Develop of Prototype System for People Recognition Based on Ear Biometrics

J.C. Córdoba Bravo[1], J.I. Torres Ordoñez[1*]
[1]*Pontificia Universidad Javeriana Cali*
c99@javerianacali.edu.co, joseivan17@javerianacali.edu.co[*]

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.

I. Introduction

In Colombia, the fingerprint identification method is the most widely used, being almost the only mechanism available for authentication. In this sense, when there are failures in the reading of the fingerprint, there may be inconveniences for people such as having to carry out more procedures or even not being able to access services that require biometric authentication; these consequences are much worse for those who have definitively lost their fingerprint. To solve the need of people who suffer from this problem, it is necessary to have an alternative method to fingerprint biometrics, which shares the qualities of being fast, secure, and reliable. Among the existing types of biometrics, there are several options that can serve as a backup, such as face, hand, iris, ear, and voice scanning. Taking into account restrictions such as health status and privacy aspects in the collection of personal data, ease of use and convenience for individuals, it was found that the best alternative for the identification of individuals is the use of ear biometrics since it has characteristics such as not being significantly affected by age, does not require contact with any surface and is not affected by the use of masks, glasses, beard, among others. According to the information presented, the main objective of this article is: Develop and evaluate a prototype ear-based biometric authentication system using Convolutional Neural Networks that functions as an alternative to a fingerprint biometric system.

II. Theoretical foundation

1) Convolutional Neural Networks (CNN): They are machine learning models based on Neural Networks. CNNs are especially useful for applications with images thanks to their convolutional layers since this layer allows extracting unique features from images such as lines, curves, and shapes. Figure 1 shows a basic architecture of a CNN. A CNN model can be trained from sets of previously labeled images, and it adjusts its weights on the neurons and their kernels to learn to extract features from the set of images. In the case of ear recognition in [1] we have that CNNs aim to obtain key features such as the edges of the ears; as well as, by applying multiple convolution layers, it allows to obtain useful features such as color, gradient, among others.

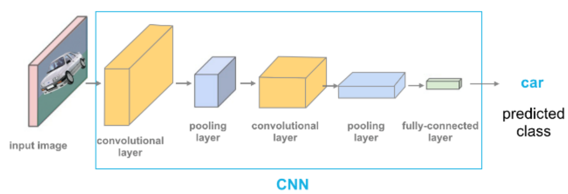

Figure 1. CNN Architecture [2]

Additionally, in [3] 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.

2) Ear Images Datasets: Ear databases available on the Internet are helpful to train models and validate their performance. In this sense, databases such as Kaggle's Ear Dataset [4] containing ear images of 164 people are used, and the AMI Ear database created by Esther Gonzales has ear images of one hundred subjects in a uniform and controlled environment [5].

3) 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 [6]. 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 [6]. This algorithm delivers the coordinates of the image where the required object is which can be extracted from the image, as shown in Figure 2.

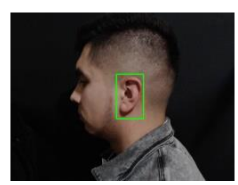
Figure 2.Haar Cascade image

4) 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. [7]. This makes it possible to use part of the previously trained model as a feature extractor and to train the following classification layers.

5) FaceNet Model: It is a deep learning convolutional neural network developed by Google in 2015 [8]. The neural network is trained on a large set of face images to produce a numerical representation of each face, called an embedding. These embeddings are then used to compare face images and find out whether they belong to the same person or not.

Figure 3. FaceNet Architecture [8]

“The architecture used by FaceNet consists of a batch standardization, with an architecture based on Zeiler&Fergus [9] style networks and the recent Inception [10] type networks” [8]. 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. In the case of ear-based biometric recognition, transfer learning can be used to use FaceNet as a feature extractor and train the classifier. This is possible because the two applications are similar in that they both deal with human skin on the head.

6) Support Vector Machines (SVM): 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 [11]. 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. 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. 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. [12]. 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. These few points used for hyperplane formation are called Support Vectors.

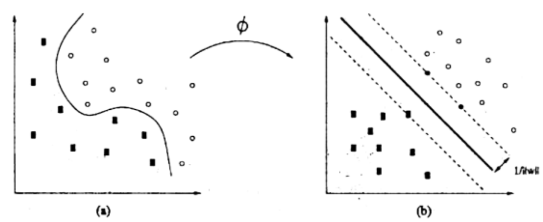

Figure 4. (a) Original data in the input space. (b) Mapped in the feature space [11].

III. Results

The CNN FaceNet and VGG 16 models are used as feature extractors. A Support Vector Machine (SVM) model is trained as a classifier. The SVM hyperparameters are calculated for each CNN model using Sklearn's Grid Search technique:

- Hyperparameters for SVM models:

Table 1. Hyperparameters with FaceNet embeddings.

Kernel	**C**	Accuracy
Linear	5	100.0%

Table 2. SVM Hyperparameters with VGG 16 embeddings.

Kernel	**C**	**Gamma**	Accuracy
RBF	5	2	42.40%

Since the FaceNet model is specialized in delivering embeddings and trained with faces that are similar to ears, it is used without other training. The VGG 16 model is retrained with the AMI database [5]. The results are presented with various metrics to measure its performance as a biometric system:

1) Accuracy: The Sklearn cross-validation technique is used. This technique partitions the entire data set into equal parts, in each iteration, a different partition is used as a test set and the rest to train the model. Thus, at each iteration the technique gives an accuracy value of the model. The average of these results is the model accuracy. For this, 120 images per user are used, for 92 users. Five partitions of the set are used in Sklearn's Cross Validation technique. The following results were obtained:

Table 3. System accuracy based on VGG16.

