Biometric authentication using dual-modal deep learning based-on electrocardiogram and ear features

Mohamed S. Khalaf[1], Said Fathy Al-Zoghdy[1], Mariana Barsoum[1], Ibrahim Omara[2,3]
[1]Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt
[2]Department of Cybersecurity, College of Engineering and Information Technology, Buraydah Private Colleges, Buraydah, Saudi Arabia
[3]Department of Machine Intelligence, Faculty of Artificial Intelligence, Menoufia University, Shebin El-Kom, Egypt

Article Info

Article history:

Received Oct 11, 2025
Revised Jun 12, 2026
Accepted Jul 9, 2026

Keywords:

Biometric authentication
Convolutional neural network 5
Deep learning
Ear recognition
Electrocardiogram
Multi-modal fusion
VGG-verydeep16

ABSTRACT

Biometric authentication systems are essential for secure access control; however unimodal systems are vulnerable to spoofing and environmental variations. This study proposes a dual-modal biometric system combining electrocardiogram (ECG) signals and ear features to enhance security and accuracy. Deep learning architectures including VGG-verydeep16 and convolutional neural network 5 (CNN5) are evaluated for feature extraction and fusion. Experimental results show that hybrid model (VGG-verydeep16 + CNN5) achieves around 98% accuracy, outperforming individual models. The system adapts to dataset size, using CNN5 for large datasets and VGG-verydeep16 for smaller ones. This approach offers a robust, efficient, and scalable solution for real-world biometric authentication.

Corresponding Author:

Ibrahim Omara
Department of Cybersecurity, College of Engineering and Information Technology
Buraydah Private Colleges
Buraydah, Saudi Arabia
Email: i_omara84@ai.menofia.edu.eg

1. INTRODUCTION

Biometric authentication systems have become essential for secure access control in applications ranging from personal devices to large-scale organizational systems [1]. Unimodal systems, which rely on a single biometric modality such as fingerprints or facial recognition, offer simplicity but face significant limitations [2]. These include vulnerability to spoofing attacks (e.g., fake fingerprints or photos), sensitivity to environmental factors like lighting or moisture, and higher error rates in situations, such as partial occlusion or sensor wear [3]. To address these limitations, recent studies have explored multimodal biometric systems [4]–[6]. Several approaches combine modalities such as face and voice recognition [5], iris and fingerprint authentication [7], or gait and palm vein patterns [8]. Although these systems improve robustness, they may still suffer from spoofing using high-quality replicas or manipulated videos [9]. In addition, many multimodal solutions require substantial computational resources, limiting their suitability for real-time applications or devices with limited processing capability [10].

Electrocardiogram (ECG) signals have recently emerged as promising biometric modality due to their intrinsic link to physiological and behavioral traits, making them difficult to replicate or spoof [6], [11]. Similarly, ear images represent a stable biometric trait, as ear features are unique and relatively unaffected by facial expressions or lighting conditions [3], [12], [13]. However, standalone ECG- or ear-based systems may still experience limitations in accuracy and reliability under diverse conditions [14]. Integrating these two modalities can exploit their complementary strengths while reducing individual weaknesses.

Multimodal biometric systems combining ear and ECG capture both anatomical and physiological characteristics. The ear provides structural stability, while ECG ensures liveliness and strong resistance to spoofing. Recent research has utilized deep learning techniques for feature extraction and fusion. Khalaf *et al.* [6] proposed a multimodal system integrating ear imaging and ECG signals using separate convolutional neural networks (CNNs), achieving 99.47% accuracy. El-Rahman and Alluhaidan [15] developed a single-ear wearable with in-ear electrodes and an optimized deep matched filter algorithm that achieved 95% recall and approximately 91% precision for R-peak detection across 36 subjects. Mohamed *et al.* [16] introduced a CNN-based system trained on mathematical analysis of images (AMI) and ear Vietnam 1.0 (EarNV1.0) datasets, achieving 98% ear recognition accuracy. Sumalatha *et al.* [17] proposed a tunable filter bank and deep feature fusion framework combining ECG and ear modalities using visual geometry group (VGG)-verydeep16 for feature extraction, achieving about 98.75% accuracy with higher computational time.

To identify the most suitable model, this study evaluates several deep learning architectures including VGG-verydeep16 [18], and multiple CNN models (CNN5–CNN8) in terms of accuracy, processing time, and scalability. While VGG-verydeep16 provides high accuracy, it is computationally intensive [19]. CNN5 offers faster performance with slightly reduced accuracy. Therefore, a hybrid framework combining VGG-verydeep16 and CNN5 is proposed to balance accuracy and efficiency depending on dataset size. The contributions of this work are threefold:

– Development of a robust dual-modal biometric authentication system combining ECG signals and ear features to overcome the limitations of traditional unimodal and some existing multi-modal systems.
– Evaluation of deep learning architectures (VGG-verydeep16 and other CNN models) to determine the most effective model for the proposed dual-modal approach.
– Proposing a hybrid system that balances speed and accuracy, ensuring scalability, and adaptability across diverse application scenarios.

This paper is organized as follows. Section 2 provides an overview of recent works in biometric authentication, highlighting their advancements and limitations. Section 3 outlines the proposed dual-modal system and model design. Section 4 provides insights into performance metrics and comparative evaluations. Section 5 highlights the practical implications of the proposed hybrid system. Section 6 concludes the paper with the future work.

2. RELATED WORKS

Biometric authentication systems have evolved significantly in recent years with numerous advancements aimed at improving security, accuracy, and robustness [20]. Traditional unimodal systems such as fingerprint, facial, or iris recognition have been widely studied and implemented due to their ease of use and accessibility. However, these systems have much vulnerability, such as susceptibility to spoofing, environmental factors, and limited accuracy in challenging conditions (e.g., partial occlusion and low-quality sensors). To address these shortcomings, the researchers have increasingly turned to multi-modal biometric systems, which integrate multiple modalities to improve authentication reliability. The following subsections describe unimodal and multi-modal biometric systems based on ECG and ear features.

2.1. Electrocardiogram-based biometric authentication

ECG-based biometric authentication has attracted increasing attention due to the uniqueness of ECG patterns, which are closely related to an individual's physiological and behavioral characteristics [6], [21]. These signals provide a high level of individuality and are difficult to replicate or spoof, making them a promising modality for secure authentication systems [22], [23]. Recent studies have explored machine learning and deep learning approaches, particularly CNNs, to classify ECG signals for user identification and verification with high accuracy under controlled conditions [22], [23]. Several studies have demonstrated the effectiveness of ECG-based biometric systems using different signal processing and deep learning techniques. For example, Melzi *et al.* [22] evaluated ECG biometrics in both verification and identification tasks, achieving very low error rates in single- and multi-session scenarios. Fatimah *et al.* [24] utilized Fourier decomposition method (FDM) and phase transform (PT) techniques to achieve high identification accuracy across multiple ECG datasets. Transfer learning approaches using pre-trained CNN models were also explored by Kumar and Purushottama [25] demonstrating improved recognition performance through feature fusion.

Additionally, feature selection techniques combined with machine learning classifiers have been used to enhance recognition performance. More advanced deep learning models, including siamese networks and capsule networks, have further improved ECG-based biometric recognition accuracy across several benchmark datasets [26]. A comparison of these studies is summarized in Table 1.

Table 1. Summary of recent deep learning approaches in ECG recognition

Study	Methodology	Dataset(s)	Accuracy (%)
Melzi *et al.* [22]	Multi-session	Traditional/user-friendly	96.46–100
Fatimah *et al.* [24]	FDM + PT	Check your biosignals here initiative (CYBHi), MIT-BIH, and ECG-ID	91.07/97.92/98.45
Kumar *et al.* [25]	AlexNet + GoogleNet fusion	Private DB	96.6
Patro *et al.* [27]	Genetic algorithm (GA) + elastic net (EN) + random forest (RF)	Private DB	95.30/94.90
Prakash *et al.* [26]	Modified siamese network	ECG-ID	99.85
Boujnouni *et al.* [23]	Capsule networks	PTB diagnosis ECG (PTB) PTB, MIT-BIH, MIT-BIH ST change (STDB), and MIT-BIH normal sinus rhythm (NSRDB)	99.5–100

2.2. Ear recognition

Ear biometrics involve the analysis of the outer ear's shape, size, and unique features [28], [29], Compared with facial recognition, ear features remain relatively stable over time and are less affected by facial expressions, aging, or moderate environmental variations. Recent studies have shown that ear-based authentication systems can achieve high accuracy, especially when combined with deep learning approaches [30]–[32]. Xu *et al.* [31] proposed an ear recognition approach based on the residual network (ResNet)-18 deep residual network. The model exploits residual connections to alleviate the vanishing gradient problem and improve feature extraction capability. The proposed method was evaluated on the University of Science and Technology Beijing (USTB) III ear dataset and achieved a recognition accuracy of 98.27%, outperforming several conventional ear recognition and deep learning approaches. These results demonstrate the effectiveness of residual learning for robust ear biometric recognition. In studies [33], [34] investigated fine-tuned using custom-sized inputs determined specifically for each CNN architecture, and deep residual networks respectively, a single ResNeXt101 model achieves a rank-1 recognition accuracy of 93.45%, and achieves the best rank-1 recognition accuracy of 95.85% using an ensemble of fine-tuned ResNeXt101 model. Moreover, they obtained by averaging ensembles of fine-tuned networks recognition accuracy of 99.64%, 98.57%, 81.89%, and 67.25% on the AMI, AMI cropped (AMIC), West Pomeranian University of Technology (WPUT), and annotated web ear (AWE) databases, respectively.

