal and vertical lines, and complex Haar features, which include more elaborate patterns with different shapes and sizes of rectangles. With this, the Haar Cascade algorithm evaluates each Haar feature in an image by segmenting cascaded regions and determines if the pattern corresponds to a specific object [17]

## 6.8 Data Augmentation

Data augmentation is a technique used in the field of machine learning when there is insufficient data to train a model. In the context of images as data, there are algorithms that make modifications in the image, creating modified copies of the original image. Even in facial recognition work, algorithms can generate new faces [18]. The Keras open-source library provides a tool called Image Data Generator that generates a copy of the modified input image. The main functions used in this work are zoom_range; generates an image randomly larger or smaller depending on the input, brightness_range; generates an image with random change of brightness, rotation range; generates an image with random change of angle or rotation. Using all these functions, a large batch of images can be generated for training a Deep Learning model.

## 6.9 Machine Learning libraries

There are open-source libraries that facilitate the development and research of new Machine Learning applications. Among them are Tensor Flow, Keras and Sklearn. These libraries include pre-trained machine learning models, tools to implement them and guidance documentation to make use of them. Tensor Flow and Keras work together to help make machine learning experiments easier to perform [19], [20]. Sklearn has useful tools for the use of models used in this work such as: creating validation and training sets, evaluating trained models, calculating their accuracy, and calculating ROC curves [21].

## 6.10 Transfer Learning

Training a new Convolutional Neural Network model requires a large amount of data, and a lot of time for training. With the transfer learning technique, it is possible to use a previously trained model and use it in another similar application without the need to train it from zero. If the model was trained on a similar application, then the outputs of the deeper layers can be used directly in the new application. [21]. This makes it possible to use part of the previously trained model as a feature extractor and to train the following classification layers.

## 6.11 FaceNet Model

FaceNet is a deep learning neural network developed by Google in 2015 [22]. Specializes in face recognition, using a transfer learning architecture to learn how to perform face recognition tasks on high-quality images. The neural network is trained on a large set of face images to produce a numerical representation of each face, called embedding. These embeddings are then used to compare images of faces to figure out if they belong to the same person or not.


***Figure 1. FaceNet Architecture [22]***

“The architecture used by FaceNet consists of a batch standardization, with an architecture based on Zeiler&Fergus [23] style networks and the recent Inception [24] type networks” [22]

This model is complemented by the Triplet loss function which consists of using a triplet composed of an anchor and a positive sample which correspond to the same identity together with a negative sample which corresponds to a different identity. This loss function makes the differences between samples of the same identity smaller, while the differences between samples of different identities larger as is shown on the **Figure 2**. [22]

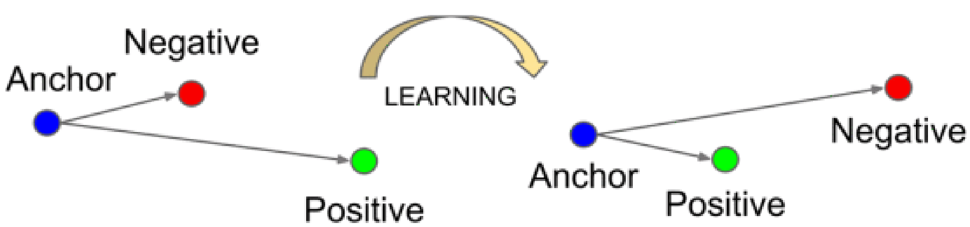



***Figure 2. FaceNet's Triplets loss function illustration [22]***

In the case of ear-based biometric recognition, FaceNet could be adapted to focus on the identification of ears instead of faces. This would be achieved through a training and validation process, in which images of ears would be used instead of images of faces. The neural network would then learn to generate embeddings for the ears and compare them to figure out whether they belong to the same person.

## 6.12 Support Vector Machines

A Support Vector Machine (SMV) is a linear sorting algorithm (but has the advantage of solving both linear and nonlinear problems), has taken on important relevance in many applications such as computer vision, character recognition, text and hypertext categorization, protein classification, natural language processing, time series analysis [25]. The starting point is a Dataset X, Y and the objective is to learn the relationship between X and Y, X are the data to be classified and Y are the labels or tags of the data [26]. Data entering the SMV are mapped into the original space of the input examples, if these are linearly separable, or into a transformed space if the data are not linearly separable [25]. The transformed space is a space of one dimension greater than that of the data, if the input data is from R2 then it is mapped to an R3 space, this space is called a feature map. [27]. In the feature space SMV calculates and creates hyperplanes that separate and maximize the margin between classes. To define the hyperplane that separates the classes, SMV only considers the training data of each class that fall just on the border of these margins. [21]. These few points used for hyperplane formation are called Support Vectors.

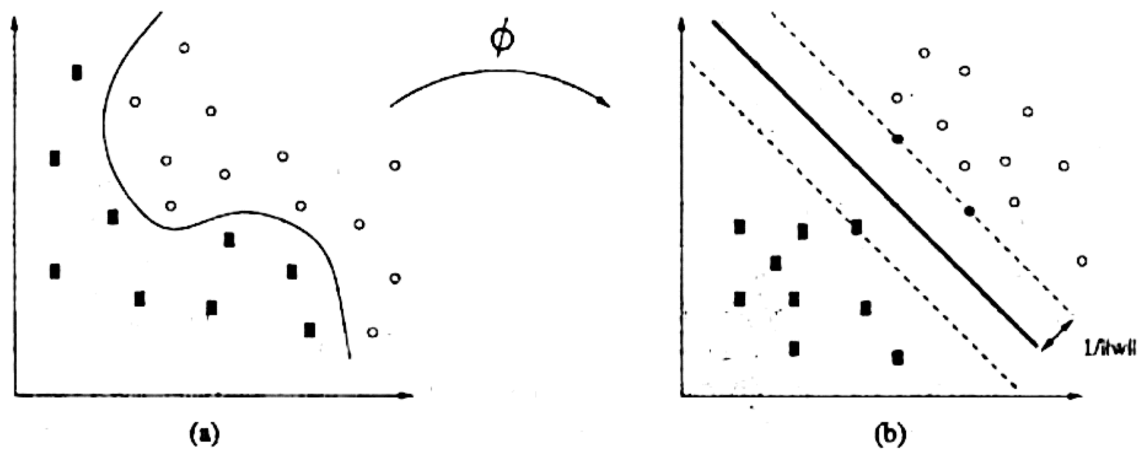



***Figure 3. (a) Original data in the input space. (b) Mapped in the feature space.***
***Source: Support Vector Machines with Applications [25]***

# 7 Development of the prototype of the Biometric Ear Recognition System

This section will discuss the entire process conducted for the development of the prototype for a biometric ear recognition system. All the aspects necessary for the preparation of the data, its processing, the implementation of CNN deep learning models, the development of the scripts necessary for the system execution, and the evaluation of the performance obtained with this prototype system will be considered.

## 7.1 Correctly obtain 2d images of the ear to create a data set of 100±10 individuals.

The acquisition of a significant database of ears of different people is needed. In this work we have a scope of 100±10 users that compose an adequate volume of data that is needed to develop and evaluate the system effectively. For this, it is necessary to find and specify the necessary criteria to obtain the data; then prepare a controlled environment to capture the images, obtain the images of the ears and finally generate the database of the users.

### 7.1.1 Identification of criteria for image capture

It is necessary to ask now: which ear will be used for this system? According to [28] about 90% of the right and left ears of people are symmetrical, this means that it would be possible to use either ear, but it is not 100% certain, since about 10% of the ears are not symmetrical. Therefore, in the system only one ear will be used as standard, the left ear. But depending on the user it could change.

To obtain an adequate image of the user's ear, it is necessary to establish the different requirements to achieve a good quality in the data, which is an especially important element for this work. These criteria include:

### 7.1.2 Image capture device:

In this section, the use of a Logitech C920 webcam was chosen for this prototype because it has very favorable characteristics for this work in terms of image capture and versatility, of which the following stand out:

**High quality resolution:** This camera has a 1080p resolution, which means that the images captured will have high quality and detail, which is especially important when working with fine details such as those found in the ear.

