Using an Accurate Multimodal Biometric for Human Identification System via Deep Learning

Saba A.Tuama[1]
sabaayad91@gmail.com

Jamila.H.Saud[2]
dr.jameelahharbi@gmail.com

Zahraa A. Alobaidy[3]
zahraaalobaidy@gmail.com

Abstract: Biometric systems for automated identification of an individual rely on behavioral or physiological variables linked with the individual. Biometric systems function in two modes: verification and identification. In the verification mode, a claimed identity is either denied or accepted, and in the identification mode, the identity of an unknown person is described. Multibiometric systems are used to establish an individual's identification by combining information supplied by several biometric sensors, samples, units, algorithms, or features. Multibiometric is an interesting and exciting research topic. It is used to identify people to improve security. Therefore, these systems are intended to prevent spoofing, facilitate continuous monitoring, enhance population coverage, and provide fault tolerance to biometric applications. This study proposes an identification system for the individual based on the ear and tongue pattern. Convolution Neural Network (CNN) extracts the essential features from the input images. This system is robust to noise, brightness variations and insensitive to rotation variation. The proposed method consists of four main stages (i.e., pre-processing, fusion, feature extraction, and finally, classification stage). The proposed method was tested on real datasets and achieved an average accuracy of 99.72% for all datasets.

Keywords: Multibiometric; Identification System; Ear Biometric; Convolution Neural Network; Tongue Pattern

[1] Dr.: Department of Computer Technology Engineering, Al-Hikma University College, Baghdad, Iraq
[2] Prof. Dr.: Department of Computer Science, College of Science, Al-Mustansiriyah University, Baghdad, Iraq
[3] Assist. Lec.: Department of Computer Technology Engineering, Al-Hikma University College, Baghdad, Iraq

1. Introduction

Biometric identification technology has been introduced thereby utilizing the behavioral or physical traits to identity a person [1]. Physiological biometrics include facial [2], hand and hand vein [3], odor, ear [4], fingerprint [5], retina [6], and DNA [7], while behavioral biometrics include gait [8], keystroke [9], signature [10] as shown in Figure 1. Any human behavioral or physiological trait can serve as a biometric characteristic so long as it satisfies the following requirements [11]:

1. Universality: Each individual should have the feature.
2. Distinctiveness: Any two individuals should be different.
3. Permanence: The feature should be sufficiently invariant.
4. Collectability: The feature should be quantitatively measurable.

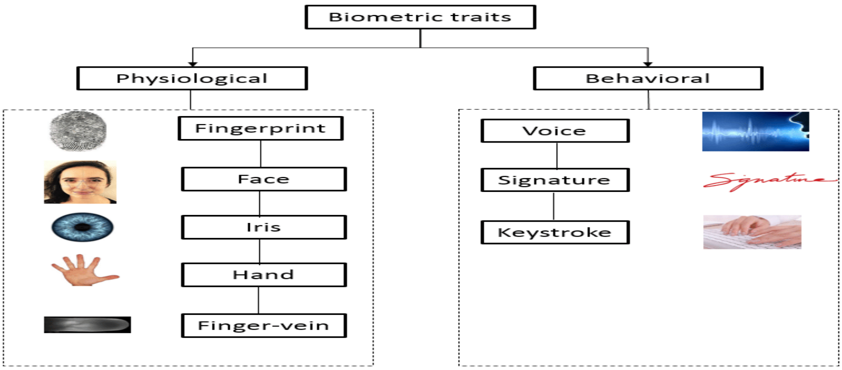


Figure 1: Classification of Biometric traits

However, for a practical biometric system (i.e., a system that employs biometrics for personal classification), other characteristics can be considered such as acceptability, circumvention, and performance [12]. Most importantly, a practical system must meet resource requirements, speed, and accuracy [13].
Different biometric characteristics are used in various applications. No single biometric can effectively meet the requirements of all applications, hence none is optimal. Each biometric has its strengths and weaknesses, and the choice of appropriate one is application-dependent [14]. Table 1 briefly compares biometric techniques based on measurements [15].
The ear is a new class of human biometric for identification with permanence and uniqueness. The ear has information-rich anatomical traits and is unaffected by ageing [16, 17]. Ear biometrics is convenient in collecting data compared to other technologies such as iris, fingerprint, and retina [18]. Moreover, tongue biometrics is a unique new biometric identification that cannot be easily forged because no

two tongues are similar [19]. The tongue is an individual vital organ well protected within the mouth and not affected by external factors [20]. The purpose of the proposed system is to investigate whether the combining of ear and tongue biometrics can attain higher performance that may not be likely using a single biometric indicator alone.

The remainder of this paper is organized as follows. Section 2 presents the multibiometric system. Section 3 presents a review of the previous studies. Section 4 presents the proposed method. Section 5 describes the various datasets that were used. Section 6 describes the use of datasets for evaluating the efficiency of the suggested algorithm, and Section 7 provides the conclusions drawn from this study.

Table 1: Comparison of various biometric technologies

Biometrics	Univer sality	Uniquenes s	Performan ce	Collect ability	Accept ability
Face	High	Low	Medium	High	High
Fingerprint	Medium	High	High	Medium	Medium
Hand Geometry	Medium	Medium	Medium	High	Medium
Iris	High	High	High	Medium	Low
Odor	High	High	High	Low	Low
DNA	High	High	High	Low	Low
Ear	Medium	Medium	High	Medium	High

2. Review of Multibiometric System

Biometric systems that depend on one source of information are called unimodal biometric systems, and they usually suffer from hacking or imposter attacks, low performance, and unacceptable error rates. The performance of the unimodal system is affected by non-universality and noisy sensor data [21]. To overcome these drawbacks, multimodal biometrics provide are used through a fusion of using two or more biometrics.

2.1 Multibiometric Fusion Mechanisms

Multimodal biometric systems provide the best recognition performance than systems based on a single biometric modality. The fusion technique is necessary and effective for combing information in the multimodal biometric system [22]. Biometric fusion occurs at different levels: (i) sensor level, (ii) feature level, (iii) score level, and (iv) decision level fusion.