Accuracy	**Average**	**Standard Deviation**
Method of Identification	42.40%	1%

Table 4. System accuracy based on FaceNet.

Method	**Average**	**Standard Deviation**
Identification	100%	0%
Verification	100%	0%

2) False Acceptance Rate (FAR) and False Rejection Rate (FRR): These metrics depend on the probability of belonging of a prediction to a class, called here as sensitivity. It is important because, if the threshold is low, there will be errors in which non-legitimate people access as if they were legitimate (FAR); and if the threshold is very high, the system will be more selective and even legitimate people may not be recognized (FRR). These two indicators should be as low as possible. Therefore, an analysis is performed using 213 samples between legitimate and non-legitimate users to obtain the measure of errors of each type for the two systems evaluated.

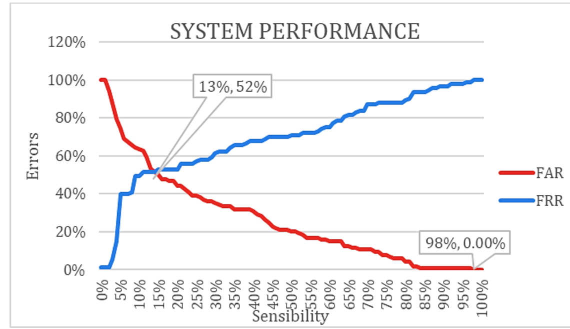

Figure 5. System performance based on FAR and FRR for VGG16

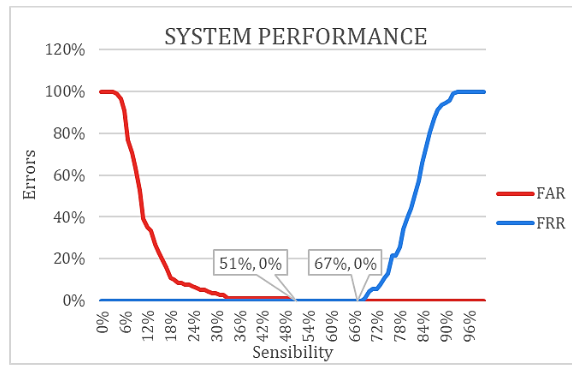

Figure 6. System performance based on FAR and FRR for FaceNet

In Figure 5 for the system based on VGG 16 the curves show that the best sensitivity value is 13%, with a FRR and FAR value of 52%. In Figure 6 for FaceNet based system the best sensitivity is between 51% and 67% where FAR and FRR are zero.

3) ROC curve: In the graph, the higher on the Y-axis, the higher the true positive rate, and on the X-axis the further to the left, the lower the false positive rate.

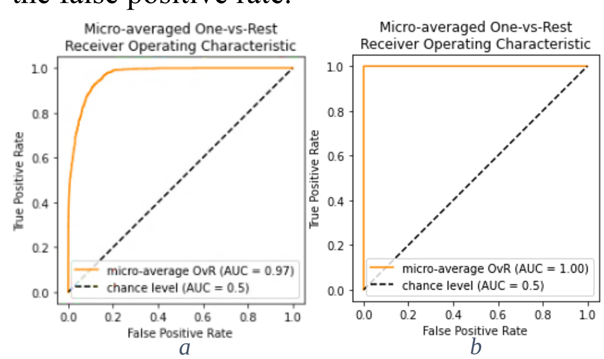

Figure 7. ROC Curves a) System based on VGG 16. b) System based on FaceNet

IV. Discussion and conclusions

The biometric system composed by the CNN FaceNet model and the linear SVM presents an ideal performance. Thanks to the techniques used to evaluate it, it is known that it is able to identify each user correctly, it doesn't pass unregistered people as users or users as unregistered, and it doesn't have false positives. However, FaceNet takes 6 seconds to calculate an embedding, which is fine for user validation, but this time is accumulated when a user registers, since 120 images are processed per user. To improve this, it is necessary to investigate how many samples per user can be reduced without losing accuracy or to train a more efficient model. On the other hand, using the VGG 16 model and SVM with RBF the results are not acceptable for a biometric system, since approximately it cannot identify 50% of the users and this error grows as the number of users increases. This poor performance has two possible causes, the first is that a retraining of this model was performed with insufficient training data (700 images), the second is that VGG 16 was originally trained with the ImageNet database [13] which is very different from this ear application.

With this, it is concluded that the main objective set for this article is achieved and making the respective improvements the best option is to use the FaceNet and SVM linear model for the identification of people using ear biometrics so that people who suffer from anomalies in their fingerprint can access services and procedures more easily.

V. References

[1] 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.

[2] Camacho, C. (2018). Convolutional Neural Networks – Machine and deep learning educator. Available: https://cezannec. github. io / Convolutional _ Neural_ Networks/.

[3] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, D. Zhang, "Biometrics Recognition Using Deep Learning: A Survey," Nov. 2019.

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

[5] E. Gonzales, L. Alvarez, L. Mazorra. "AMI Ear Database.". Universidad de las Palmas, Available: https://ctim.ulpgc.es/ research_ Works /ami_ear_ database/ (accessed Feb. 14, 2023).

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

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

[8] F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07-12-June-2015, Mar. 2015, doi: 10.1109/CVPR.2015.7298682.

[9] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," Lecture Notes in Computer, vol. 8689 doi: 10.48550/arxiv.1311.2901.

[10] 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.

[11] E. J. Carmona, "Tutorial sobre Máquinas de Vectores Soporte (SVM)", Universidad Nacional de Educación a Distancia, Nov. 2016.

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

[13] "Keras: The Python deep learning API." https://keras.io/ (accessed Feb. 14, 2023).