**Automatic focus:** It also has an autofocus that automatically adjusts the focus to keep the image clear and sharp. This is important when working with small objects such as ears.

**Compatibility:** This model is compatible with a wide variety of operating systems and software, which makes it easily adaptable to different needs and configurations for this work. Additionally, it makes it very versatile for this work since it is compatible with the software tools used in the development process.

**Accessibility:** It is also an affordable camera compared to other high-quality options on the market, which makes it an excellent choice for a prototype.

### 7.1.3 Environment for capturing images.

In the development of this work, a first approach was made to take the images of people´s profile and extract from them the respective ear by recording a video in which the people made movements of the head to generate variations in the angle of capture of the ear. Subsequently, the extraction of the ear image was performed with a script that uses OpenCV and a Haar Cascade detector that generates the boxes containing the ear to be subsequently cropped and stored.

However, the method used evidenced flaws due to the presence of elements in the background of the image since the images were captured in the wild. These elements were detected by the script and resulted in a lot of unwanted data. **Figure 4** shows some of these detection errors.

***Figure 4. Examples of Haar Cascade detection errors***

Considering these errors in the detection of the ear, it is concluded that this approach is not suitable for an automatic operation of a biometric system, however, 92 videos of people were captured, which are useful to create the proposed database.

To achieve an automatic and fast operation (not requiring human intervention), the characteristics of a controlled environment in which the biometric system captures the photo and extracts the user's ear without the faults presented above are established. The following specifications are defined:

**Background:** It is necessary to avoid elements different from the user’s head to improve the automatic detection of the ear, so using a background is the best option. In this case a black background is used.

**Capture distance:** Since the size of the ear is small, it is also necessary to specify the distance from which the image will be captured, in this case 20cm allows to obtain a decent quality image without being uncomfortable for the capture.

**Illumination:** Another important aspect is to have good illumination in the environment of this system, for the case of this prototype a light source located in the room’s ceiling is used.

Once these requirements are specified, the next step is to prepare the environment for the image capture. The Logitech C920 camera is connected to the computer for data storage and processing. A black canvas is also used for the background and the shooting distance is adjusted manually at first. Subsequently, a tripod is used to fix the camera with the possibility of adjusting it to the user's head height.

***Figure 5. Image capture platform***

### 7.1.4 Image capture and database generation

With the capture platform, the user's photo and ear are obtained. This is processed depending on whether the user is being registered or using the system (see section 7.3.3). The image capture process consists of the use of a Python script using OpenCV that displays the camera visualization on the screen and allows the user to capture the image by pressing the g key. The script enters the captured image into the Haar Cascade detector and the box that has the ear is cropped. With this, the original image is displayed on the screen as well as the ear obtained or a warning in case no ear has been detected. With this, it is possible to see if the image obtained is adequate. Then, the options to save the capture or discard it are enabled.

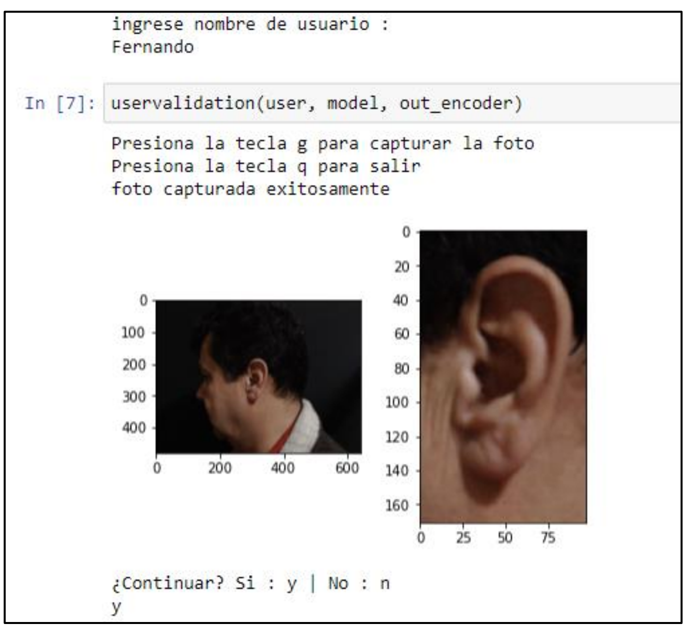



***Figure 6. Capturing Images***

This script developed for the biometric system data collection is also used to process the 92 captured videos. For each user video, 5 different ear images are captured and a database of 92 people's ears is formed (Annex 11.2). This is an especially important part of this work since it allows the development of the next stages and the testing of the operation of this prototype recognition system.

## 7.2 Process images to improve the exposure of patterns necessary for the identification of persons.

It is necessary to ensure that the system correctly identifies the users when they use it. For this purpose, the data to be used for the development of the system must be preprocessed.

### 7.2.1 Adjusting the ear image to be used on a CNN model.

In this work we make use of CNN models that process images to extract their features, each CNN model with this functionality has different requirements for its input image. In this work two CNN models are used, and the following image processing is done:

**Color configuration:** Depending on the libraries used to process the image it can change if it is RGB or BGR, in this work we used the OpenCV library to manage the images, so the image is loaded in BGR, and it is necessary to convert it to RGB.

**Image scaling:** Image scaling is to adjust the dimension of an image to a specific size in pixels, it is common to find that a CNN model requires the image of a certain dimension. The OpenCV library functions and the Keras library with nearest interpolation are used for this purpose.

**Scaling pixel values:** This involves converting the pixel values of an image into a specific range, such as 0 to 1 or -1 to 1, to improve the accuracy and performance of machine learning models. In this paper we convert pixel values to float32.

**Normalize pixel values in all channels:** Normalizing the pixels of an image is useful in machine learning tasks, such as image classification, since it helps to reduce the variance and improve the accuracy of the model. In this work we perform the Z-score transformation which adjusts to a common scale with zero mean and unit standard deviation.

### 7.2.2 Data augmentation

Thanks to the proposed controlled environment and the data capture script it is possible to propose an automatic and more effective user registration compared to the first approach made (user video). For this it is necessary the Data Augmentation which allows to simulate the possible changes in the user's posture in front of the camera and to have more images to train the model. The user then only must position himself adequately for a single photo with the conditions described in the previous sections.

The data augmentation performed in this work consists of generating random variations of rotation, brightness, zoom in and zoom out. The tool used for this is Image Data Generator from Keras. Examples of these variations are shown in **Figure 7**.