The sensor level is also known as image-level fusion (for image-based biometrics or data level fusion [23]). In this level, multiple samples are combined to form a single sample. This type of fusion is applicable only if the multiple sources represent samples of the single biometric trait obtained either using a single sensor or different compatible sensors [24]. Feature level fusion is achieved by combining various feature sets extracted from multiple biometric sources [25]. The matching module compares the extracted feature set with the stored templates with the aid of a classifier or matching algorithm for giving matching scores [26]. While in the decision module, the matching scores are utilized either to verify a user's identity or identify an enrolled user [27].

2.2 Multibiometric Categories

Recognition systems using multimodal biometric traits can be designed to operate in one of the following integration scenarios, as shown in Figure 2:

1. Multi-Sensor: Different sensors can capture the same biometrics or body parts. Information and combined them using the sensor level fusion technique [28].
2. Multi-Modal: Different sensors capture different biometric modalities or body parts from the same person, e.g., retinal and face, which demands other sensors. This category can be more expensive; due to it requests multiple sensors that sense various biometric traits [29].
3. Multi-Sample: Multiple samples of the same biometric are collected during the enrolment and recognition phases. For example, multiple face samples are taken from the same person [30].
4. Multi-Algorithmic: The same sensor is utilized, however, its output is supplied to proposed using multiple algorithms for matching and feature extraction [31].
5. Multi-Instance: The use of the same sensor for capturing several instances of the same biometric trait or the same body part. With different face poses or images [32].

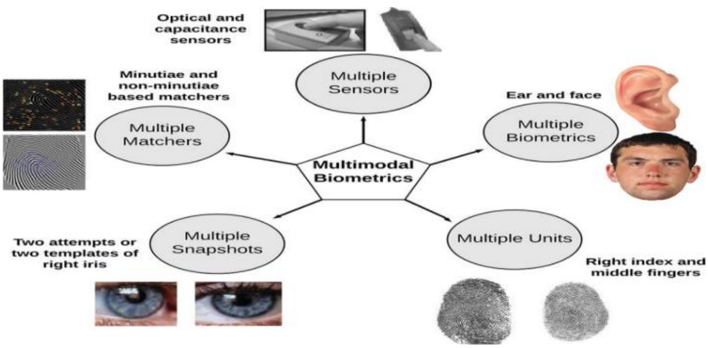


Figure 2: Types of multibiometric recognition

3. Literature Review

In this section, some of the previous works relevant to the multibiometric system are discussed below:

Darwish et al. (2009) [33], presented a multimodal biometric system for personal identification using face and ear. The proposed method segmented ear and face images followed by features extracting step based on Principal Component Analysis (PCA) algorithm called Eigenears and Eigenfaces. The proposed system was tested on several databases (such as the Massachusetts Institute of Technology (MIT) and the ORL dataset). The first dataset contains a collection of ear images and facial images from 40 individuals. At the same time, the second dataset is composed of 10 face images per 15 individuals. The overall recognition performance of the proposed system in terms of accuracy was 92.24%.

Tayyaba et al. (2013) [34], proposed a novel ear identification system, where three modules (database, image processing, and identification) are employed. The system extracts useful information and removes noise from the input images by the image processing module. The processing module consists of different phases: image acquisition, image enhancement, grayscale conversion, segmentation, thresholding, thinning, and template matching. The training set consists of 24 ear images. Experimental results showed that the system produces 98% accuracy.

Kaur et al. (2015) [35], presented a biometrics system based on signature, speech, and tongue. The experimental results showed the accuracy of this system is 88.75%, with 0.05% of FRR and with 0.06 % of FAR.

Songze et al. (2016) [36], introduced a multimodal recognition method based on profile face and ear feature fusion. The proposed method was evaluated by using the Collection E database of the University of Notre Dame, it includes 464 images from 114 individuals, the result was the recognition rate of profile face was 91.15%, the recognition rate of an ear was 93.81%, and the recognition rate of feature fusion was 99.12%.

Ali and Mariam (2017) [37], proposed a hybrid human recognition system based on face, palm print and ear images. The system was tested using the Sc Face dataset, consisting of 1440 images from 60 individuals. The method achieved a 97.4% recognition rate using face, ear, and palm fusion.

Thivakaran et al. (2019) [38] proposed a multimodal biometric system for authenticating a person's by utilizing fingerprint and ear. The proposed method was evaluated using the CASIA database (fingerprint images) and IIT Delhi database (ear images), and their system gave 95.66% accuracy with features fusion of both ear and fingerprint while achieving 80.21% accuracy when used only ear biometric, and achieved 85.01% accuracy when utilized only the fingerprint biometric.

Adebayo et al. (2021) [39] proposed “A Design and Simulation of Ear and Tongue Based Biometrics for Attendance Management System”. PCA was used for feature extraction, and Self Organizing Feature Map was utilized to train and test the system. The performance of their proposed method was evaluated based on Equal Error Rate (EER) and recognition accuracy. The experimental results showed the accuracy of this system is 99.9%.

4. Proposed Method

The suggested system aims to identify persons by using multibiometric (ear and tongue prints). The presented system offers enough stability for rotation and brightness variations and provides a high accuracy rate. It consists of four stages: (i) pre-processing for enhancing and segmentation the extracted Region of Interesting (RoI), (ii) fusion, (iii) feature extraction, (iv) classification. Figure 3 displays the layout of the suggested multibiometric system. Each module of the proposed method is described in the following subsections.