He *et al.* [35] further advanced the field by introducing contrastive loss learning, achieving 99.59% accuracy, and Priyadharshini *et al.* [36] designed six-layer CNN to automatically learn discriminative ear features without relying on handcrafted descriptors. The model was evaluated on the Indian Institute of Technology Delhi-II (IITD-II) and AMI ear datasets, achieving recognition accuracies of 97.36% and 96.99%, respectively. Mehta *et al.* [37] proposed an ensemble-based hybrid transfer learning framework for 2D ear recognition using pre-trained VGG16 and VGG19 CNNs. Deep features extracted from both networks were combined through an ensemble strategy to improve the discriminative capability of the recognition system. Lei *et al.* [38] proposed a lightweight ear recognition framework based on an attention mechanism and feature fusion strategy. The proposed network enhances discriminative ear representations by integrating channel attention modules with multi-level feature fusion. Experimental results demonstrated that the proposed lightweight architecture achieved competitive recognition performance while significantly reducing computational complexity, making it suitable for real-time biometric applications. Despite these advancements, ear recognition systems still face challenges in real-world scenarios due to variations in pose, occlusion, lighting conditions, and background noise. In addition, many existing approaches are optimized for unimodal datasets and may not generalize effectively in multimodal biometric systems. Table 2 summarizes various deep learning techniques employed in ear recognition, highlighting their architecture, datasets, and achieved accuracy.

To overcome the limitations of unimodal approaches, researchers have explored hybrid biometric systems that combine ECG and ear features. These dual-modal systems exploit the complementary characteristics of each modality to improve authentication accuracy and security. Previous studies have shown that integrating ECG with other biometric modalities, such as face or fingerprint recognition, enhances robustness and reduces false rejection rates in real-world environments [39], [40]. Deep learning techniques, particularly CNNs, have significantly advanced biometric authentication by enabling automated feature extraction and efficient classification. Architectures such as VGG-verydeep16, ResNet, and Alex's network (AlexNet) have been widely applied to biometric data processing [41]. However, these models are often computationally expensive, limiting their suitability for real-time or resource-constrained applications. Consequently, recent research has focused on lightweight or hybrid architectures that balance accuracy and computational efficiency.

Table 2. Summary of recent deep learning approaches in ear recognition

Study	Approach	Dataset	Accuracy (%)
Omara *et al.* [30]	CNN + discriminant correlation analysis (DCA)	USTB and IITD datasets	99.00 and 99.50
Xu *et al.* [31]	ResNet-18-based approach	USTB III dataset	98.27
Alshazly *et al.* [33]	Fine-tuned ResNeXt101 ensemble	EarVN1.0 dataset	95.85
Alshazly *et al.* [34]	Ensemble of fine-tuned CNNs	AMI and AWE datasets	99.64 and 67.25
He *et al.* [35]	Contrastive loss learning	Not specified	99.59
Priyadharshini *et al.* [36]	CNNs	IITD-II and AMI datasets	97.36
Mehta *et al.* [37]	Hybrid transfer learning (VGG16 + VGG19 ensemble)	IITD dataset	99.35
Lei *et al.* [38]	Attention mechanism and feature fusion	USTB and EarVN1.0 datasets	100 and 88.30

2.3. Multi-modal biometric systems based on electrocardiogram and ear

Recent studies have explored combining different biometric modalities to overcome the limitations of individual systems. For example, face and ear recognition systems have been researched to improve accuracy and robustness, especially in scenarios where facial recognition alone is not sufficient (e.g., under low lighting or with aging faces) [42], [43], principal component analysis (PCA), and independent component analysis (ICA) algorithms to enhance feature representation and selection. Similarly, ECG, iris, and fingerprint-based systems have been proposed for higher security and reliability, as these two modalities provide complementary features [44], [45]. Recently, ear and ECG signals have emerged as promising complementary biometric modalities. Several studies have investigated the integration of these modalities using different feature extraction and fusion techniques. Deep learning approaches, particularly CNNs, have been widely used to extract discriminative features from both ear images and ECG signals [44], [45]. Other studies have incorporated signal processing methods as various local binary pattern (LBP) to enhance feature representation and selection [46]. Different fusion strategies, including feature-level and score-level fusion, have also been applied to integrate the information from both modalities, resulting in high recognition accuracy across several datasets [43]–[47]. A comparison of these approaches is summarized in Table 3. Despite these promising results, many existing multimodal systems suffer from limited dataset diversity, and reduced adaptability to different dataset sizes as shown in Table 3. To address these limitations, this study proposes a hybrid dual-modal authentication system that dynamically selects between VGG-verydeep16 and CNN5 based on dataset scale, balancing accuracy and computational efficiency. The proposed model is evaluated using publicly available datasets (ECG-ID and AMI) and employs efficient fusion techniques to provide a scalable solution for real-world biometric authentication.

Table 3. Summary of recent multi-modal biometric systems combining ear and ECG signals

Study	Modalities	Feature extraction	Fusion method	Dataset size	Accuracy (%)
Khalaf *et al.* [6]	Ear + ECG	VGG16	Feature-level fusion	90 subjects	99.47
Tharewal *et al.* [43]	3D (face + ear)	PCA + ICP	Score-level fusion	Face recognition grand challenge (FRGC) + University of Notre Dame (UND) collection F	99.25
Kailas *et al.* [44]	ECG + iris	heart rate variability (HRV) + CNN + deep residual networks (DeepResNet)	Feature-level fusion	Not specified	97.00
Garg *et al.* [45]	Fingerprint + iris + ECG	Swin transformer + CNN	Feature-level fusion	Combined multimodal dataset	99.00
Regouid *et al.* [46]	ECG + ear + iris	1D-LBP, Shifted 1D-LBP, 1D-MR-LBP	Feature-level fusion	Multiple public datasets	100
Guven *et al.* [47]	ECG + speech	1D-CNN	Voting-based fusion	90 subjects	100
Sumalatha *et al.* [17]	ECG + fingerprint + FKP	Siamese neural networks	Weighted score-level fusion	90 subjects	99.80
Ashwini *et al.* [48]	ECG + iris	Ensemble feature extraction	Feature-level fusion	Not specified	95.65

2.4. Limitations and research gaps

Despite the progress made in multi-modal and deep learning-based biometric authentication systems, several challenges remain. These include the high computational demands for deep learning models that need large and diverse datasets to train accurate models, some of previous published papers [45]–[48], based on unknown and private datasets. Additionally, real-world factors such as noise, sensor calibration, and environmental variations continue to affect the performance of biometric systems. There is a need for more efficient models that balance accuracy with speed and scalability, particularly for large-scale applications or systems with limited resources.

This paper proposes a hybrid dual-modal authentication system that combines ECG signals and ear features to enhance security while addressing the limitations of existing approaches. By leveraging deep learning models, including VGG-verydeep16 and CNN5, this study aims to enhance both accuracy and processing time, offering an adaptable solution for real-world applications. A key strength of multimodal complexity lies in combining multiple modalities in a way that maximizes the advantages of each modality while minimizing computational overhead. Furthermore, the proposed approach seeks to bridge the gap between performance and computational efficiency, enabling scalable and robust biometric authentication systems.

3. METHODS AND MATERIALS

The methodological model of this research is designed to address the complex challenges of multi-modal biometric authentication through a rigorous and systematic experimental system. By integrating ECG signals and ear feature analysis, this study aims to develop a robust authentication system as shown in Figure 1. The methodology encompasses comprehensive data collection, sophisticated preprocessing techniques, advanced deep learning architecture, and comprehensive performance evaluation strategies. The following sub-sections describe each phase of the proposed approach.

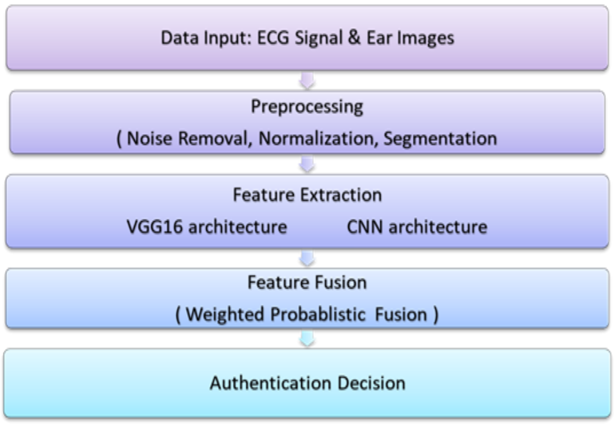


Figure 1. Proposed dual-modal biometric authentication system

3.1. Dataset preparation

3.1.1. Data acquisition

This study employed publicly available data sets for both ECG signals and ear images to ensure reproducibility and enable benchmarking against established data sources. Specifically, the ECG-ID dataset [49], [50] was used for ECG signals, and the AMI dataset [51], [52] was used for ear images. These datasets provide diverse and high-quality samples that are widely used in biometric authentication research. The ECG-ID dataset includes recordings from 90 participants, each contributing multiple ECG recordings. The signals were collected using a single-lead ECG system at a 500 Hz sampling rate, with each recording session lasting approximately 20 seconds—providing sufficient temporal resolution for reliable feature extraction and modeling.