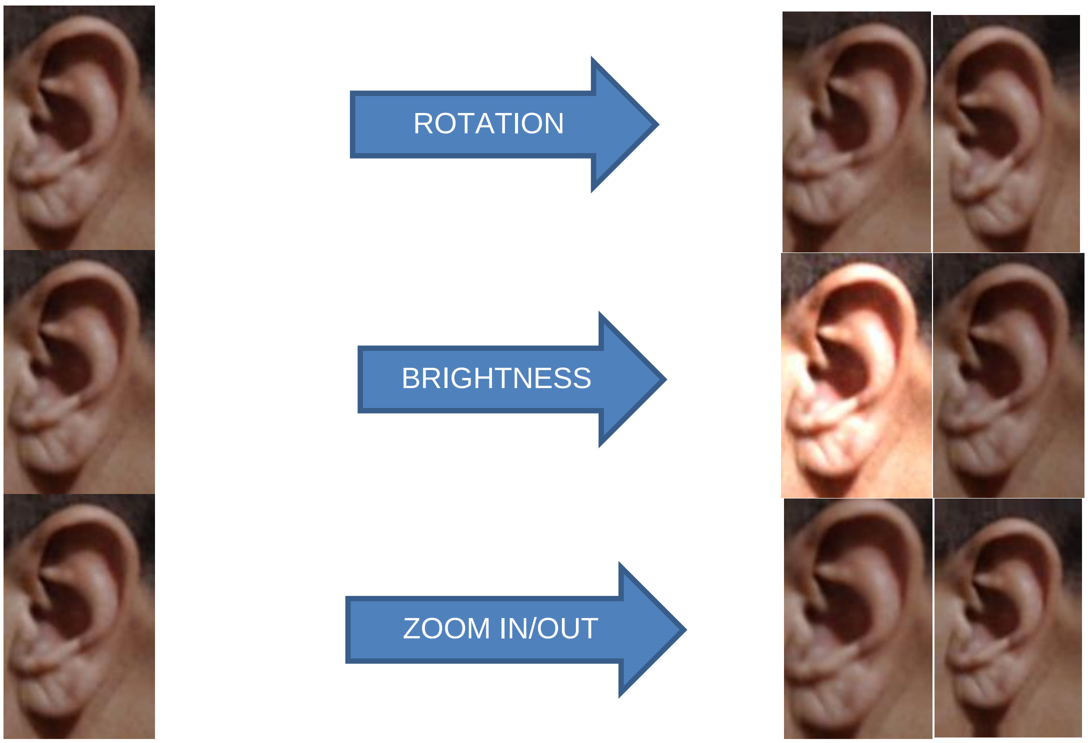



***Figure 7. Examples of data augmentation***

This data augmentation process is only used in the user registration stage, since it is necessary for the designed system to have the capacity to classify them adequately in their validation. The modifications made are presented below:

**Rotation:** This modification simulates the tilt of the user's head. The argument of rotation_range receives as input an integer in degrees for random rotations. In this work we perform data augmentation with change in random rotation up to 15 degrees in both directions.

**Brightness:** This modification takes into account possible differences in ambient lighting. The brightness_range argument receives a list of two float-type numbers that are a range of brightness variance, a value of 1.0 is brightness neutral, if it is less than 1.0 the image is darkened and if it is greater than 1.0 the image will be brightened. In this work a range of 0.6 to 1.4 is used.

**Zoom in/out:** This modification takes into account possible changes in the photo taking distance due to bad positioning of the user by randomly moving the ear closer or further away. The zoom_range argument receives a list of two floating numbers which is a range for zooming in or out, if the range is less than one it zooms the ear in and if it is greater than one it zooms the ear out. In this work we use 0.94 to 0.86 for zoom in and 1.02 to 1.11 for zoom out.

This data augmentation process generates 120 ears from one ear of the user, all with random changes as presented below.

## 7.3 Algorithm design for biometric ear identification.

This section deals with the most important topic for this work which is how to use the ear to extract unique features and use them to make the person recognition system by using ear biometrics. In this sense, feature extraction is examined using two CNN models which are VGG16 and FaceNet. Subsequently, a classifier for user recognition based on the features obtained is developed.

### 7.3.1 Feature extraction using CNN models.

To differentiate ears, it is necessary that the machine learns their unique characteristics, for this purpose the feature extraction function of Convolutional Neural Networks is used. Thanks to the convolutional layers, the CNN generates a Feature Map (FM from now on), a vector with numerical elements that represent the characteristics of the input object, in this case, the image of an ear.

The CNN uses the FM as input to its hidden layers stage to classify among all its categories, but in this work, they will serve as input to the SVM classifier. Therefore, the FM is calculated inside a CNN. With the Keras library, the last hidden layers can be removed, and the convolutional layers can be exposed. The retraining of the last classification layers is called transfer learning. Here we examine the feature extraction process using the CNN models VGG16 and FaceNet.

#### 7.3.1.1 VGG16 Model

To use the VGG16 model, the following input criteria are specified for image preprocessing: Input size of 224 x 224 pixels and BGR format. This preprocessing is performed using the Keras library, which performs the respective normalization and standardization of the data to meet the requirements of this model.

The model is then retrained to learn how to extract ear features and provide an FM that can be subsequently used by the classifier. For this purpose, the AMI Ear Database [14] containing 700 ear images is used. The training is performed with the following hyperparameters:

**Optimizer:** Adam optimizer is used since it is a popular choice for neural network training due to its efficiency and good performance on a variety of deep learning problems, and its default parameter settings work well in most applications [29].

**Loss function:** The categorical crossentropy loss function is used since it is suitable for multi-class classification problems, such as the multiple user's ear recognition of this work.

**Epochs:** ten epochs were used to achieve good results without overfitting the data used for training.

The result of this training is presented in **Figure 8** where the summary of training accuracy and loss is plotted on the following learning curves.

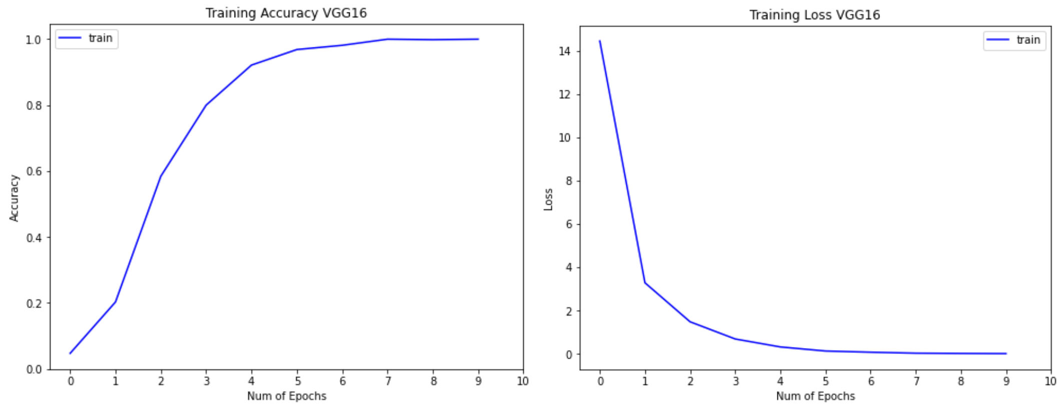



***Figure 8. Learning curves VGG16***

As a result, the model has as input the image of the ear and returns a one-dimensional vector of one hundred numerical float32 elements corresponding to the feature vector representation of the input image.

#### 7.3.1.2 FaceNet Model

To make use of the FaceNet model it is necessary to make the following adjustments to the image: RGB color format, resize to 160x160, convert pixels to float32, perform Z-score transformation with mean and standard deviation.

For this model, considering the characteristics of the FaceNet architecture and the characteristic of the Triplet Loss function, it is determined to use the feature vector or embedding generated by the model in its base form and enter it to the classifier and thus perform the learning transfer.

The model returns a one-dimensional vector of 128 numerical float32 elements corresponding to the embeddings that will be used by the classifier.

### 7.3.2 Design of ear classifier model based on the extracted features.

#### 7.3.2.1 Classifier training

As a classifier, a Classification Support Vector Machine (C-SVM) is used. This classifier must be able to identify which user corresponds to an ear in its input, so it is trained with the embeddings provided by the CNN model. With the increase of data for each user, 120 embeddings are created, of which 80% were taken for training and 20% for validation. In terms of data, out of 120 user ears, the classifier is trained with ninety ears and the remaining thirty ears are used to validate the classifier. An example of the format of the classification set entering the SVM is as follows:

X_train = [[Embedding1], [Embedding2], [Embedding3], [Embedding4], [Embedding5], [Embedding6]]
Y_train = [ [user1], [user1], [user1], [user2], [user2], [user2] ]

To split the data into training and validation, it is possible to build a function by hand or use the Sklearn library that returns a training vector and a validation vector.

#### 7.3.2.2 Hyperparameter Optimization

An SVM for classification has several types and hyperparameters that depend on the data with which it is used. Hyperparameters are parameters that the model does not learn automatically during training, instead they must be entered in its definition. These hyperparameters modify the model and influence the prediction, so it is important to set the right hyperparameters.

To find the proper hyperparameters a technique called Grid Search of Sklearn is used, with this, a dictionary with the parameters is entered and the technique trains models with the hyperparameters, using Cross Validation calculates its accuracy and delivers the parameters that had the best accuracy.