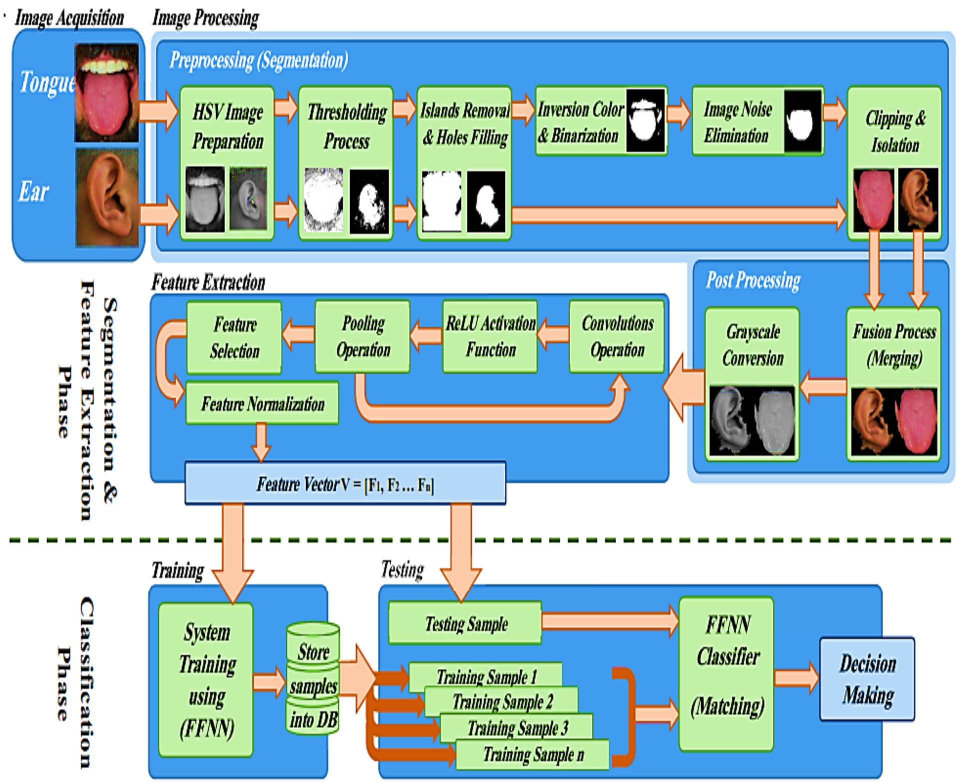


Figure 3: The detailed structure of the proposed system

4.1 Processing Stage

This stage aims to extract the ear/tongue region (RoI) from color images by applying many steps. Figure 4 presents the steps of the pre-processing stage for both ear and tongue images. These steps have been illustrated in our previous articles [40, 41] except the clipping step, which is utilized to allocate ear/tongue location in the resulted images and disregard background region; this is done by applying the clipping process; this process is done via making four detach scans (i.e., along with the four directions: up, down, right, & left) with a given width from the centre of every direction for capturing the first hit of a white pixel in it. After that, the locations of the four-hit points are utilized to determine the ear/tongue area coordinates. Figure 4 displays the resulting image after applying the clipping process.

(a) (b) (c) (d)

Figure 4 :Clipping process; (a) Ear image, (b) After clipping, (c) Tongue image, (d) After clipping.

4.2 Fusion Stage

After extracting RoI from the two input images (i.e., ear and tongue), the fusion process was applied to concatenate the two images into one image. First, the size of the resulting image would be determined by Equation (1) and (2) [42]:

$$\boldsymbol{Nw} = (\boldsymbol{Wid_e} + \boldsymbol{Wid_t}) \qquad \text{eq(1)}$$

$$\boldsymbol{Nh} = \boldsymbol{max}(\boldsymbol{High_e}, \boldsymbol{High_t}) \qquad \text{eq(2)}$$

where Nw and Nh represent the width and height of the merged image, respectively, and Wid_e and Wid_t represent the width of ear image and width of tongue image respectively. $High_e$ and $High_t$ stand for the height of the ear and tongue images, respectively, and max represents the maximum function. Then, the extracted ear pattern was first added to the new image. After that, the extracted tongue print was added to it, for each of the images would be added to the new image with starting X, Y coordinates of 0. Figure 5 shows the resulting image after applying the fusion process.

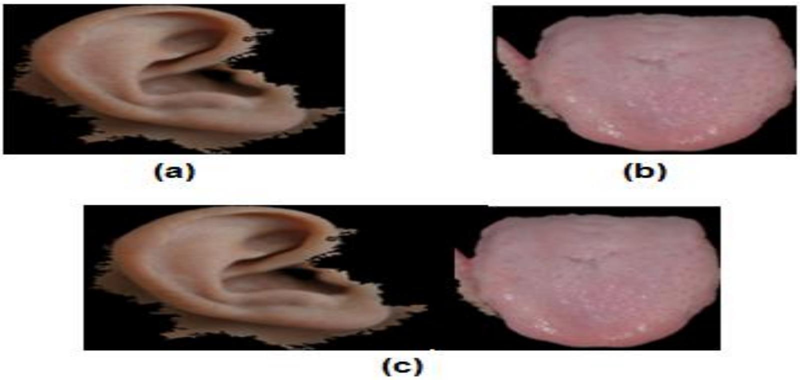


Figure 5: Merging process result; (a) Ear pattern Image, (b) Tongue print Image, (c) Resulting image after the merging

4.3 Feature Extraction Stage

Convolution Neural Network (CNN) is a popular deep learning algorithm used in many applications such as image classifications, image recognition, feature extraction, object detections, etc., [43]. CNN is a feedforward multi-layered network that consists of four layers: (i) convolution, (ii) activation, (iii) pooling, and (iv) fully connected [44], as shown in Figure 6. CNN is used to extract the essential features from the input image.