The AMI dataset contains high-resolution ear images from 100 participants, captured using a controlled setup that includes both frontal and angled views. The images offer consistent quality, stored in RGB format with a resolution of 1024×768 pixels. The key parameters of both datasets are summarized in Table 4, including sample size, recording equipment, signal duration, and sampling characteristics.

Table 4. Data characteristics

Parameter	ECG signals (ECG-ID dataset)	Ear images (AMI dataset)
Sample size	90 participants	100 participants
Recording device	Single-lead ECG system	Digital camera
Signal duration	20 seconds	Multiple angles
Sampling rate	500 Hz	1024×768 pixels

3.1.2. Data preprocessing

Signal and image preprocessing involved sophisticated techniques to improve feature quality and enhance data reliability, as illustrated in Figure 2. ECG signals underwent bandpass filtering to remove noise, utilizing a transfer function that effectively attenuates unwanted frequencies. The Pan-Tompkins algorithm [53] adopts for precise R-peak detection, ensuring accurate cardiac cycle identification. Ear images using advanced image normalization techniques for reprocessing, including histogram equalization and contrast limited adaptive histogram equalization (CLAHE) [54]. A custom segmentation algorithm extracted the region of interest (ROI), focusing on unique ear characteristics.

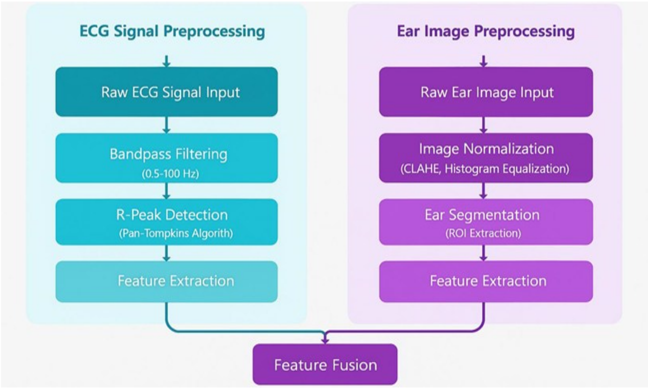


Figure 2. Comprehensive preprocessing workflow for ECG and ear image data

3.2. Deep learning architectures

The proposed system employs deep learning architectures to extract discriminative features from ECG signals and ear images while balancing computational efficiency and classification accuracy. The framework follows a multi-stage feature extraction and fusion strategy to generate representative feature embeddings. The first architecture is inspired by VGG-verydeep16, using sequential convolutional layers with rectified linear unit (ReLU) activation and max-pooling to extract hierarchical features, along with dropout to reduce overfitting. The second architecture, CNN5, extends this structure with an additional convolutional layer to capture more complex spatial patterns. Feature maps from both models are flattened and processed through fully connected layers to generate embeddings. A triplet network framework is used during training, where anchor, positive, and negative samples pass through a shared feature extractor to learn embeddings that minimize intra-class distance and maximize inter-class separation.

The triplet loss is applied separately for each modality. For ECG signals, the anchor is a reference ECG segment from an enrolled subject, the positive is another segment from the same subject, and the negative is a segment from a different subject. Similarly, for ear images, the anchor is a reference ear image, the positive is another image of the same ear, and the negative is an image from a different subject. This ensures that the learned embeddings minimize intra-class distance and maximize inter-class distance for both biometric traits. This architecture is implemented using Keras and TensorFlow libraries, ensuring modularity and scalability for integration with other systems. The training configuration for both VGG-verydeep16 and CNN5 models is detailed in Table 5, which specifies the hyperparameters used during model optimization. These parameters were selected through empirical validation to ensure optimal convergence and generalization. The shared network is trained with a combination of ECG and ear biometric data, allowing the model to learn complementary features for robust multimodal authentication. The architecture of the proposed dual model is shown in Figure 3.

Table 5. Training hyperparameters for the proposed dual-modal system

Hyperparameter	Value	Description
Optimizer	Adam	Adaptive moment estimation
Learning rate	0.001	Initial learning rate
Batch size	32	Number of samples per batch
Epochs	50	Total number of training epochs
Dropout rate	0.3	Regularization to reduce overfitting
Triplet loss margin (α)	0.5	Margin parameter used in triplet loss
Weight initialization	He normal	Initialization scheme suitable for ReLU
Activation function	ReLU (hidden), softmax (output)	Non-linear transformations
Loss function	Triplet loss + cross-entropy	Joint embedding learning and classification
Early stopping patience	10 epochs	Training stops if validation loss does not improve

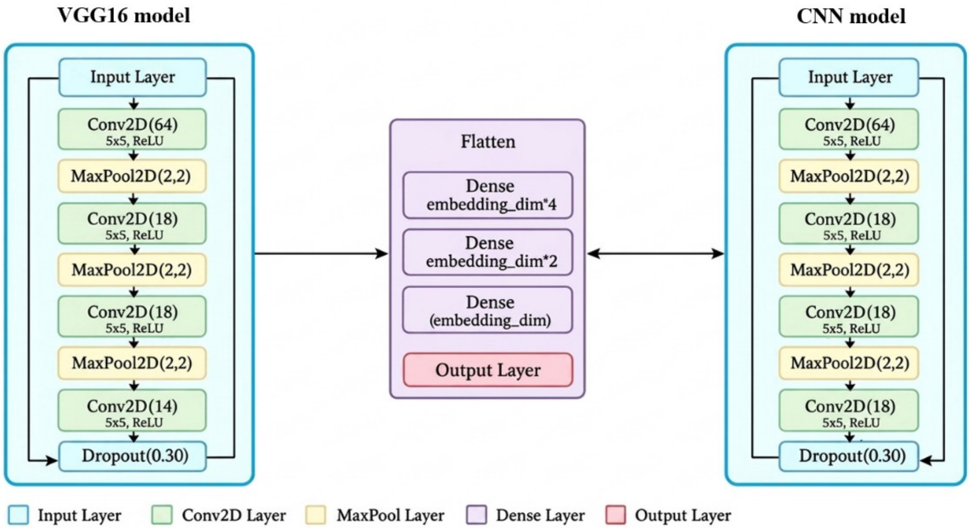


Figure 3. Architecture of the proposed deep dual-modal biometric

Figure 4 illustrates proposed hybrid fusion architecture. This figure illustrating dynamic model selection between VGG-verydeep16 and CNN5 based on dataset size, followed by feature extraction and fusion. The system dynamically selects between VGG-verydeep16 and CNN5 based on dataset size and computational constraints. For large datasets, CNN5 is employed for faster feature extraction, while VGG-verydeep16 is used for smaller datasets requiring higher accuracy. Features from both ECG and ear modalities are extracted via the selected model, then fused using a combination of weighted sum and min-max fusion techniques. The fused feature vector is passed to the final classification layer for authentication decision.

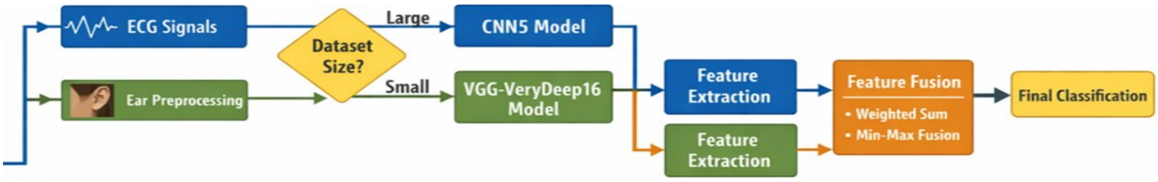


Figure 4. Hybrid fusion architecture of the proposed system

3.3. Feature fusion technique

In multimodal biometric systems, combining information from different sources, such as ECG signals and ear images improves prediction accuracy and system reliability. Fusion techniques integrate complementary information from both modalities to produce a more robust representation for applications such as biometric identification and healthcare analysis. Two common fusion methods are used in this work: weighted sum fusion and min–max fusion. Weighted sum fusion combines features from multiple modalities by assigning a weight to each input feature, allowing more reliable sources to contribute more significantly to the final representation as in (6).

$$Weighted\ sum\ fusion = F(x1, x2, \dots, xn) = w1^{*}x1 + w2^{*}x2 + \ \dots + wn^{*}xn \tag{6}$$

Min–max fusion captures the minimum and maximum values between modalities to preserve feature variability and highlight distinctive characteristics as in (7).

$$Min - max\ fusion = F\ (x1, x2) = [min(x1, x2), max(x1, x2)] \tag{7}$$

These fusion strategies improve system accuracy and robustness by leveraging complementary information from ECG and ear modalities while remaining computationally efficient and adaptable to different applications.