The hyperparameters used in this work are:

**Gamma:** Kernel coefficient for 'rbf', 'poly' and 'sigmoid'. The gamma parameter controls the flexibility of the SVM model and its ability to fit the training data.

**C:** Regulation parameter. The strength of the regularization is inversely proportional to C. It must be strictly positive. The penalty is a penalty of l2 squared. determines the amount of classification error allowed in the model.

**Coef0:** Independent term in the kernel function. Only significant in 'poly' and 'sigmoid'. It is used to adjust the relative importance of higher degree coefficients in the kernel function.

**Kernel Linear:** Linear SVM, used when the data can be separated linearly and has no significant nonlinear features.

**Kernel Poly:** Polynomial kernel, used when the data are not linearly separable, and a polynomial decision boundary is required.

**Kernel RBF:** It is used when the shape of the decision boundary is not known. This kernel uses a radial distance function to transform the data to a high-dimensional feature space, where a nonlinear decision boundary can be found.

**Kernel Sigmoid:** This kernel is used when the data is expected to have a sigmoid shape.

### Hyperparameter optimization for classifier with FaceNet

Evaluated hyperparameters:

| Kernel | C | Gamma | Coef0 |
|---|---|---|---|
| Linear | [0.3, 0.7, 1, 2, 5, 10, 50, 100] | --- | --- |
| Poly | | [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 5] | [-1, -0.5, 0.0, 0.5, 1, 1.5, 2, 4, 6] |
| RBF | | | --- |
| Sigmoid | | | [-1, -0.5, 0.0, 0.5, 1, 1.5, 2, 4, 6] |

***Table 1. Hyperparameters evaluated for SVM with FaceNet.***

Best hyperparameters found:

| Kernel | C | Accuracy |
|---|---|---|
| Linear | 5 | 100.00% |

***Table 2. Best hyperparameters for SVM with FaceNet.***

### Hyperparameter optimization for classifier with VGG16

Evaluated hyperparameters:

| Kernel | C | Gamma | Coef0 |
|---|---|---|---|
| Linear | [0.3, 0.7, 1, 2, 5, 10, 50, 100] | --- | --- |
| Poly | | [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 5] | [-1, -0.5, 0.0, 0.5, 1, 1.5, 2, 4, 6] |
| RBF | | | --- |
| Sigmoid | | | [-1, -0.5, 0.0, 0.5, 1, 1.5, 2, 4, 6] |

***Table 3. Hyperparameters evaluated for SVM with VGG16.***

Best hyperparameters found:

| Kernel | C | Gamma | Accuracy |
|---|---|---|---|
| RBF | 5 | 2 | 42.40% |

***Table 4. Best hyperparameters for SVM with VGG16***

## 7.3.3 Develop algorithms for biometric ear enrollment and recognition.

With the models and database ready, it is necessary to create scripts to make use of it, Python is used for this. The basis of a biometric system is the identification of registered users, so the biometric system needs an algorithm to role users, this algorithm will store the user data necessary for subsequent identification, and the algorithm that makes use of the data to make decisions: allow the access to registered users and deny unregistered users. Based on this, two algorithms are developed that satisfy the bases of the biometric system explained: user validation algorithm and user identification algorithm.

### 7.3.3.1 User validation algorithm

The identity of the person is confirmed by comparing the user's ear with a set of "negative" example ears, i.e., ears of people who are not registered and should not have access. So, the classifier will only have two options: YES: Belongs to the set of user ears, and NO: Belongs to the set of negative ears. This method generates an independent model for each user. The method consists of two stages: training and validation.

**Training stage:**

This script is created for user registration, every time a new user registers this algorithm is executed and the model trained for that user is saved to disk. **Figure 9** shows a block diagram of the algorithm. On the left side of the block diagram enters the user data, ear extraction, data augmentation and image processing, on the right side enters a set of 'negative' ear images which are preprocessed, the set of negative images was made using the data sets [14] [13]. Both sets of user and negative ears go through the CNN model for feature extraction in embeddings.

An SVM model is trained with the hyperparameters found in the previous section with the two sets of embeddings. This trained SVM is saved and can discriminate whether an ear belongs to the user or not. As there is a different model for each user, it is saved on disk with one code by user.

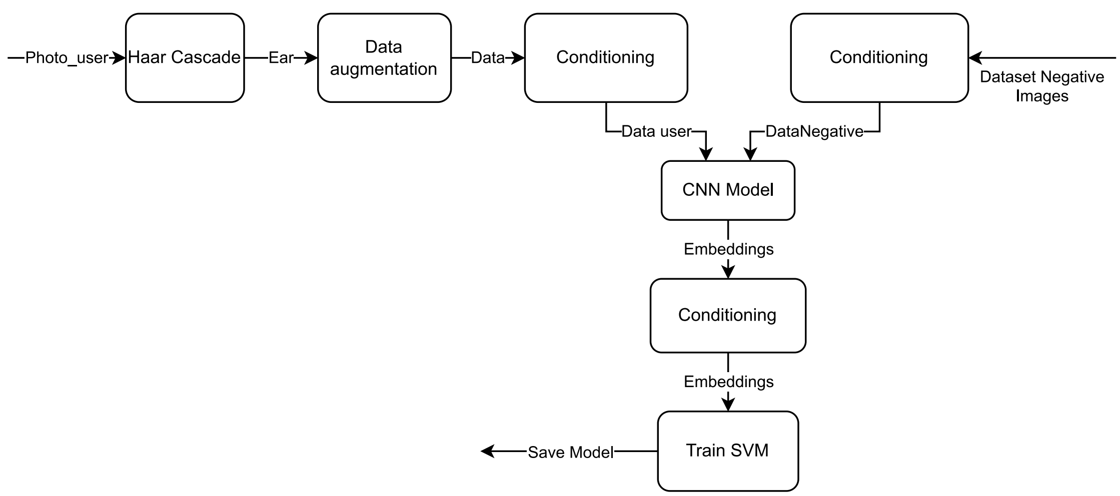



*Figure 9. User validation algorithm. Training stage*

In the block diagram of Figure 9 there are blocks called "conditioning", in these blocks the data is transformed to be used in the next block. In the conditioning stage before the CNN model the transformations are made to the images explained previously in section 7.3.1. Another data conditioning is made to the embeddings before entering the SVM, here the numerical values of the embeddings are normalized with the L2 norm.

**Validation Stage**

This script uses the trained models when a user needs to validate. The input is the user code, name, and photo. The output is whether it is the user or not. The block diagram is shown in **Figure 10**.

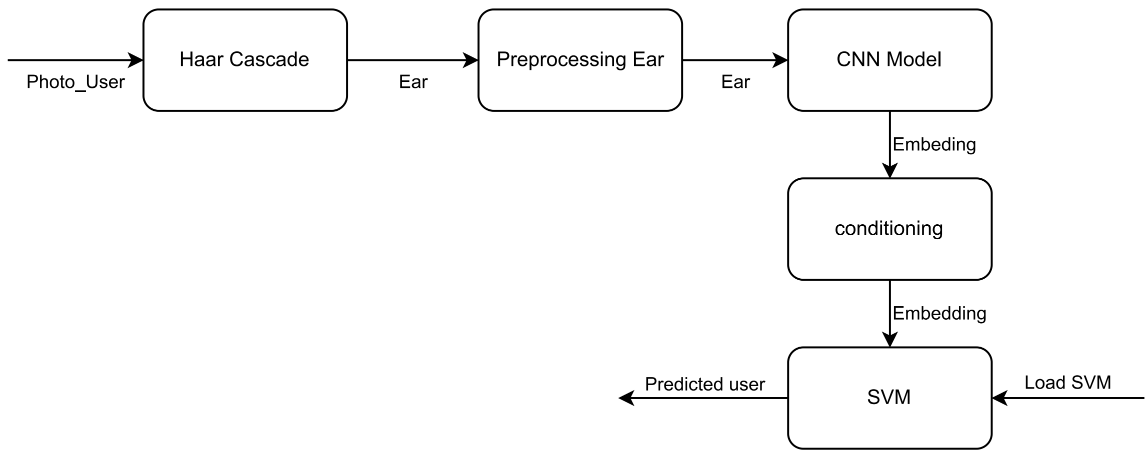



***Figure 10. User validation algorithm. Validation Stage***

To use the system, first the user's photo is captured, the respective classifier model is loaded with the entered data, the embedding of the ear to be predicted is obtained, conditioned, and used in the classifier model to validate or not the user.