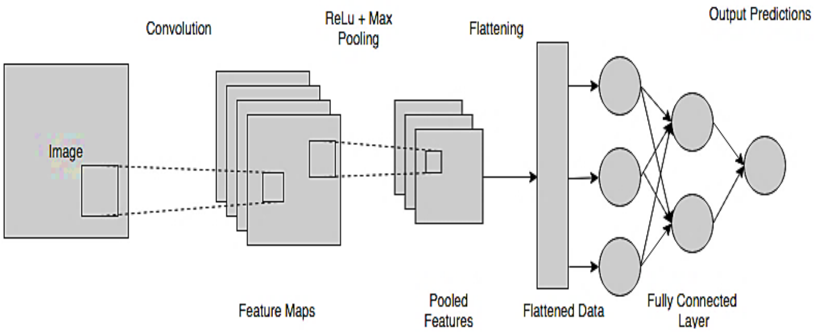


Figure 6: A general convolution neural network architecture

The resulted image from the previous steps (i.e., fusion stage) is converted to a greyscale image using Equation (3) [45]:

$$\boldsymbol{Gr}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathbf{1}}{\mathbf{3}}\left(\boldsymbol{P}_{\boldsymbol{red}}(\boldsymbol{x},\boldsymbol{y}) + \boldsymbol{P}_{\boldsymbol{green}}(\boldsymbol{x},\boldsymbol{y}) + \boldsymbol{P}_{\boldsymbol{blue}}(\boldsymbol{x},\boldsymbol{y})\right) \qquad \text{eq(3)}$$

where Gr(x,y) is the grayscale image, and $\boldsymbol{P}_{\boldsymbol{red}}(,)$, $\boldsymbol{P}_{\boldsymbol{green}}(,)$ and $\boldsymbol{P}_{\boldsymbol{blue}}(,)$ are pixel values of red, green, and blue channels, respectively. Figure 7 shows the result after the process of grey image preparation.

After that many operation is applied to extract the essential feature from the grey image, and the operations are: (i) Convolutions, (ii) Non-linearity activation function, (iii) Pooling, and (iv) Feature selection. Firstly, the convolutional layer comprises a set of convolutional kernels used for the extraction of feature map from the input image [46]. In the convolution layer; the law's mask is applied to extract texture features and using one or more types of law's masks (e.g., level (LL), edge (EE), and spot (SS)), and it produces specific information with each combination of these masks [47]. The result of the convolution operation is called a feature map.

Secondly, the activation layer is used as one of the activation functions to introduce the non-linearity into neural networks. This process is accomplished by applying the Relu activation function using the following equation [48]:

$$\boldsymbol{Relu}(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} \mathbf{0} & \boldsymbol{if\ Conv_{Img}}(\boldsymbol{x},\boldsymbol{y}) \leq \mathbf{0} \\ \mathbf{1} & \boldsymbol{Otherwise} \end{cases} \qquad \text{eq(4)}$$

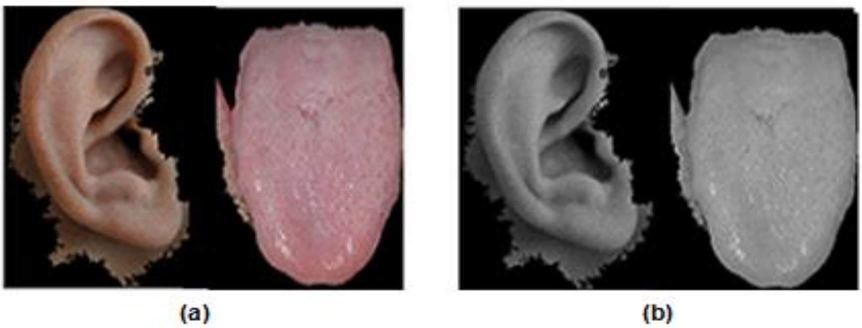


Figure 7: Grey Image Preparation; (a) Before, (b) After

where $Relu\ (,)$ represents the output image after applied Relu activation function, $Conv_{Img}(,)$ represents the image generated by a convolution operation. Also, x and y are the coordinates of the image.

Thirdly, the pooling layer is used for sub-sampling the obtained feature maps; this layer is also known as the sub-sampling layer [49]. Max pooling (or maximum pooling) is used in pooling operation, which calculates the maximum value for each patch of the feature map.
Finally, the feature selection process is used to remove the less important or irrelevant features to achieve better accuracy and reduce the computational cost. This operation is done by calculating some statistical features from the input features map (i.e., Pool_Img ()) and selecting these statistical features to be used as texture features for the classification stage. The statistical features are: (i) Mean (μ), (ii) Variance (σ2), (iii) Standard Deviation (Std), (iv) Energy, (v) Skewness, (vi) Kurtosis, (vii) Entropy, (viii) Contrast, (ix) Dissimilarity, and (x) Homogenous. All these statistical features are determined using the following equations [50, 51]:

$$\mu = \frac{\sum_{i=0}^{W-1}\sum_{j=0}^{H-1} P(i,j)^2}{(W-1)\times(H-1)} \qquad eq(5)$$

$$\sigma^2 = \frac{1}{(W-1)\times(H-1)}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(P(i,j)-\mu)^2 \qquad eq(6)$$

$$Std = \sqrt{\frac{1}{(W-1)\times(H-1)}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(P(i,j)-\mu)^2} \qquad eq(7)$$

$$Energy = \sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(P(i,j))^2 \qquad eq(8)$$

$$Skewness = \frac{\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(P(i,j)-\mu)^3}{(W-1)\times(H-1)\times std^3} \qquad eq(9)$$

$$Kurtosis = \frac{\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(P(i,j)-\mu)^4}{(W-1)\times(H-1)\times std^4} \qquad eq(10)$$

$$Entropy = -1 \times \left(\sum_{i=0}^{W-1}\sum_{j=0}^{H-1} P(i,j)\times log_2\big(P(i,j)\big)\right) \qquad eq(11)$$

$$Contrast = \sum_{i=0}^{W-1}\sum_{j=0}^{H-1} P(i,j)\times (i,j)^2 \qquad eq(12)$$

$$\boldsymbol{Dissimilarity} = \sum_{i=0}^{W-1}\sum_{j=0}^{H-1} \boldsymbol{P(i,j)} \times |\boldsymbol{i-j}| \qquad \text{eq(13)}$$

$$\boldsymbol{Homogenous} = \sum_{i=0}^{W-1}\sum_{j=0}^{H-1} \frac{\boldsymbol{p(i,j)}}{\boldsymbol{1+(i-j)^2}} \qquad \text{eq(14)}$$

where W and H represent the width and height of the pooling image, respectively, i and j are the coordinates of the pooling image, respectively, and p(,) are pixels values in the image (pooling()).
The Convolutions, Non-linearity activation function, and Pooling are repeated more one time to get acceptable results.