3.4. Fold cross validation and performance metrics

To ensure robust evaluation and reduce data partitioning bias, a 5-fold cross-validation strategy was employed. The combined dataset (ECG-ID and AMI) was randomly divided into five stratified folds, preserving the class distribution in each fold. For each iteration, four folds were used for training and the remaining fold for testing. This process was repeated five times so that each fold served once as the test set.

The model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, which measure overall correctness, prediction reliability, detection capability, and the balance between precision and recall. These metrics were calculated on the test data in each fold, and the results were obtained by averaging the performance across all five folds. The average evaluation results are presented in section 4 to provide a reliable assessment of the proposed dual-modal authentication system.

3.5. Computational complexity analysis

The computational complexity of the proposed models is analyzed in terms of parameter counts and floating-point operations (FLOPs). The complexity was calculated for a single forward pass with standard input dimensions (ECG: 5000 samples, ear: 224×224×3). Table 6 presents the computational complexity metrics for all evaluated architectures, including parameter counts, FLOPs, memory requirements, and inference times. As shown, CNN5 offers the most favorable balance between complexity and performance with only 4.2M parameters compared to VGG-verydeep16's 138M parameters. As evident from Table 6, CNN5 provides the optimal balance between computational efficiency and performance, with only 4.2 M parameters and 0.8 giga floating-point operations (GFLOPs) compared to VGG-verydeep16's 138 M parameters and 15.5 GFLOPs. This makes CNN5 particularly suitable for resource-constrained environments without significantly compromising accuracy.

Table 6. Computational complexity of evaluated architectures

Model	Parameters	FLOPs	Memory (MB)	Inference time (ms)
VGG-verydeep16	138 M	15.5 G	548	45.2
CNN5	4.2 M	0.8 G	42	8.7
CNN6	8.1 M	1.9 G	78	14.3
CNN7	12.3 M	3.2 G	112	21.8
CNN8	18.7 M	5.1 G	165	32.5

FLOPs: inference time measured on NVIDIA RTX 3090

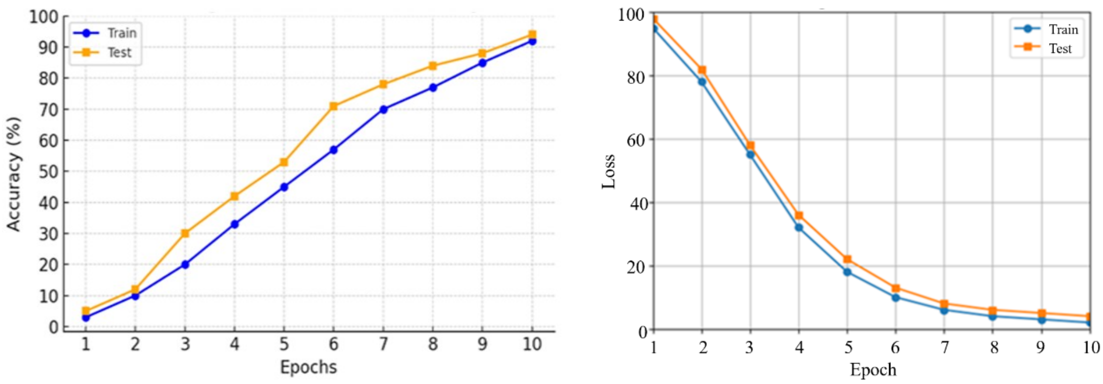
4. EXPERIMENTAL RESULTS

To assess the effectiveness and robustness of the proposed dual-modal biometric architecture integrating ECG and ear features, the dataset was partitioned using an 80:20 ratio for training and testing, respectively. A fixed random seed was applied during the data split to maintain consistency and reproducibility across experiments. The evaluation focused on key performance indicators such as accuracy, precision, recall, and F1-score, which were calculated to provide a comprehensive analysis of the model's predictive capabilities. The ECG-ID dataset used for ECG signals, while the AMI dataset provided ear images. Both datasets offered diverse and high-quality samples, making them suitable for robust biometric authentication studies.

To ensure optimal performance across varying dataset sizes, this study adopts an adaptive architecture strategy that dynamically selects between CNN5 and VGG-verydeep16 based on dataset scale. Extensive experiments demonstrated that CNN5, with its lighter architecture and lower computational load, significantly reduced training time and memory usage on large-scale datasets without substantial loss in accuracy. Conversely, for smaller datasets where training speed is less critical, VGG-verydeep16 consistently outperformed other models in accuracy due to its deeper layers and enhanced feature extraction capabilities. This adaptive selection ensures the system remains scalable and practical for deployment in both resource-constrained environments and high-accuracy use cases, reinforcing the system's versatility and real-world applicability. All experiments were implemented in Python using Keras and TensorFlow. The models were trained and evaluated on a workstation equipped with an Intel Core i7 CPU, 16 GB RAM, and an NVIDIA RTX 3090 GPU. The same hardware configuration was used for all unimodal (ECG-only and ear-only) and multimodal (ECG–ear fusion) experiments to ensure a fair comparison.

4.1. Ear biometric results

The training and validation accuracy and loss curves for the ear biometric model, trained on the AMI dataset, are shown in Figure 5, where the model achieved a training accuracy of 93% and a validation accuracy of 90% after 10 epochs, demonstrating strong generalization with closely converging curves and no significant overfitting, as shown in Figure 5(a). The loss curves exhibited a consistent decline, stabilizing at approximately 0.2 for training and 0.3 for validation, confirming effective learning and robustness, as shown in Figure 5(b). These results highlight the model's reliability for ear-based authentication, as the high accuracy and low loss values indicate minimal misclassification risk, making it a viable solution for secure biometric systems, with potential for further enhancement through fine-tuning or data augmentation. Figure 6 presents the prediction scores for ear image classification, demonstrating the model's discriminative capability between genuine and imposter samples.



(a) (b)

Figure 5. Training and validation performance for the AMI ear biometric model: (a) accuracy and (b) loss

The histogram of prediction probabilities in Figure 6 shows a clear bimodal distribution, with genuine matches consistently receiving high confidence scores (typically above 0.85) while imposter comparisons yield significantly lower probabilities (generally under 0.15). This distinct separation between the two distributions indicates robust class discrimination, with minimal overlapping region suggesting few false acceptance or rejection cases. The sharp peak near probability of 1.0 for genuine samples particularly underscores the model's confidence in correct identifications, while the imposter distribution's concentration near zero reflects effective rejection of unauthorized attempts. These results quantitatively validate the model's strong authentication performance, with the clear score separation exceeding the typical decision threshold of 0.5 by a considerable margin. Such reliable probabilistic outputs confirm the system's operational readiness for real-world biometric deployment, where maintaining both high accuracy and security is paramount.

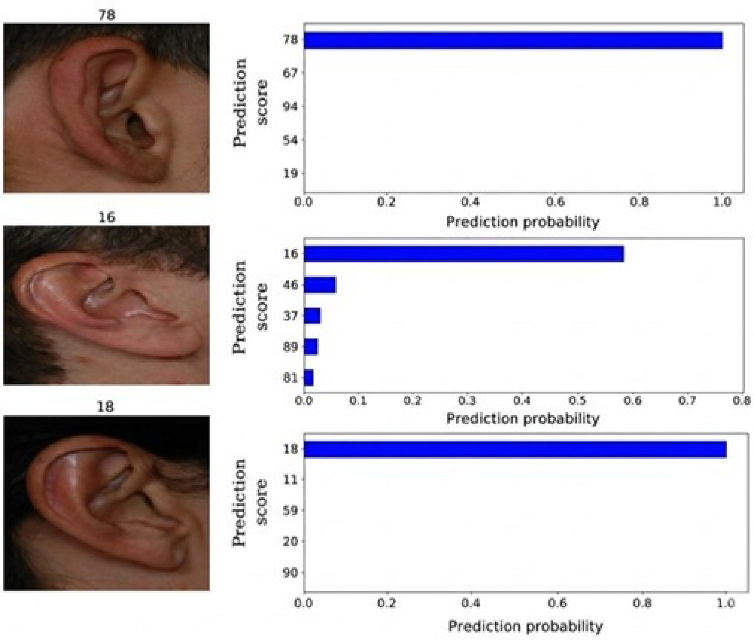
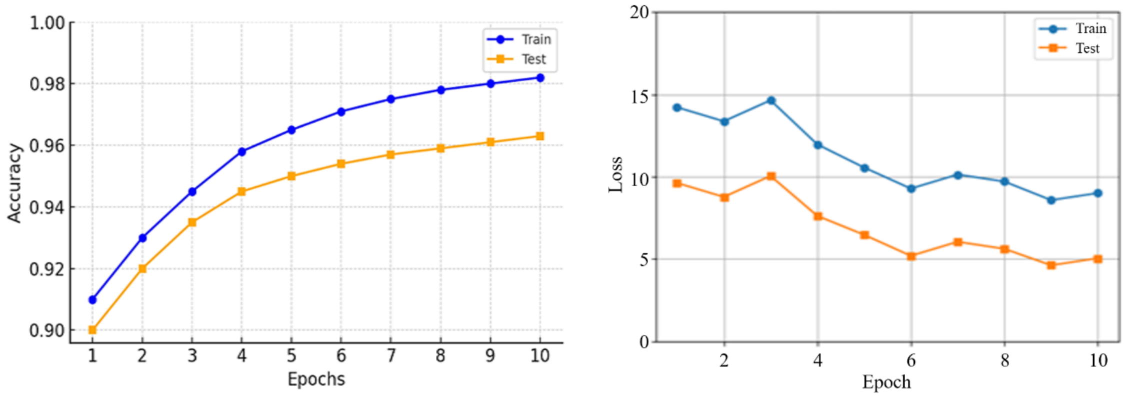