### 7.3.3.2 User identification algorithm

In this identification algorithm the system returns a prediction of the name and probability of belonging to one of the registered users, in this work ninety-two users are registered for testing. The approach of this identification algorithm is designed to register a batch of users at the same time. After the embeddings of all users are obtained, the classifier is trained and stored on disk to be subsequently used. This method has three stages: Registration, Training and Validation.

**Registration stage**

This registration stage is where the user's embeddings are obtained and saved to disk. The algorithm begins extracting the user's ear image samples, performs data augmentation, preprocesses the ear images, obtains the embeddings of the user's ears, and saves them to disk, as shown in the block diagram in **Figure 11**. The saved user data is used in the next stage.

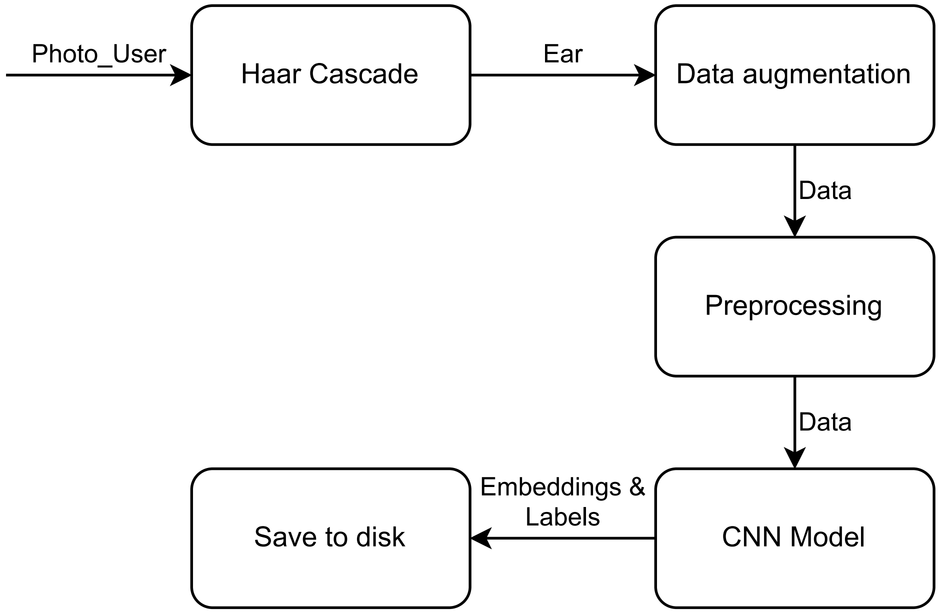



***Figure 11. User identification algorithm. Registration stage***

## Training stage

At this stage, the user embeddings stored on disk are joined and used for training the classifier SVM. Its block diagram is shown in **Figure 12.**

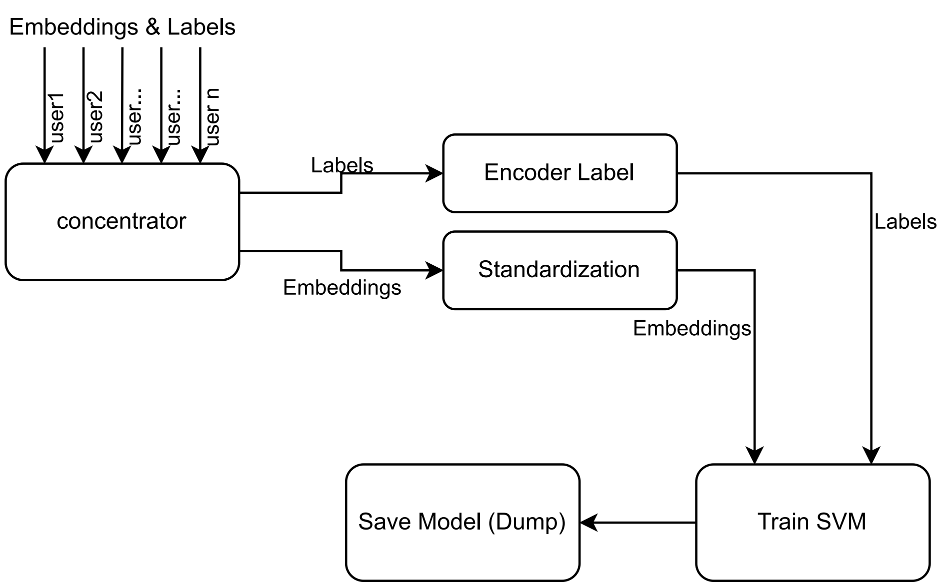



***Figure 12. User identification algorithm. Training stage***

From the union of all saved embeddings, an array of embeddings and of labels is created, the embeddings are normalized, and the labels are encoded to train the SVM. The classifier model already trained with all the registered users is saved on disk for use in the next stage.

**Validation Stage**

This is the stage where the classifier model already trained with the registered users is used. The only difference between this script and the validation method script is that in this one a classifier model is used for a batch of registered users, while in the previous method a classifier model is used for a single user. The block diagram of the algorithm is shown in **Figure 13**.

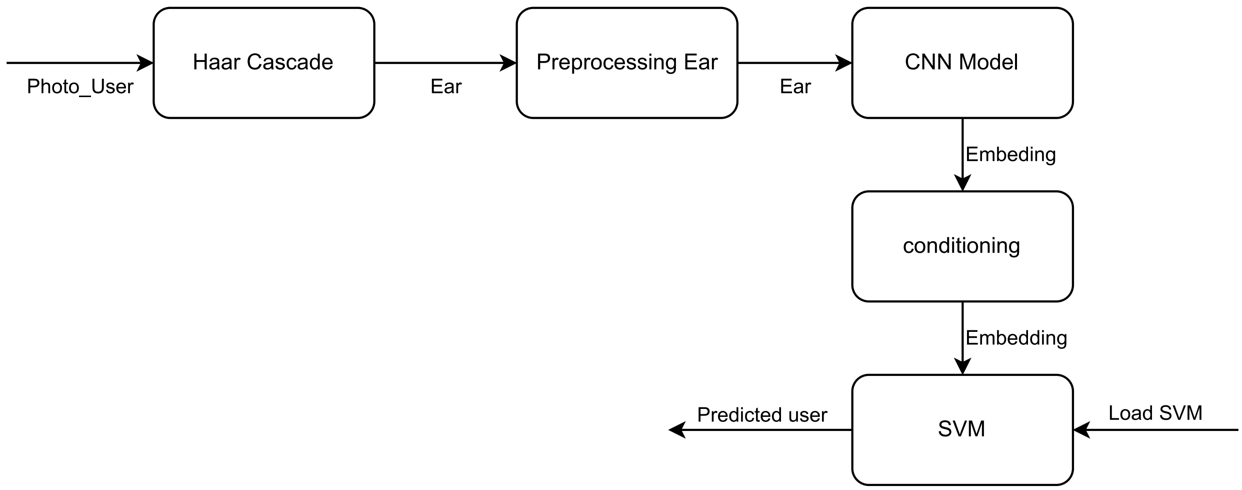



***Figure 13. User identification algorithm. Validation Stage***

The input is the user's photo and name, the output returned by the classifier model is the predicted name of the user, and the belonging to class probability; this prediction must match with enough probability to the expected user for a successful identification of this person.