4.4 Classification Stage

In this stage, the extracted features vector is fed as inputs to the fully connected feedforward neural network [52, 53], which uses as a classification layer to predict the best labels to describe the image. But before that due to the extracted features vector is on different scales, so to minimize the effect of large values of the features vector, it is important to apply features normalization on the features vector before feeding it to the classification stage for training and testing. In the proposed system, the max-min normalization is applied to convert all values of the features vector (i.e., SFV ()) into a specific range (0, 1) using the following equation [54]:

$$\boldsymbol{Xi_{new}} = \frac{\boldsymbol{X_i - X_{Min}}}{\boldsymbol{X_{Max} - X_{Min}}} \qquad \text{eq(15)}$$

where Xi_{new} is the obtained normalized feature component, Xi is the input feature component, X_{min} represents the minimum value in the features vector, while X_{max} represents the maximum value in the features vector.
The classification stage consists of two steps: (i) the training step in which the system is trained using a training dataset for neural network enrolment, and (ii) the testing step to check the performance of the system using a test dataset and using the set of weights generated in the training step.
Datasets

5. Datasets

The proposed system was tested using Tongue and Ear Images Datasets (TEID). TEID is consists of three sub-datasets: (i) First Dataset (Ds1), (ii) Second Dataset (Ds2), (iii) Third Dataset (Ds3).
Ds1 consists of a Collection of Tongue and Ear Images (CTEI). It was collected by taking both the snaps of tongue and ear from students and staff of the vision and health management departments at Al-Mansour Technical Medical Institute.

The dataset was acquired ear and tongue images from 30 different subjects, all of them ageing is range from 19 to 43 years old. All the images were taken using three models of Nikon camera (i.e., Nikon D3200 and Nikon D5100)), under different lighting conditions with the subject placed at a distance of about 0.5 meters from the camera, and we used a 55mm focal length for the ear and tongue images. The resolution of images from camera1 and camera2 are (690×1012) and (572×838) pixels, respectively, in JPEG format. While, Ds2 consists of two datasets (AMI Ear Database and Tongue Dataset1). The AMI Ear Database is publicly available, and it was downloaded from the standard reference database [55] were taken 27 ear images, each image has size (522×732) pixels in JPEG format. The Tongue Dataset1(TD1) was taken from an online site that is publicly available [56] which is consisted of 27 tongue images, each image has size (406× 345) pixels in JPEG format.

In addition, Ds3 has consisted of two datasets (Tongue Dataset 2 and Ear Image Dataset). Tongue Dataset 2 (TD2) was taken from a master student [57], which is consisted of 30 tongue images; each image has size of (712× 720) pixels in JPEG format. The with a size of Ear Image Dataset (EID) was taken from a master student [58], which is consisted of 30 ear images, each image has size (273×328) pixels in JPEG format.

Each image was rotated six times (with rotation steps of 5 degrees) in both clockwise and anticlockwise directions. A total of 12 images variants have been produced from each tongue and ear image. Thus, a dataset consisting of 2088 images (i.e., 1044 for tongue and 1044 for ear images) is established. Figure8 shows some of the ear and tongue samples of the considered datasets.

C1S1

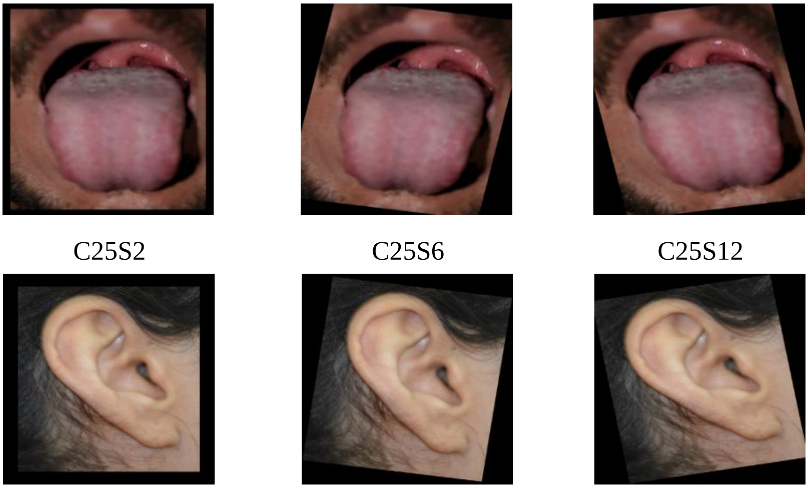
C1S5

C1S12

C10S2

C10S6

C10S11



C25S2 C25S6 C25S12

C47S1 C47S6 C47S11

Figure 8: Samples of taken ear/tongue images as materials for system performance tests. The images belong to 3 different datasets.

6. Results and Discussion

There are many parameters in the CNN algorithm that play a vital role in recognition rate; they are (i) The type and size of filters in convolution operation, (ii) The selected type of pooling operation, (iii) The size of sliding in the pooling operation, and (iv) The number of repetitions of CNN operations. The best-attained recognition result is met when the type of filter is Law's Edge filter (EE). The test results listed in Table 2 show the effect of choosing the type of law's filter in convolution operation on the recognition rate on three different datasets. Figure 9 shows a confusion matrix in the proposed system.

	Actual Values	
Predicted Values	1041	2
	1	0

Figure 9: Confusion matrix of the proposed system

In addition to this, the filter size in the convolution operation, the type pooling and the size of sliding in the pooling operation show effects on the recognition rate, and the test results show that the best result is obtained when the filter size is taken 3×3 with max-pooling and sliding size 2×2. Table 3 shows the effects of some parameters of CNN operations on the recognition rate.

Table 2: The effect of choosing different Law's filter types on the recognition rate

Dataset Type	Filter Type	Recognition Rate (%)
Ds1	LL	87.5
	SS	84.72
	EE	99.72
Ds2	LL	96.29
	SS	94.44
	EE	100
Ds3	LL	95.55
	SS	95.27
	EE	99.44

Table 3: The effectiveness of different parameters in CNN operations on the accuracy rate of the identification system.