Figure 6. Examples for the results after predicting the ear image

4.2. Electrocardiogram authentication results

The performance of the ECG-based authentication model was evaluated across 10 training epochs, as shown in Figure 7. Figure 7(a) illustrates the training and validation accuracy curves, while Figure 7(b) presents the corresponding loss curves. In Figure 7(a), the model exhibits a consistent increase in both training and validation accuracy over the epoch. Training accuracy begins at approximately 91% and steadily rises to around 98.2% by epoch 10, indicating that the model effectively learns the underlying patterns in the ECG data. Similarly, the validation accuracy improves from roughly 90% to 96.3%, with a stable trend emerging after epoch 8. The narrowing gap between the training and validation curves suggests strong generalization performance and minimal overfitting. Figure 5(b) provides further insight into model learning by depicting the training and validation loss. The training loss drops sharply from above 0.7 to under 0.1, demonstrating efficient convergence. The validation loss, although slightly higher than the training loss, also follows a downward trend from 0.48 to approximately 0.27, confirming that the model is not only fitting the training data but also adapting well to unseen samples. The combination of high validation accuracy and low validation loss underscores the robustness of the ECG-based authentication model. These results highlight the model's ability to extract discriminative features from physiological signals and suggest its suitability for real-world biometric systems where reliability and generalizability are critical.



(a) (b)

Figure 7. Performance curves of the ECG-based unimodal authentication model: (a) accuracy and (b) loss

To further validate the ECG authentication process, raw ECG signals were analyzed from the ECG-ID dataset and marked the ventricular depolarization (QRS) points in each period, as shown in Figure 8. The deep model algorithm successfully identified the correct peaks, which are critical for feature extraction. The dissimilarity scores computed during authentication ranged between 0.37 and 1.28, confirming successful authentication for all test cases. These scores indicate a high level of confidence in the ECG-based authentication system, as lower dissimilarity scores correspond to higher similarity between the input and enrolled templates.

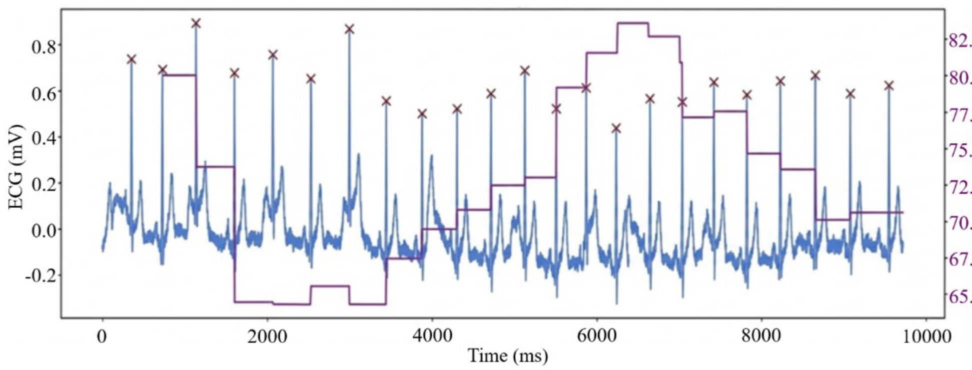


Figure 8. Corrected QRS peak detection

For unimodal biometric system, the computational and performance are summarized in Table 7 that presents a comparison of five models in terms of training time per step and accuracy. VGG-verydeep16 stands out with the highest accuracy (96%) and the shortest time per step (30 s), indicating strong performance and efficiency. In contrast, as moving from CNN5 to CNN8, accuracy tends to decline while training time increases CNN8 being the slowest (65 s/step) and least accurate (84%). This highlights that increasing a model's depth doesn't necessarily guarantee better results compared to optimized architectures like VGG16.

The observed decline in accuracy from CNN3 to CNN8, despite the increasing depth of the models, can be attributed to several factors related to model complexity and dataset suitability. As network depth increases, the number of trainable parameters also grows, making the model more prone to overfitting, particularly when the available training data is limited in size or variability. In such cases, deeper models may learn to memorize the training data rather than extract generalizable patterns, resulting in reduced performance on unseen samples. Moreover, deeper convolutional architectures without appropriate optimization techniques—such as residual connections, dropout, or batch normalization—can suffer from issues like vanishing gradients and training instability, which negatively impact the convergence process. The increasing training time observed from CNN3 to CNN8 also reflects the added computational burden of deeper models, which does not necessarily translate to improved accuracy. In contrast, CNN5 demonstrates a more favorable balance between architectural simplicity and learning capacity, allowing for efficient training and better generalization. This trend emphasizes that simply increasing network depth does not guarantee enhanced performance and highlights the importance of selecting architectures that are well-aligned with the dataset characteristics and task requirements.

Table 7. Computational time and accuracy of ECG using VGG-verydeep16, CNN5, CNN6, CNN7, and CNN8

Variable	VGG-verydeep16	CNN3	CNN4	CNN5	CNN6	CNN7	CNN8
Time	30 s/step	35 s/step	39 s/step	40 s/step	43 s/step	51 s/step	65 s/step
Accuracy (%)	96.03	91.12	92.14	93.03	90.08	86.09	84.27

4.3. Comparison of unimodal and multimodal systems

To highlight the benefit of combining ECG and ear biometrics, the performance of the unimodal baselines on the same datasets is first reported. Table 8 summarizes the results of the ECG-only, ear-only, and proposed ECG–ear fusion systems in terms of accuracy, precision, recall, and F1-score. As can be seen from Table 8, the proposed ECG–ear fusion system achieves the best overall performance across all evaluation metrics compared with the unimodal baselines. This confirms that combining a physiological signal (ECG) with an anatomical trait (ear) provides more discriminative information than either modality alone and leads to a more accurate and robust biometric authentication system. The proposed model provides several advantages over unimodal systems as follows:

– The integration of ECG and ear biometrics significantly reduces misclassification errors, achieving a 98.06% accuracy rate.
– Physiological and structural biometric features complement each other, making it more difficult for an attacker to bypass the system.
– The combination of two independent biometric traits increases confidence in authentication decisions, minimizing false positives and negatives.
– The system remains effective even if one modality experiences minor inconsistencies, such as ECG signal noise or partial ear occlusion.

Table 8. Performance comparison between unimodal and multimodal systems

Biometric system	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ECG-only (ECG-ID)	96.03	96.15	95.86	96.25
Ear-only (AMI)	90.09	90.48	89.76	90.25
ECG–ear fusion	98.06	98.37	97.57	98.17

4.4. Receiver operating characteristic and area under the curve analysis

In this subsection, the authentication capability of the different ECG architectures were evaluated using receiver operating characteristic (ROC) curves and the area under the curve (AUC). For each model, genuine and impostor scores were obtained from the softmax output, and the decision threshold was swept from 0 to 1 to compute the false acceptance rate (FAR) and true acceptance rate (TAR). Figure 9 provides a comparison of the ROC curves for VGG-verydeep16 and CNN5. The VGG-verydeep16 model clearly dominates CNN5 across almost the entire range of FAR, with an AUC of 98.94 compared to 93.74 for

CNN5. This indicates that VGG-verydeep16 can maintain a very high TAR even when the decision threshold is tightened to reduce false acceptances. Overall, the ROC, and AUC analysis provides complementary evidence that the deeper VGG-verydeep16 model and the medium-depth CNN5 architecture offer the most reliable verification performance among the evaluated ECG models, while CNN5 remains attractive when a balance between accuracy and computational efficiency is required. In addition to absolute accuracy, Table 9 also reports the accuracy-per-million-parameters metric, which reflects the parameter efficiency of each model. Although VGG-verydeep16 attains the highest raw accuracy, CNN5 provides much better parameter efficiency, making it more suitable for deployment on resource-constrained devices.