# 8 Analyze the performance of the models.

This section presents the evaluation of the system considering the two models used based on different metrics to figure out the performance for these models and for the system in general. With this, a comparison is made between the two CNN models used to determine the best one.

## 8.1 Accuracy

To find the accuracy of a model, the Sklearn Cross Validation technique is used. This technique uses the entire data set to train and validate the model through iterations. The technique partitions the dataset, and, in each iteration, a different partition is used as the test set and the rest of the partitions to train the model. Thus, in each iteration the technique gives an accuracy value of the model. The average of these results is the model accuracy.

To evaluate the classifier trained with embeddings, the full set of user data, i.e., 120 images per person, is used. Five partitions are used for the technique to train and validate the classifier five times. The results were as follows:

### 8.1.1 System based on VGG16.

| ACCURACY | **AVERAGE** | **STANDARD DEVIATION** |
|---|---|---|
| METHOD OF IDENTIFICATION | 42.40% | 1% |

***Table 5. System accuracy based on VGG16.***

This accuracy value clearly does not satisfy the requirement of greater than 80% accuracy for the scope of this work.

### 8.1.2 System based on FaceNet.

| METHOD | **AVERAGE** | **STANDARD DEVIATION** |
|---|---|---|
| IDENTIFICATION | 100% | 0% |
| VERIFICATION | 100% | 0% |

***Table 6. System accuracy based on FaceNet.***

The Cross Validation technique gives 100% accuracy values for both the user identification and verification methods. This means that the system based on the FaceNet model can correctly identify all registered users.

## 8.2 False Acceptance Rate (FAR) and False Rejection Rate (FRR)

The performance of the biometric system is then evaluated based on the False Acceptance Rate (FAR) and False Rejection Rate (FRR) depending on the sensitivity of the system. This sensitivity is set up as a threshold in the probability delivered by the classifier from which a user is recognized as legitimate. This sensitivity defines the security and reliability of the system since, if the threshold is low, the system will validate more people and there will be errors in which non-legitimate people access as if they were legitimate (FAR); and in the opposite case, if the threshold is very high, the system will be more selective and even legitimate people may not be recognized (FRR). Therefore, to achieve the best system performance, these indicators must be as low as possible. Therefore, an analysis is performed using 213 samples between legitimate users and non-legitimate users to obtain the measure of errors of each type for the two systems evaluated.

### 8.2.1 System based on VGG16.

The results obtained by the recognition system based on VGG16 are shown in **Figure 14**, where an unsatisfactory performance can be evidenced since the system must have a threshold higher than 98% to eliminate false acceptance errors, which in turn causes the false rejection rate to be close to 100%, i.e., it could not recognize users properly. It can also be shown that with a threshold of 13% the system could recognize users, but it would not be a secure system since the false accepted rate reaches 52% which means a vulnerability on the system since non-legitimate people could even access as if they are legitimate by mistake. This yields the system based on VGG16 to have a bad performance.

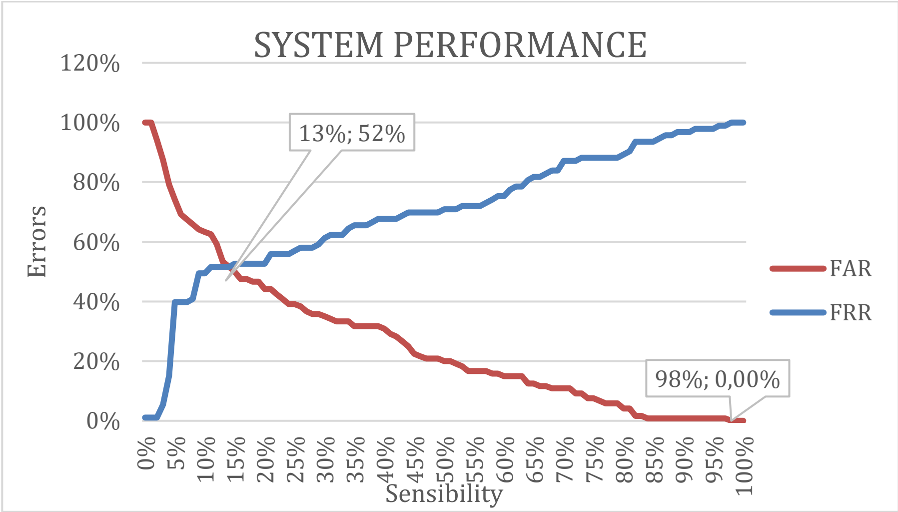



*Figure 14. System performance based on FAR and FRR for VGG16*

### 8.2.2 System based on FaceNet.

The same test is performed with the FaceNet-based recognition system and the results are shown in **Figure 15**. In this case it is possible to notice a big difference in the performance results with respect to the previous system.

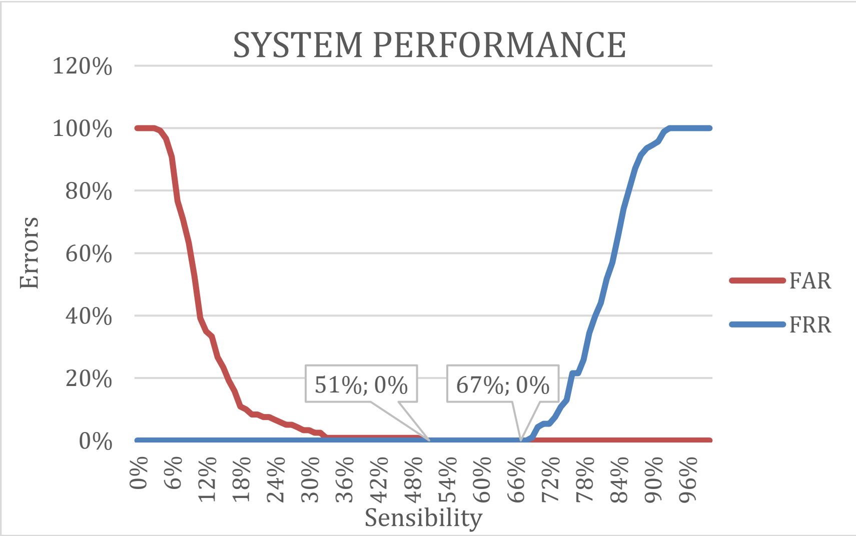



*Figure 15. System performance based on FAR AND FRR for FaceNet*

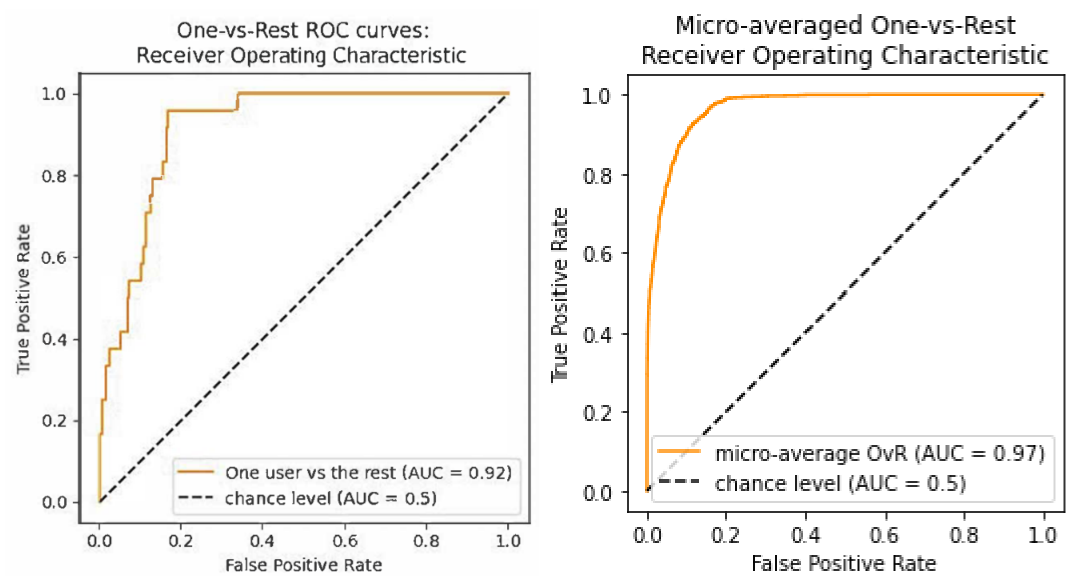
Based on the graph, the FaceNet-based recognition system performance is ideal when the probability threshold is between 50% and 68%, since the obtained values of errors (FAR and FRR) in this range are equal to 0. This means that the system can correctly identify users and prevent non-legitimate users from being taken as legitimate. This makes the system much more secure and robust for the task of biometric recognition of people.