Convolutional Kernel Size	Pooling Type	Sliding Size	Recognition Rate (%)
3×3	Max pooling	2×2	99.72
3×3	Max pooling	4×4	91.38
3×3	Average pooling	2×2	93.33
3×3	Average pooling	4×4	87.50
5×5	Max pooling	2×2	85.27
5×5	Max pooling	4×4	84.44
5×5	Average pooling	2×2	88.05
5×5	Average pooling	4×4	85.50

We can notice from the experimental result that the best recognition rate is attained when the number of repetitions of CNN operations (no_itr) is 2 for DS1 and DS2, whilst no itr=1 for DS3; as shown in Table 4. So, the obtained average recognition rate from the proposed system is 99.72%.

The outcomes of the conducted experiments indicated the need for the feature normalization process before the classification stage to keep features contrast in one level for getting a higher accuracy rate of an individual's identification. Also, clipping ear/ tongue images in the pre-processing stage is helpful to decrease the time consumption of the subsequence stages (i.e., feature extraction and classification stage). This method is robust to image rotation and variation in size. The proposed method extracts a small number of discriminative features so that the method is algorithmically simple and requires low computational complexity. The achieved results of recognition accuracy by the methods recently published in the literature have been compared to the obtained results by our proposed system, as listed in Table 5.

Table 4: The effect of the number of repetitions of CNN operations on the recognition rate.

Dataset Type	Filter Type	Repetition No.	Recognition Rate (%)
Ds1	EE	1	94.16
		2	99.72
		3	97.50
Ds2	EE	1	96.60
		2	100
		3	99.07
Ds3	EE	1	99.44
		2	97.77
		3	98.33

Table 5: Comparison of the recognition rate of our proposed system with some other previously proposed methods

Reference	Biometrics Type	Biometric Traits	Recognition Rate (%)
[33]	Multimodal	Face and Ear	92.24
[34]	Unimodal	Ear	98
[35]	Multimodal	Speech, Signature, and Tongue	88.75
[37]	Multimodal	Face, Ear, and Palm	97.4
[38]	Multimodal	Fingerprint and Ear	95.6
The Proposed System	Multimodal	Ear and Tongue	99.72

7. Conclusions

Biometrics is the most relevant means of authenticating and identifying persons in a fast and reliable way via unique biological characteristics. Biometric traits are unique to every person and remain unchanged throughout an individual's lifetime. These features made biometrics a promising solution to society. In this paper, a robust multimodal biometric system by using a fusion of two biometric traits (ear and tongue) at the image fusion level. The proposed system is algorithmically simple, less complex, and robust against rotation and variance in size. CNN is applied to get invariant features against any rotation to address the effect of rotation. The outcomes of the conducted tests on three various datasets indicate that the average recognition rate obtained with the proposed system is 99.72% for all datasets.

References

[1] Sahana, J. S., Tarun R. R. and Manjunath C. R., (2019). "Multi Biometric Recognition System," International Journal of Computer Science and Mobile Computing, vol.8, issue.6, pp. 89-9.

[2] Tsekeridou S. and Pitas I., (1998). "Facial Feature Extraction in Frontal Views Using Biometric Analogies," In 9th European Signal Processing Conference (EUSIPCO 1998), IEEE, pp. 1-4.

[3] Wang, K., Zhang Y., Yuan Z. and Zhuang D., (2006). "Hand Vein Recognition based on Multi Supplemental Features of Multi-Classifier Fusion Decision," In 2006 International Conference on Mechatronics and Automation, IEEE, pp. 1790-1795.

[4] A. Benzaoui, A. Hadid and A. Boukrouche, "Ear Biometric Recognition Using Local Texture Descriptors," Journal of electronic imaging, vol. 23, no. 5, 2014.

[5] Jain, A. K., Bolle R. and Pankanti S., (2006). "Biometrics: Personal Identification in A Networked Society," Springer Science & Business Media, vol. 479.

[6] Tuama S. A. and George L. E., (2016). "An Efficient Method for Automatic Human Recognition Based on Retinal Vascular Analysis," Research Journal of Applied Sciences, Engineering and Technology, vol. 12, no.1, pp.122-128.

[7] França L. T., Carrilho E. and Kist T. B, (2002). "A review of DNA sequencing techniques. Quarterly reviews of biophysics", Vol. 35, No .2, 169-200.

[8] Gafurov D., Snekkenes E. and Buvarp, T.E., (2006). "Robustness of Biometric Gait Authentication Against Impersonation Attack," In OTM Confederated International Conferences, On the Move to Meaningful

Internet Systems, Springer, Berlin, Heidelberg, pp. 479-488.

[9] R. A. Maxion and K. S. Killourhy, (2010). "Keystroke Biometrics with Number-Pad Input," In 2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN), IEEE, pp. 201-210.

[10] G. Bailador, C. Sanchez-Avila, J. Guerra-Casanova and A. Santos Sierra, (2011). "Analysis of pattern recognition techniques for in-air signature biometrics," Pattern Recognition, vol. 44, no. 10-11, pp.2468-2478.

[11] D. Kumar, C. Deepika and K. Chadha, (2014). "Biometrics and Face Recognition," International Journal of Innovative Research in Technology (IJIRT), ISSN: 2349-6002, vol. 1, issue 5.

[12] B. Meghna, R. Ashok and N. J. Patel, (2014). "Analyzing of Various Biometric System Predicted on Different Criteria Affected by Various Factor," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, issue 12, pp. 757-766.

[13] P. Salil, P. Sharath and K. J. Anil, (2003). "Biometric Recognition: Security and Privacy Concerns," IEEE Computer Society, vol. 1, no. 2, pp. 33-42.

[14] D. Mellado, L. E. Sánchez, M. E. Fernández and M. Piattini, (2012). "IT Security Governance Innovations: Theory and Research," A book published by IGI Global, USA , ISBN: 1466620846, pp. 210.

[15] M. Kalyani and B. Samayita, (2013). "Comparative Study of Different Biometric Features," International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, Issue 7.