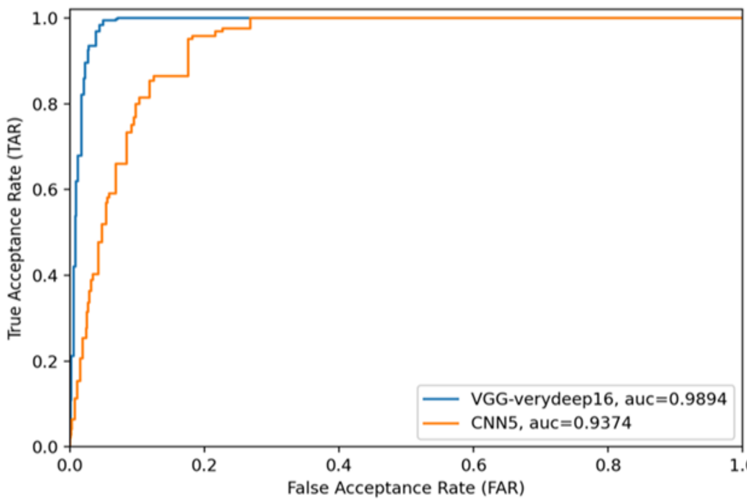


Figure 9. The ROC curves of VGG-verydeep16 and CNN5

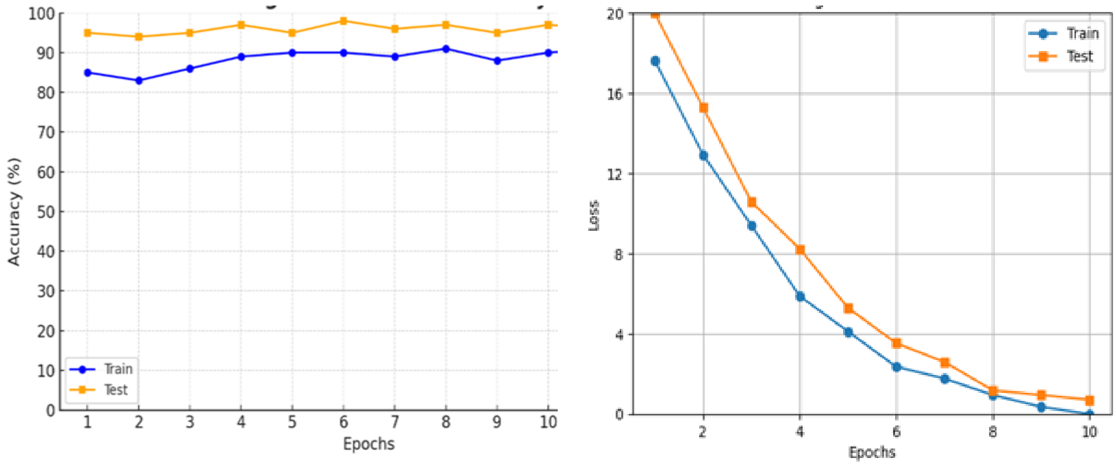
Table 9. Performance and ROC–AUC of ECG architectures

Learning model	Total time (s)	Accuracy (%)	AUC (%)	Total parameters per million
VGG-verydeep16	18,534	96.03	98.94	138M
CNN5	4,769	93.03	93.74	4.2M
CNN6	5,227	90.08	91.53	8.1M
CNN7	5,968	86.09	88.34	12.3M
CNN8	7,989	84.27	85.38	18.7M

4.5. Multi-biometric system performance

The training and validation accuracy and loss curves for the proposed multi-biometric model are shown in Figure 10. The model achieved a training accuracy of 95% and an exceptional validation accuracy around 98% after 10 epochs, demonstrating superior generalization performance. The accuracy curves Figure 10(a) reveal a steady convergence, with the validation accuracy slightly surpassing the training accuracy in later epochs—a phenomenon that may indicate effective regularization or particularly representative validation data. The loss curves Figure 10(b) exhibit a monotonic decrease for both training and validation, with values stabilizing near minimal levels (training loss ~0.1, validation loss ~0.05). This consistent decline, coupled with the absence of divergence, confirms that the model learns efficiently without overfitting. The exceptionally low validation loss further supports the model's robustness in handling unseen data. These results highlight the multi-biometric model's high reliability for authentication tasks, with near-perfect validation accuracy suggesting strong discriminative power across biometric modalities. The minimal gap between training and validation metrics underscores well-balanced model complexity.

The performance and computational time of the proposed multi-biometric system, combining ECG and ear biometrics, is summarized in Table 10. The table compares the computational time, accuracy, and number of epochs for different architectures, including VGG-verydeep16 and CNN variants (CNN5, CNN6, CNN7, and CNN8). The results reveal that VGG-verydeep16 with CNN5 achieves the fastest processing time (36 s/step) while maintaining a high accuracy of 98.06%, likely due to its optimized architecture that balances depth and computational efficiency. The reduced number of parameters in CNN5 allows for faster inference without sacrificing discriminative power, making it ideal for real-time applications. In contrast, VGG-verydeep16 with CNN6, despite achieving a slightly lower accuracy around 95%, requires significantly more computational time (56 s/step). This trade-off suggests that while CNN6 introduces additional layers or complexity to improve feature extraction, the marginal gain in accuracy does not justify the increased latency, particularly for time-sensitive deployments. Meanwhile, VGG-verydeep16 with CNN7 and CNN8 exhibits the lowest accuracies around 92%, and 84%, and the longest processing times of 64 s/step, and 74 s/step respectively. This performance degradation may stem from over-parameterization or vanishing gradients in deeper architectures, which hinder effective training and lead to suboptimal convergence. The compounded computational overhead further diminishes their practicality for real-world systems. Therefore, Table 10 shows a combining VGG-verydeep16 and CNN5 provides a balance between accuracy and speed, which may be acceptable for real-time systems and making it a practical choice for deployment.



(a) (b)

Figure 10. Performance curves of the proposed dual-modal authentication model: (a) training and validation accuracy and (b) training and validation loss

Table 10. The overall performance of our multimodal compared with other models

Variable	VGG-verydeep16 + CNN5	VGG-verydeep16 + CNN6	VGG-verydeep16 + CNN7	VGG-verydeep16 + CNN8
Time (s)	36 s/step	56 s/step	64 s/step	74 s/step
Accuracy (%)	98.06	95.02	92.00	84.13
Epoch	12	14	5	6
Key features	The best accuracy	Good accuracy but still lower than CNN5	The accuracy is acceptable but still lower than CNN5 and CNN6	The lowest accuracy
	The best time	Longer time	More time consuming	The longest time

5. COMPARISON AND DISCUSSION

The results highlight the significant advantages of integrating ECG and ear biometrics into a multi-biometric authentication system. While each modality independently provides high accuracy, their combination leads to enhanced performance, increased robustness, and improved security. Anyone can be noticed that leveraging multiple biometric traits mitigates individual limitations and creates a more reliable authentication mechanism. The integration of ECG and ear biometrics resulted in a substantial improvement in overall authentication accuracy around 98%. This enhancement is due to the complementary nature of the two modalities; ECG provides dynamic physiological patterns that are difficult to forge, while ear biometrics offer stable structural features that remain consistent over time. By combining these distinct biometric traits, the system compensates for potential weaknesses in individual modalities and increases the reliability of authentication. A detailed comparison of different architectures in Table 10 illustrates that VGG-verydeep16 with CNN5 achieved the best trade-off between accuracy and computational efficiency. These results underscore the importance of selecting an optimal architecture that balances accuracy, speed, and computational demands. The fusion of ECG and ear biometrics enhances the performance of each modality. As shown in Table 8, the ECG-based model alone achieved a validation accuracy of 96.03%, but when integrated into the multi-biometric system, its effective contribution improved due to the added layer of security from ear biometrics. Similarly, the ear biometric mode that achieves 90.09% accuracy independently, benefited from ECG integration as physiological-based authentication adds a dynamic verification layer. This cross-reinforcement reduces the likelihood of false positives and false negatives, leading to a more robust authentication system. The ear biometric model also demonstrated strong classification confidence, distinguishing genuine from imposter samples with high probability. By combining both modalities, the overall authentication process benefits from increased precision and resilience against spoofing attacks, which are more likely to succeed in unimodal systems. The proposed system is compared against state-of-the-art approaches leveraging different biometric combinations. Several studies have explored the fusion of different biometric traits, including face, voice, iris, fingerprint, ear, and ECG, to improve recognition performance. Table 11 presents a comparative analysis of our proposed system with recent multi-biometric models.

The authors [45], [46] achieved the highest accuracy of 99.00% and 100%, respectively, by utilizing fusion of Swin transformer and CNN-based, or fusion various LBP feature extraction methods with feature selection and fusion, highlighting the effectiveness of learning techniques in multi-biometric authentication based on three modalities. Similarly, Guven *et al.* [47] employed a 1D-CNN for feature extraction, reaching an accuracy of 100% but with limited dataset. The proposed framework [44] applied pre-processing techniques for noise removal and quality enhancement, followed by HRV feature extraction from ECG signals in both the time and frequency domains. Then, ensemble of CNN and DeepResNet was employed for classification and achieved 97.00%. The authors [17], [48] have exploited CNN and handcrafted features with ensemble feature extraction to achieve good results 99.80% and 95.65%, respectively. However, most of previous multimodal systems have been evaluated on smaller datasets. The proposed dual-modal system leverages a large-scale dataset with promising result of 98.06%, ensuring better generalization. Additionally, the model maintains a balance between computational efficiency and accuracy while VGG-verydeep16 delivered superior accuracy, CNN5 offered a more balanced performance with faster computation time and reliability in real-world applications. Security and robustness are further reinforced by integrating physiological and morphological biometric traits, reducing vulnerability to attacks. While the proposed system demonstrates superior performance, several areas for improvement remain.