## 8.3 ROC curve

The ROC curve is a good way to evaluate the performance of a biometric system [30] In the graph, the higher on the Y-axis the higher the True Positive Rate, and on the left X-axis the lower the False Positive Rate.

### 8.3.1 System based on VGG16.

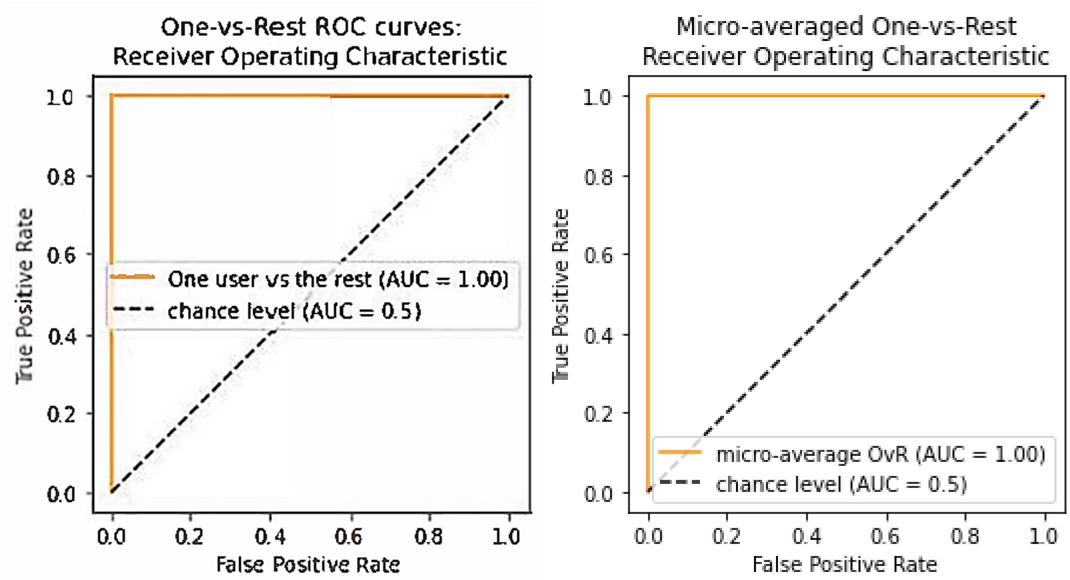
**Figure 16**(a) shows the ROC curve of one user vs. the rest, and in (b) an average ROC curve of all classes or users in this case. In the upper left corner, the curve is away from the corner which means that the model has a considerable number of false positives, the AUC value is the area under the curve, an area under the curve (AUC) value greater than 0.9 is generally good, but for the requirements of this biometric system where the system is expected to be able to recognize all registered users is not so good.




a) b)

***Figure 16. ROC curve for system based on VGG16. a) One user vs. the rest b) ROC averaged over all users.***

### 8.3.2 System based on FaceNet.




a) b)

***Figure 17. ROC curve for FaceNet-based system. a) One user vs. the rest b) ROC averaged over all users.***

**Figure 17** shows that the ROC curve is close to the upper left corner, where false positives are minimal and true positives are maximal. This means that there are no wrong identifications by the biometric system designed with this model and classifier.

## 8.4 Best Recognition System

In this work experiments with two CNN models, VGG16 and FaceNet were made, both models deliver an embedding of similar format, the VGG16 model was retrained with images of ears and FaceNet was used in its base form. With the embedding provided by each model, a classifier for person recognition based on SVM was trained and the results of the evaluation of these models were obtained using metrics such as: Accuracy, FAR, FRR, and ROC Curve. These results showed a clear superiority of the FaceNet model over the VGG16 model in all the tests performed.

The CNN FaceNet model is robust, trained with a lot of data and although it is specialized in obtaining face features it has proven to be highly effective in delivering ear embeddings that allow ear recognition in a correct and secure way. On the other hand, the VGG16 model retrained with ear images did not perform well in this task. Among the causes of this is that the amount of training data was not enough to obtain an accurate model, and another is that originally the VGG16 model was trained with the ImageNet database [20] which is very different from the application performed in this work which consists of ears.

Considering the above, it is determined that the prototype system developed will be based on the FaceNet model to perform biometric ear recognition.

## 8.5 Analysis of the strengths and weaknesses of the proposed system

**Strengths:**

- **User privacy:** The proposed system does not store personal information, not even the user's own ear, the system stores the user's ear in a coded way as embedding. To identify a user, it is necessary, the embedding, model and several lines of code that condition the embedding.

- **Model accuracy:** According to the results obtained by the model evaluation with cross-validation, the system can validate the user correctly almost 100% of the time. This gives reliability to the proposed system.

- **Non-invasive:** What is needed to use the proposed system is just a profile picture of the user, this takes no more than a minute. Once the photo is taken, the system takes care of the rest.

- **Security:** In the traditional fingerprint biometrics method, it is possible to breach security using trickery by copying another person's fingerprint. In the proposed method, considering also that it is not yet used, it is very difficult to copy an ear of another person, since the complete form is offered, neither could an image of the correct ear be able to be in front of the capturing machine since it would be easily seen by the system operator.

**Weaknesses:**

- **The Haar Cascade algorithm is not 100% accurate:** Depending on the user it may not detect his ear correctly. This can be corrected by changing the Haar Cascade parameter. For this reason, the proposed system has a confirmation statement, where the operator checks if the ear is well detected, if is correct, the system continues with the rest.

- **Waiting time:** Although tests performed with the FaceNet-based system show very satisfactory results, the model takes 6 seconds to calculate an embedding for one ear, if the number of ears to be processed increases, this time accumulates. VGG16 does it in about one second, but it does not have enough performance. The SVM model for classifying between users used in both cases takes no more than one second to deliver its prediction.

- **Computational capacity:** Especially in the case of registering several people at the same time, the calculation of the embeddings can take several minutes (depending on the number of users) and that needs computers with enough processing capabilities.

# 9 Conclusions and future work

## 9.1 General conclusions and answers to the stated objectives

- The performance of the proposed biometric system prototype based on FaceNet convolutional neural network and linear SVM represented by the Accuracy, ROC curve, FAR, and FRR metrics shows that it´s a secure and trustful way to perform the task of people recognition based on ear biometrics. Since it does not accept illegitimate users as does not reject legitimate users. This yields a particularly good prototype that meets the objective of this work.

- To ensure that the images of the ears have an adequate quality to capture their unique characteristics, a controlled environment was implemented where the distance from the camera to the ear is 20cm, the background is black and has sufficient brightness.

- The implementation of a controlled environment makes the results and the performance of the developed system to be particularly good and that shows the importance of standardizing the data captured on this kind of system that requires high security.

- The procedure to capture the ear from the image was enabled using the Haar Cascade Detector that is an extremely useful tool to obtain the data for the dataset and to get the ear image for the system application. Additionally, since the ear detection algorithm is not completely accurate, for some users the parameters of the Haar Cascade detector need to be changed to detect the ear correctly.

- In this work, an ear database was created in a semi-controlled environment, containing 5 ear images per person, with a total of 92 people. The ear images were captured directly from the person's profile. This database was very useful for the development and evaluation of the performance of the proposed system. This is also very important since this information can be useful for further developments in this field of research.