[16] M. Yichao, Z. Huang, X. Wang, and K. Huang, (2020). "An overview of multimodal biometrics using the face and ear." Mathematical Problems in Engineering 2020, pp: 1-17. Article ID 6802905 | https://doi.org/10.1155/2020/6802905.

[17] N. Boodoo and R.Subramania, (2009). "Robust multi biometric recognition using face and ear images. International Journal of Computer Science and Information Security (IJCSIS), vol. 6, no. 2.

[18] K. P. Tripathi, (2011). "A Comparative Study of Biometric Technologies with Reference to Human Interface," International Journal of Computer Applications, vol. 14, no. 5, pp. 10-15.

[19] J. Nadeem, T. Radhika and S. Nithya, (2017). "Tongue Prints in Biometric Authentication: A Pilot Study," Journal of Oral and Maxillofacial Pathology, vol. 21, issue 1, pp. 176- 179.

[20] T. Sivakumar, S. Nair, G. Zacharias, M. Nair and A. Joseph, (2018).

"Identification of Tongue Print Images for Forensic Science and Biometric Authentication," Journal of Intelligent & Fuzzy Systems, vol. 34, no. 3, pp. 1421-1426.

[21] F. Besbes, H. Trichili and B. Solaiman, (2008). "Multimodal Biometric System Based on Fingerprint Identification and Iris Recognition," In April 2008 3rd international conference on information and communication technologies: from theory to applications, IEEE, Syria, pp. 1-5.

[22] Y. M. Fouda, (2012). "Fusion of Face and Voice: An Improvement," International Journal of Computer Science and Network Security, vol. 12, no. 4.

[23] N. Radha and A. Kavitha, (2011). "Rank Level Fusion Using Fingerprint and Iris Biometrics," Indian Journal of Computer Science and Engineering (IJCSE), vol. 2, no. 6, pp. 917-923.

[24] A. Ross and A. K. Jain, (2007). "Fusion Techniques in Multibiometric Systems," In Face Biometrics for Personal Identification, Springer, Berlin, Heidelberg, pp. 185-212.

[25] A. Nada and H. H. Al-Baity, (2020). "Deep Learning Approach for Multimodal Biometric Recognition System Based on Fusion of Iris, Face, And Finger Vein Traits", Sensors 20, no. 19.

[26] S. C. Dass, K. Nandakumar and A. K. Jain, (2005). "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," In International conference on audio-and video-based biometric person authentication, Springer, Berlin, Heidelberg, pp. 1049-1058.

[27] O. O. Muhtahir and P. H. Gerhard, (2016). "Unimodal and Multimodal Biometric Sensing Systems: A Review," IEEE, vol. 4, pp. 7532- 7555.

[28] H. AlMahafzah and M. Z. AlRwashdeh, (2012). "A Survey of Multibiometric Systems," International Journal of Computer Applications, vol. 43, no. 15, pp. 36-43.

[29] R. Gad and A. EL-SAYED, (2015). "Multi-Biometric Systems: A State of the Art Survey and Research Directions," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 6, no. 6, pp. 128-138.

[30] A. Teoh, S. A. Samad and A. Hussain, (2004). "Nearest Neighbourhood Classifiers in a Bimodal Biometric Verification System Fusion Decision Scheme," Journal of Research and Practice in Information Technology, vol. 36, no. 1, pp. 47-62.

[31] M. Imran, A. Rao and G. H. Kumar, (2010). "Multibiometric Systems: A Comparative Study of Multi-Algorithmic and Multimodal Approaches," Procedia Computer Science, vol. 2, pp.207-212, pp. 208-212.

[32] H. AlMahafzah and M. Z. AlRwashdeh, (2012). "A Survey of Multibiometric Systems," International Journal of Computer Application, pp. 1-8.

[33] A. A. Darwish, R. Abd-Elghafar and A. A. Fawzi, (2009). "Multimodal Face and Ear Images," Journal of Computer Science, vol. 5, no. 5, pp.374-379.

[34] S. Tayyaba, M. W. Ashraf, N. Afzulpurkar, A. Hussain, M. Imran and A. Noreen, (2013). "A Novel Ear Identification System for Security Applications," International Journal of Computer and Communication Engineering, vol. 2, no. 2.

[35] G. Kaur and D. Singh, (2015). "A Novel Biometric System Based on Hybrid Fusion Speech, Signature and Tongue," International Journal of Computer Applications, vol. 119, no. 7, Pp. 1-10.

[36] L. Songze and Q. Min, (2016). "Multimodal Recognition Method based on Ear and Profile Face Feature Fusion," International Journal of Signal Processing, Image Processing and Pattern Recognition,vol.9, no.1, pp.33-42.

[37] M. M. Ali and M. S. Mariam, (2017). "Hybrid Recognition System under Feature Selection and Fusion," International Journal of Computer Science Trends and Technology (IJCST), vol. 5, issue 4.

[38] T. K. Thivakaran, S. V. Padira, A. S. Kumar and S. Reddy, (2019). "Fusion Based Multimodel Biometric Authentication System using Ear and Fingerprint," International Journal of Intelligent Engineering and Systems, , vol. 12, no. 1, pp. 62-73.

[39] O. Y. Adebayo, A. A. Adigun, T. T. Adeyemo and E. O. Isola, (2021). "Design and Simulation of Ear and Tongue Based Biometrics for Attendance Management System," UNIOSUN Journal of Engineering and Environmental Sciences, vol. 3, no. 1, pp. 12-17.

[40] Tuama S. A. and Saud J. H., (2020). "The Use of HSV Color Model for Subtle Ear Region Extraction," Iraqi Journal of Science, pp.1215-1222.

[41] Tuama S. A. and Saud J. H., (2019). "An Efficient Segmentation Method for Automated Tongue Extraction using HSV Color Model," International Journal of Advanced Science and Technology (IJAST), NADIA, vol. 133, pp. 1-10.

[42] Bigün, J., Bigun, J., and Gustavsson, T., (2003). "Image Analysis: 13th Scandinavian Conference, SCIA 2003, Halmstad, Sweden, June 29-July 2, 2003: Proceedings", Vol. 2749. Springer Science & Business Media.