Table 11. Comparative analysis of multi-biometric (ECG+Ear) authentication systems

Study	Biometric modalities	Feature extraction method	Fusion approach	Dataset size/ subject	Accuracy (%)
Kailas *et al.* [44]	ECG + Iris	HRV + CNN + DeepResNet	Feature-level fusion	Not specified	97.00
Garg *et al.* [45]	Fingerprint + Iris + ECG	Swin transformer + CNN	Feature-level fusion	Combined multimodal dataset	99.00
Regouid *et al.* [46]	ECG + Ear + Iris	1D-LBP, Shifted 1D-LBP, 1D-MR-LBP	Feature-level fusion	Multiple public datasets	100
Guven *et al.* [47]	ECG + Speech	1D-CNN	Voting-based fusion	90 subjects	100
Sumalatha *et al.* [17]	ECG + Fingerprint + FKP	Siamese neural networks	Weighted score-level fusion	90 subjects	99.80
Ashwini *et al.* [48]	ECG + Iris	Ensemble feature extraction	Feature-level fusion	Not specified	95.65
Proposed model	ECG + Ear	CNN (VGG-verydeep16 with CNN5)	Feature-level fusion	Subject: 90 ECG samples: 1000, Ear samples 700	98.06

6. CONCLUSION

This study has successfully demonstrated the effectiveness of integrating ECG signals and ear features into a robust dual-modal biometric authentication system. By leveraging the complementary nature of these modalities with ECG providing dynamic physiological patterns that are difficult to forge and ear biometrics offering stable structural features. Our comprehensive evaluation of deep learning architecture revealed that while VGG-verydeep16 delivered superior accuracy, CNN5 offered a more balanced performance with faster computation time. The proposed hybrid system adaptively employs either model based on dataset size, effectively addressing the trade-off between accuracy and processing speed. This model not only enhances security through multi-layered verification but also demonstrates resilience against spoofing attacks and environmental variations that typically compromise unimodal systems. Future work will focus on implementing adaptive ECG filtering techniques, developing pose-invariant ear recognition methods, and validating the system across diverse demographic groups. As biometric authentication continues to evolve, this research provides a significant contribution toward more secure, efficient, and reliable access control solutions applicable across various domains including healthcare, finance, and secure facilities.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohamed S. Khalaf	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			
Said Fathy Al-Zoghdy	✓	✓		✓	✓	✓		✓		✓		✓		
Mariana Barsoum	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓		
Ibrahim Omara		✓		✓	✓	✓	✓		✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

[1] J. Ashbourn, *Guide to biometrics for large-scale systems*. London, United Kingdom: Springer London, 2011, doi: 10.1007/978-0-85729-467-8.

[2] O. Aiadi, B. Khaldi, A. Korichi, M. Chaa, M. B. Bezziane, and I. Omara, “Fusion of deep and local gradient-based features for multimodal finger knuckle print identification,” *Cluster Computing*, vol. 27, no. 6, pp. 7541–7557, Sep. 2024, doi: 10.1007/s10586-024-04352-3.

[3] I. Omara and R. F. Soliman, “CIoEBT: cancelable internet of ear biometric things based – a novel deep metric learning approach,” *Expert Systems with Applications*, vol. 297, Feb. 2026, doi: 10.1016/j.eswa.2025.129439.

[4] N. Alay and H. H. Al-Baity, “Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits,” *Sensors*, vol. 20, no. 19, 2020, doi: 10.3390/s20195523.

[5] A. Abozaid, A. Haggag, H. Kasban, and M. Eltokhy, “Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion,” *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16345–16361, Jun. 2019, doi: 10.1007/s11042-018-7012-3.

[6] M. S. Khalaf, S. F. El-Zoghdy, M. Barsoum, and I. Omara, “A novel multimodal biometric system based on deep fusion of ECG and ear,” *Journal of Flow Visualization and Image Processing*, vol. 31, no. 2, pp. 53–76, 2024, doi: 10.1615/JFlowVisImageProc.2024051591.

[7] K. Vishi and S. Y. Yayilgan, “Multimodal biometric authentication using fingerprint and iris recognition in identity management,” in *2013 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2013, pp. 334–341, doi: 10.1109/IIH-MSP.2013.91.

[8] S. S. Patel *et al.*, “Enhancing biometric authentication through score-level fusion of gait and palm vein modalities,” in *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems*, Aug. 2024, pp. 1–7, doi: 10.1109/IACIS61494.2024.10721667.

[9] A. Guesmi, M. A. Hanif, B. Ouni, and M. Shafique, “Physical adversarial attacks for camera-based smart systems: current trends, categorization, applications, research challenges, and future outlook,” *IEEE Access*, vol. 11, pp. 109617–109668, 2023, doi: 10.1109/ACCESS.2023.3321118.

[10] I. Omara, A. Hagag, S. Chaib, G. Ma, F. E. A. El-Samie, and E. Song, “A hybrid model combining learning distance metric and dag support vector machine for multimodal biometric recognition,” *IEEE Access*, vol. 9, pp. 4784–4796, 2021, doi: 10.1109/ACCESS.2020.3035110.

[11] M. Hammad, P. Pławiak, K. Wang, and U. R. Acharya, “ResNet-attention model for human authentication using ECG signals,” *Expert Systems*, vol. 38, no. 6, Sep. 2021, doi: 10.1111/exsy.12547.

[12] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, “EarEcho: using ear canal echo for wearable authentication,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–24, 2019, doi: 10.1145/3351239.

[13] Z. Wang *et al.*, “Computing with smart rings: a systematic literature review,” in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2025, pp. 1–54, doi: 10.1145/3749480.

[14] C. K. H. Ne, J. Muzaffar, A. Amlani, and M. Bance, “Hearables, in-ear sensing devices for bio-signal acquisition: a narrative review,” *Expert Review of Medical Devices*, vol. 18, pp. 95–128, Dec. 2021, doi: 10.1080/17434440.2021.2014321.

[15] S. A. El-Rahman and A. S. Alluhaidan, “Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments,” *PLOS ONE*, vol. 19, no. 2, Feb. 2024, doi: 10.1371/journal.pone.0291084.

[16] Y. Mohamed, Z. Youssef, A. Heakl, and A. B. Zaky, “Advancing ear biometrics: enhancing accuracy and robustness through deep learning,” in *2024 Intelligent Methods, Systems, and Applications*, Jul. 2024, pp. 437–442, doi: 10.1109/IMSA61967.2024.10652851.

[17] U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak, “Multimodal biometric authentication: a novel deep learning framework integrating ECG, fingerprint, and finger knuckle print for high-security applications,” *Engineering Research Express*, vol. 7, no. 1, Mar. 2025, doi: 10.1088/2631-8695/ad9aa0.

[18] S. Tammina, “Transfer learning using VGG-16 with deep convolutional neural network for classifying images,” *International Journal of Scientific and Research Publications*, vol. 9, no. 10, Oct. 2019, doi: 10.29322/IJSRP.9.10.2019.p9420.

[19] A. V. Jonnalagadda, H. A. Hashim, and A. Harris, “Comprehensive and comparative analysis between transfer learning and custom built VGG and CNN-SVM models for wildfire detection,” in *6th International Conference on Intelligent Computing in Data Sciences*, 2024, doi: 10.1109/ICDS62089.2024.10756303.

[20] M. Hammad, A. M. Iliyasu, I. A. Elgendy, and A. A. A. El-Latif, “End-to-end data authentication deep learning model for securing IoT configurations,” *Human-centric Computing and Information Sciences*, vol. 12, 2022, doi: 10.22967/HCIS.2022.12.004.

[21] M. Hammad and K. Wang, “Parallel score fusion of ECG and fingerprint for human authentication based on convolution neural network,” *Computers & Security*, vol. 81, pp. 107–122, Mar. 2019, doi: 10.1016/j.cose.2018.11.003.

[22] P. Melzi, R. Tolosana, and R. V. Rodriguez, “ECG biometric recognition: review, system proposal, and benchmark evaluation,” *IEEE Access*, vol. 11, pp. 15555–15566, 2023, doi: 10.1109/ACCESS.2023.3244651.

[23] I. El Boujnouni, H. Zili, A. Tali, T. Tali, and Y. Laaziz, “A wavelet-based capsule neural network for ECG biometric identification,” *Biomedical Signal Processing and Control*, vol. 76, 2022, doi: 10.1016/j.bspc.2022.103692.

[24] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, “Biometric identification from ECG signals using fourier decomposition and machine learning,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, doi: 10.1109/TIM.2022.3199260.

[25] T. R. S. Kumar and T. L. Purushottama, “ECG authentication reinforced by pretrained convolutional neural networks,” *2023 International Conference on Network, Multimedia and Information Technology (NMITCON),* Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/NMITCON58196.2023.10275834.

[26] A. J. Prakash, K. K. Patro, S. Samantray, P. Pławiak, and M. Hammad, “A deep learning technique for biometric authentication using ECG beat template matching,” *Information*, vol. 14, no. 2, 2023, doi: 10.3390/info14020065.

[27] K. K. Patro, A. J. Prakash, M. J. Rao, and P. R. Kumar, “An efficient optimized feature selection with machine learning approach for ECG biometric recognition,” *IETE Journal of Research*, vol. 68, no. 4, pp. 2743–2754, 2022, doi: 10.1080/03772063.2020.1725663.