- The performance of a VGG16 model retrained with ear imaging was evaluated, the results of the system with this model are not satisfactory and do not achieve the requirement of greater than 80% accuracy for this application.

- With the FaceNet model, calculating an embedding takes about 6 seconds, and validating a user takes about 7 seconds. The waiting time to be validated is acceptable since only one embedding of the candidate user needs to be computed for use, but the problem lies in registering the user. The model was trained with more than a hundred images per user, which translates into more than 8 minutes to register a user.

- Although the method of comparing the user's ear with a set of negative ears gave good results, it has limitations, as negative ears were included in the training. Depending on the case, the examples of negative ears might not be sufficient and result in a false positive.

## 9.2 Future research directions

- Future work can reduce the waiting time to register by calculating how many images are sufficient to train the model without decreasing its performance. It is also possible to search for or train a more efficient specialized ear model than the ones used in this work.

- Create a more reliable ear detection algorithm that can recognize the ear of any user and does not recognize non-ear objects.

- Link the user database to an internet storage, so as not to fill up local disks and to have options to restore data in case of loss.

- Even the scope of this work is to provide a biometric alternative for people who do not have fingerprints or lost them; this approach on ear biometric systems can be scaled to more scenarios and populations.

- The database constructed in this work is limited in the population variety, in future works to improve the performance and scalability of the system it is necessary to train it with a database with population variety.

- Design and build the photo capture structure so that the camera is located at the precise distance, and easily adjusted to the user's height, as well as having the proposed black background.

# 10 References


[1] "Identificación biométrica: cada vez con más usos en la vida cotidiana- Registraduría Nacional del Estado Civil." https://www.registraduria.gov.co/Identificacion-biometrica-cada-vez.html (accessed Aug. 18, 2021).

[2] M. Osvaldo and A. Cuello Videla, "Adermatoglifia: Una mutación genética que impide la formación de huellas dactilares Adermatoglyphia: a genetic mutation that prevents the formation of fingerprints," 2016. [Online]. Available: www.somecrimnl.es.tl66Fechaderecepción:11/08/2015Fechadeaceptación:25/11/2015

[3] N. Sarfraz, "Adermatoglyphia: Barriers to Biometric Identification and the Need for a Standardized Alternative," *Cureus*, Feb. 2019, doi: 10.7759/CUREUS.4040.

[4] J. Silva *et al.*, "Pérdida del Dactilograma Durante la Senilidad," *International Journal of Morphology*, vol. 36, no. 1, pp. 258–261, Mar. 2018, doi: 10.4067/S0717-95022018000100258.

[5] M. Gutiérrez, Á. María, C. Peñuela, and A. Camilo, "Panorama de la biometría de huellas dactilares en Colombia", Universidad Piloto de Colombia, Bogotá, Colombia, 2013.

[6] S. Nikose and H. K. Meena, "Ear-biometrics for human identification," *Proceedings - 2020 Advanced Computing and Communication Technologies for High Performance Applications, ACCTHPA 2020*, pp. 8–13, Jul. 2020, doi: 10.1109/ACCTHPA49271.2020.9213190.

[7] A. Mancera (2013, feb 27). "Entidades públicas usarán las huellas para documentación." https://www.asuntoslegales.com.co/actualidad/entidades-publicas-usaran-las-huellas-para-documentacion-2032975 (accessed Sep. 01, 2021).

[8] N. Fernández Valencia and I. Mecatrónico, "Identificación de las características biométricas de la mano humana mediante visión por computador para el diseño de un prototipo," 2015, Accessed: Oct. 12, 2021. [Online]. Available: http://repositorio.utp.edu.co/dspace/handle/11059/5635

[9] J. J. Sánchez, "Biometría y la seguridad informática en los métodos de autenticación. - 10596/39060," *UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA – UNAD*, 2020. https://repository.unad.edu.co/handle/10596/39060 (accessed Oct. 12, 2021).

[10] M. Burge and W. Burger, "Ear biometrics in computer vision," *Proceedings - International Conference on Pattern Recognition*, vol. 15, no. 2, pp. 822–826, 2000, doi: 10.1109/ICPR.2000.906202.

[11] Valverde-Rebaza, Jorge. (2007). "Detección de bordes mediante el algoritmo de Canny." https://www.researchgate.net/publication/267240432_Deteccion_de_bordes_mediante_el_algoritmo_de_Canny (accessed Sep. 22, 2021).

[12] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometrics Recognition Using Deep Learning: A Survey," Nov. 2019, Accessed: Sep. 22, 2021. [Online]. Available: http://arxiv.org/abs/1912.00271

[13] R. Bajaj, “Ear Dataset | Kaggle.” https://www.kaggle.com/datasets/bajajra/ear-dataset (accessed Mar. 29, 2023).

[14] V. Truong Hoang, “EarVN1.0,” vol. 4, 2020, doi: 10.17632/YWS3V3MWX3.4.

[15] E. Gonzales, L. Alvarez, L. Mazorra. “AMI Ear Database.”. Centro de I+D de Tecnologias de la Imagen Universidad de Las Palmas de G.C. https://ctim.ulpgc.es/research_works/ami_ear_database/ (accessed Feb. 14, 2023).

[16] “OpenCV.” https://opencv.org/ (accessed Feb. 14, 2023).

[17] P. Viola and M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features,” 2001.

[18] X. Wang, K. Wang, and S. Lian, “A survey on face data augmentation for the training of deep neural networks,” *Neural Comput Appl*, vol. 32, no. 19, pp. 15503–15531, Oct. 2020, doi: 10.1007/S00521-020-04748-3.

[19] “TensorFlow.” https://www.tensorflow.org/?hl=es-419 (accessed Feb. 14, 2023).

[20] “Keras: the Python deep learning API.” https://keras.io/ (accessed Feb. 14, 2023).

[21] J. Brownlee, “Deep Learning for Computer Vision. Image Classification, Object Detection and Face Recognition in Python,”, v 1.4. 2019.

[22] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 815–823, Mar. 2015, doi: 10.1109/CVPR.2015.7298682.

[23] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8689 LNCS, no. PART 1, pp. 818–833, Nov. 2013, doi: 10.48550/arxiv.1311.2901.

[24] C. Szegedy *et al.*, “Going Deeper with Convolutions,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 1–9, Sep. 2014, doi: 10.48550/arxiv.1409.4842.

[25] E. J. Carmona, “Tutorial sobre Máquinas de Vectores Soporte (SVM)”, Accessed: Feb. 14, 2023. [Online]. Available: https://www.researchgate.net/publication/263817587

[26] J. M. Moguerza and A. Muñoz, “Support Vector Machines with Applications,” *https://doi.org/10.1214/088342306000000493*, vol. 21, no. 3, pp. 322–336, Aug. 2006, doi: 10.1214/088342306000000493.

[27] G. A. BETANCOURT, “LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs),” *Scientia et Technica*, vol. 1, no. 27, Apr. 2005, doi: 10.22517/23447214.6895.

[28] P. Yan and K. W. Bowyer, “Empirical evaluation of advanced ear biometrics,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2005-September, 2005, doi: 10.1109/CVPR.2005.450.

[29] P. L. Lagari, L. H. Tsoukalas, and I. E. Lagaris, “Variance Counterbalancing for Stochastic Large-scale Learning,” *International Journal on Artificial Intelligence Tools*, vol. 29, no. 5, Aug. 2020, doi: 10.1142/S0218213020500104.

[30] A. Jain, P. Flynn, A. Ross. “Handbook of Biometrics,” *Handbook of Biometrics*, Springer, 2008, doi: 10.1007/978-0-387-71041-9.

# 11 Attachments

## 11.1 GitHub repository with the code of this work:

https://github.com/J0S3dj3/Jose-Ivan.git

## 11.2 Generated Ear Images Dataset

https://drive.google.com/drive/folders/1kPnnPPjYd7fOa-ADqzUicUaQVFRkHwry?usp=sharing