[43] W. Song and J. Cai, "End-To-End Deep Neural Network For Automatic Speech Recognition," Standford CS224D Reports, Pp. 1-8, 2015.

[44] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin and V. K. Asari, (2019). "A State of the Art Survey on Deep Learning Theory and Architectures," Electronics, vol. 8, no. 3, pp. 292.

[45] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Third Edition, 2008.

[46] Q. Ke, J. Liu, M. Bennamoun, S. An, F. Sohel and F. Boussaid, (2018). "Computer Vision For Human-Machine Interaction," In Computer Vision for Assistive Healthcare, Academic Press, pp. 127-145.

[47] S. Dash and U. R. Jena, (2017). "Multi-resolution Laws' Masks Based Texture Classification," Journal of applied research and technology, vol. 15, no. 6, pp.571-582.

[48] C. Fan, (2016). "Survey of Convolutional Neural Network," Indiana University, pp. 1-22.

[49] Y. D. Li, Z. B. Hao and H. Lei, (2016). "Survey of Convolutional Neural Network," Journal of Computer Applications, vol. 36, no. 9, pp. 2508-2515.

[50] A. Materka and M. Strzelecki, (1998). "Texture Analysis Methods –A Review," Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels, vol.10, no. 1.97, pp. 1-33.

[51] M. Rachidi, C. Chappard, A. Marchadier, C. Gadois, E. Lespessailles and C. L. Benhamou, (2008). "Application of Laws' Masks to Bone Texture Analysis: An Innovative Image Analysis Tool in Osteoporosis," in 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, pp. 1191-1194.

[52] T. L. Fine, "Feedforward neural network methodology, (2006). "A book published by Springer Science and Business Media.

[53] D. Svozil, V. Kvasnicka and J. Pospichal, (1997). "Introduction to Multi-Layer Feed-Forward Neural Networks," Chemometrics and intelligent laboratory systems, vol. 39, no. 1, pp.43-62.

[54] K. Sanjaya, S. Panda and K. J. Prasanta, (2014). "A Smoothing Based Task Scheduling Algorithm for Heterogeneous Multi-Cloud Environment," 3rd

IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE, Waknaghat.

[55] E. Gonzalez, L. Alvarez and L. Mazorra. CTIM. R & D Center for Image Technologies, University of Las Palmas, (2000). http://www.ctim.es/research_works/ami_ear_database /subset-1.zip.

[56] M. Shi, G. Li, F. Li and Y. Wang. C2g2fsnake: A Novel Approach for Automated Tongue Segmentation Considering Prior Knowledge", submitted to Science in China (Information Science) (under review), (2010). http://levis.ton gji.edu.cn/gzli/code/C2G 2FSnake.zip.

[57] M. A. Yousif and J. H. Saud, (2019). "Tongue Print Features Extraction by Gabor Filters Family," Al-Mustansiriyah Journal of Science, vol. 13, issue 1, pp. 82-93.

[58] S. Salman and J. H. Soud, (2018). "Deep Learning Machine using Hierarchical Cluster Features", Al-Mustansiriyah Journal of Science, vol. 29, issue 3, Pp. 82-93.

استخدام المقاييس الحيوية المتعددة لتحديد هوية الانسان عبر التعلم العميق

صبا اياد طعمه[1] جميلة حربي سعود[2] زهراء عبدالحميد العبيدي[3]
sabaayad91@gmail.com dr.jameelahharbi@gmail.com zahraaalobaidy@gmail.com

المستخلص: أنظمة المقاييس الحيوية لتعرف الالي على الافراد تعتمد على المتغيرات السلوكية او الفسيولوجية المرتبطة بالافراد. تعمل أنظمة المقاييس الحيوية بأسلوبين: التحقق والتعريف للهوية. أسلوب التحقق، يتم فيه رفض او قبول الهوية اما أسلوب التعريف يتم فيه وصف هوية الشخص الغير مُعَرف. تُستخدم الأنظمة متعددة المقاييس الحيوية لتحديد هوية الفرد من خلال جمع المعلومات التي توفرها المتحسسات الحيوية او العينات او الوحدات او الخوارزميات او الخواص والميزات. تعد أنظمة متعدد المقاييس الحيوية موضوعاً بحثياً مثيراً للاهتمام. ويتم استخدامها للتعرف على هوية الأشخاص لغرض تحسين وزيادة مستوى امن وحماية المعلومات. من المتوقع ان استخدام هذه الأنظمة تساعد على منع الانتحال وتسهيل عملية المراقبة المستمرة كذلك تحسين التغطية السكانية و توفير حل للإخطاء في تطبيقات المقاييس الحيوية. ولهذه الأسباب تم تقديم هذه الدراسة لاقتراح نظام تعريف هوية الفرد بالاعتماد على نمط الاذن واللسان. تستخدم الشبكة العصبية الالتفافية (ش.ع.ا) لاستخلاص واستخراج الميزات والخواص الأساسية من الصور المدخلة. هذا النظام يعد قوي وفعال مع الضوضاء واضاءة السطوع وهو غير حساس لتغيرات في الدوران. الطريقة المقترحة تتكون من اربع مراحل رئيسية هي (المعالجة الأولية، الدمج، استخراج الميزات وأخيرا. مرحلة التصنيف). تم اختبار الطريقة المقترحة على ثلاث مجاميع مختلفة من البيانات الحقيقية وحققت معدل دقة يساوي ٩٩,٧٢٪ لجميع البيانات في المجموعات الثلاثة.

الكلمات المفتاحية: المقاييس الحيوية المتعددة، نظام تحديد الهوية، البيومترية للأذن، الشبكة العصبية الالتفافية، البيومترية للسان

[1] مدرس دكتور: قسم هندسة تقنيات الحاسوب - كلية الحكمة الجامعة – بغداد - العراق
[2] استاذ دكتور: قسم علوم الحاسوب - كلية العلوم - الجامعة المستنصرية - بغداد - العراق
[3] مدرس مساعد: قسم الهندسة تقنيات الحاسوب - كلية الحكمة الجامعة - بغداد - العراق