[28] I. Omara, F. Li, H. Zhang, and W. Zuo, “A novel geometric feature extraction method for ear recognition,” *Expert Systems with Applications*, vol. 65, pp. 127–135, 2016, doi: 10.1016/j.eswa.2016.08.035.

[29] I. Omara, G. Ma, and E. Song, “LDM-DAGSVM: learning distance metric via dag support vector machine for ear recognition problem,” in *2020 IEEE International Joint Conference on Biometrics*, 2020, pp. 1–9, doi: 10.1109/IJCB48548.2020.9304871.

[30] I. Omara, X. Wu, H. Zhang, Y. Du, and W. Zuo, “Learning pairwise SVM on hierarchical deep features for ear recognition,” *IET Biometrics*, vol. 7, no. 6, pp. 557–566, Nov. 2018, doi: 10.1049/iet-bmt.2017.0087.

[31] X. Xu, S. Cao, and L. Lu, “Ear recognition based on residual network,” *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2021)*, vol. 89, pp. 971–979, 2022, doi: 10.1007/978-3-030-89698-0_99.

[32] H. Li, W. Zhao, Y. Zhang, and E. Zio, “Remaining useful life prediction using multi-scale deep convolutional neural network,” *Applied Soft Computing*, vol. 89, Apr. 2020, doi: 10.1016/j.asoc.2020.106113.

[33] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, “Deep convolutional neural networks for unconstrained ear recognition,” *IEEE Access*, vol. 8, pp. 170295–170310, 2020, doi: 10.1109/ACCESS.2020.3024116.

[34] H. Alshazly, C. Linse, E. Barth, S. A. Idris, and T. Martinetz, “Towards explainable ear recognition systems using deep residual networks,” *IEEE Access*, vol. 9, pp. 122254–122273, 2021, doi: 10.1109/ACCESS.2021.3109441.

[35] P. He, H. Li, H. Wang, and R. Zhang, “Detection of computer graphics using attention-based dual-branch convolutional neural network from fused color components,” *Sensors*, vol. 20, no. 17, Aug. 2020, doi: 10.3390/s20174743.

[36] R. A. Priyadharshini, S. Arivazhagan, and M. Arun, “A deep learning approach for person identification using ear biometrics,” *Applied Intelligence*, vol. 51, no. 4, pp. 2161–2172, Apr. 2021, doi: 10.1007/s10489-020-01995-8.

[37] R. Mehta, A. S. Akbari, and K. K. Singh, “Ensemble-based hybrid transfer approach for an effective 2D ear recognition system,” *IEEE Access*, vol. 12, pp. 155733–155746, 2024, doi: 10.1109/ACCESS.2024.3485514.

[38] Y. Lei, D. Pan, Z. Feng, and J. Qian, “Lightweight human ear recognition based on attention mechanism and feature fusion,” *Applied Sciences*, vol. 13, no. 14, 2023, doi: 10.3390/app13148441.

[39] J. S. A.- Falconi, H. Al Osman, and A. E. Saddik, “ECG and fingerprint bimodal authentication,” *Sustainable Cities and Society*, vol. 40, pp. 274–283, 2018, doi: 10.1016/j.scs.2017.12.023.

[40] I. Odinaka *et al.*, “ECG biometric recognition : a comparative analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, 2012.

[41] S. M. Almabdy and L. A. Elrefaei, “An overview of deep learning techniques for biometric systems,” *Studies in Computational Intelligence*, vol. 912, pp. 127–170, 2021, doi: 10.1007/978-3-030-51920-9_8.

[42] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. T. Ahmed, “Past, present, and future of face recognition: a review,” *Electronics*, vol. 9, no. 8, pp. 1–53, 2020, doi: 10.3390/electronics9081188.

[43] S. Tharewal *et al.*, “Score-level fusion of 3D face and 3D ear for multimodal biometric human recognition,” *Computational Intelligence and Neuroscience*, pp. 1–9, Apr. 2022, doi: 10.1155/2022/3019194.

[44] A. Kailas, M. Girimallaih, M. Madigahalli, V. K. Mahadevachar, and P. K. Somashekarappa, “Ensemble of convolutional neural network and DeepResNet for multimodal biometric authentication system,” *International Journal of Electrical and Computer Engineering*, vol. 15, no. 4, pp. 4279–4295, Aug. 2025, doi: 10.11591/ijece.v15i4.pp4279-4295.

[45] R. Garg, P. Pathak, and M. P. Singh, “A multimodal biometric recognition system based on fingerprints, iris and ECG via swin transformer and CNN model,” *Systems and Soft Computing*, vol. 7, Dec. 2025, doi: 10.1016/j.sasc.2025.200369.

[46] M. Regouid, M. Touahria, M. Benouis, and N. Costen, “Multimodal biometric system for ECG, ear and iris recognition based on local descriptors,” *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22509–22535, 2019, doi: 10.1007/s11042-019-7467-x.

[47] G. Guven, U. Guz, and H. Gürkan, “A novel biometric identification system based on fingertip electrocardiogram and speech signals,” *Digital Signal Processing*, vol. 121, Mar. 2022, doi: 10.1016/j.dsp.2021.103306.

[48] K. Ashwini, G. N. K. Murthy, S. Raviraja, and G. A. Srinidhi, “A novel multimodal biometric person authentication system based on ECG and iris data,” *BioMed research international*, 2024, doi: 10.1155/2024/8112209.

[49] T. S. Lugovaya, “Biometric human identification based on electrocardiogram,” *M.Sc. thesis*, Department of Applied Mathematics and Computer Science, Electrotechnical University “LETI”, Saint-Petersburg, Russian Federation, 2005.

[50] A. L. Goldberger *et al.*, “PhysioBank, physioToolkit, and physioNet,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, doi: 10.1161/01.CIR.101.23.e215.

[51] A. Jain *et al.*, “Insect identification in the wild: the AMI dataset,” in *Computer Vision (ECCV 2024)*, Cham, Switzerland: Springer, 2025, pp. 55–73, doi: 10.1007/978-3-031-72913-3_4.

[52] E. Gonzalez, L. Alvarez, and L. Mazorra, "AMI ear database," *Universidad de Las Palmas de Gran Canaria*, 2025. [Online]. Available: https://ctim.ulpgc.es/research_works/ami_ear_database/
[53] L. Sathyapriya, L. Murali, and T. Manigandan, "Analysis and detection R-peak detection using modified pan-Tompkins algorithm," in *2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies*, 2015, pp. 483–487, doi: 10.1109/ICACCCT.2014.7019490.
[54] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 38, no. 1, pp. 35–44, Aug. 2004, doi: 10.1023/B:VLSI.0000028532.53893.82.

BIOGRAPHIES OF AUTHORS

Mohamed S. Khalaf is currently a Ph.D. student in Artificial Intelligence. He received his B.Sc. degree in Mathematics and Computer Science from Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt. His research interests include artificial intelligence, machine learning, biometric authentication systems, and signal and image processing. He can be contacted at email: msoror429@gmail.com.

Prof. Dr. Said Fathy Al-Zoghdy received his B.Sc. (Hons.) and M.Sc. degrees in Computer Science from the Faculty of Science, Menoufia University, Egypt, in 1993 and 1997, respectively. He obtained his Ph.D. in 2004 from University of Tsukuba, Japan. At Menoufia University, he has held successive academic positions, serving as a demonstrator, assistant lecturer, lecturer, associate professor, and since 2018 as a professor in the Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt. He has also served as a professor at Taif University, Saudi Arabia, and held a postdoctoral position at the School of System Informatics, Kobe University, Japan. His research interests include distributed and parallel computing, swarm intelligence, cloud computing, machine learning, forgery detection, security, and cryptography, with particular emphasis on load balancing in distributed systems, modeling and simulation, grid computing, and the design and analysis of parallel algorithms. He can be contacted at email: elzoghdy@science.menofia.edu.eg.

Dr. Mariana Barsoum is a lecturer at Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt. Her research areas of interest include visualization, bioinformatic, artificial intelligent, machine learning, deep learning, image processing, and forgery detection. She can be contacted at email: mariana_barsoum@yahoo.com.

Assoc. Prof. Dr. Ibrahim Omara is received the B.Sc. degree (Hons.) in Mathematics and Computer Science in July 2005, and the M.Sc. degree in Computer Science in May 2012 from Menoufia University, Egypt. He obtained the Ph.D. degree in Computer Science from the School of Computer Science and Technology, Harbin Institute of Technology (HIT), China, in October 2018. From April 2007 to October 2018, he served as a demonstrator and assistant lecturer at the Faculty of Science, Menoufia University. He also held a postdoctoral position at Huazhong University of Science and Technology, Wuhan, China. He is currently an associate prof. in the Department of Machine Intelligence, Faculty of Artificial Intelligence, Shebin El-Kom, Menoufia University. His research interests include computer vision, machine learning, particularly metric learning algorithms—and biometrics, and a focus on ear and face recognition. He can be contacted at email: i_omara84@ai.menofia.edu.eg.