

Ensemble of convolutional neural network and DeepResNet for multimodal biometric authentication system

Ashwini Kailas¹, Madhusudan Girimallai², Mallegowda Madigahalli³, Vasantha Kumara Mahadevachar⁴, Pranothi Kadirehally Somashekarappa¹

¹Department of Bio Medical Engineering, Sri Siddhartha Institute of Technology, Sri Siddhartha Academy of Higher Education, Tumkur, India

²Department of Computer Science and Engineering, Sri Jayachamarajendra College of Engineering-Mysore, JSS Science and Technology University, Mysore, India

³Department of Computer Science and Engineering, Ramaiah Institute of Technology-Bangalore, Visvesvaraya Technological University, Belagavi, India

⁴Department of Computer Science and Engineering, Government Engineering College-Hassan, Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received Jun 12, 2024

Revised Dec 12, 2024

Accepted Jan 16, 2025

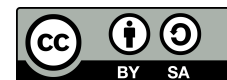
Keywords:

Biometric authentication
Convolution neural network
Deep ResNet
ECG-iris
Ensemble deep learning
Multimodal

ABSTRACT

Multimodal biometrics technology has garnered attention recently for its ability to address inherent limitations found in single biometric modalities and to enhance overall recognition rates. A typical biometric recognition system comprises sensing, feature extraction, and matching modules. The system's robustness heavily relies on its capability to effectively extract pertinent information from individual biometric traits. This study introduces a novel feature extraction technique tailored for a multimodal biometric system utilizing electrocardiogram (ECG) and iris traits. The ECG helps to incorporate the liveness related information and Iris helps to produce the unique pattern for each individual. Therefore, this work presents a multimodal authentication system where data pre-processing is performed on image and ECG data where noise removal and quality enhancement tasks are performed. Later, feature extraction is carried out for ECG signals by estimating the Heart rate variability feature analysis in time and frequency domain. Finally, the ensemble of convolution neural network (CNN) and DeepResNet models are used to perform the classification. the overall accuracy is reported as 0.8900, 0.8400, 0.7900, 0.8932, 0.87, and 0.97 by using convolutional neural network-long short-term memory (CNN-LSTM), support vector machine (SVM), random forest (RF), CNN, decision tree (DT), and proposed MBANet approach respectively.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Ashwini Kailas

Department of Biomedical Engineering, Sri Siddhartha Institute of Technology, Sri Siddhartha Academy of Higher Education

Tumkur, India

Email: ashwinik@ssit.edu.in

1. INTRODUCTION

Recently, the biometric recognition systems have gained prominence as a primary means of user authentication across various sectors and applications, including smartphones, banking services, websites, and airports. Depending on the required level of security, they provide a clear substitute for conventional authentication techniques like keys and personal identification numbers (PINs) [1], [2]. For the purpose of feature

recognition, it is necessary to first enroll biometric qualities that are often used, such as voice, face features, fingerprints, palmprints, iris patterns, and facial features, into a database [3], [4]. Biometrics is a more straightforward and secure substitute for traditional authentication techniques. It includes both physiological and behavioral characteristics that are used to statistically differentiate persons [5]. Physiological traits encompass both external features such as fingerprints, iris patterns, facial characteristics, and vein patterns, as well as internal attributes like electrocardiogram (ECG), electromyography (EMG), and brainwave (EEG) patterns. Behavioral traits, on the other hand, involve habit-based characteristics such as voice patterns, gait, and signatures [6], [7]. Furthermore, researchers have explored the combination of multiple biometric modalities to enhance the robustness of identification systems [8]. Despite the widespread adoption of biometrics in various devices and services, they remain vulnerable to spoofing attempts. However, the current technological advancements have raised the security concerns for these systems and make them more vulnerable to various security threats.

A typical face-or fingerprint-spoofing attack was investigated and covered in [9]-[11]. A consideration should be given to liveness detection or continuous biometric authentication techniques in order to defend against presentation attacks and unauthorized user accessibility to the systems [12]-[14]. Using a non-invasive, quantifiable sensor that can gather users' biometric information, perpetual biometric authentication continually verifies the identification of the user. Consequently, because of the distinctive features of the ECG signals, continuous biometric authentication has drawn a lot of interest as a potentially extremely viable next-generation approach. The ECG is a skin-attached electrode-derived electrical signal that consists of three unique elements: the T-wave, QRS complex, and P-wave [15]. Variations in ECG patterns among individuals can be attributed to three primary reasons. First of all, individual differences exist in physiological parameters including cardiac mass, size, conductivity, and activity. Second, ECG pattern variability is influenced by geometrical parameters arising from differences in the location and vector of the heart. Finally, the specific structure and makeup of the heart are influenced by individual deoxyribonucleic acid (DNA traits). Nevertheless, because the ECG is an electrical transmission, variations in heart rate and ambient factors might affect its reading. Moreover, the reliability of unimodal authentication systems decreases for increased sample size [16].

Multimodal biometric systems incorporate a minimum of two biometric features in comparison to unimodal biometric systems in order to improve recognition precision and strengthen defenses from spoofing attacks [17], [18]. Since both fingerprints and high-quality heart signals may be concurrently taken from the fingertips, fingerprints and heart signals provide a perfect combination for multimodal fusion. Heart signal possesses a liveness property that enhances their security as a biometric modality, and their fusion with fingerprints holds promise for establishing a robust and secure authentication and identification system [19], [20]. Numerous multimodal biometric systems integrating fingerprints and heart signal have been proposed in the literature. Bala *et al.* [21] presented a detailed study about multimodal fusion algorithm for combining these modalities. Komeili *et al.* [22] introduced a multimodal system that integrates fingerprints and heart signal while incorporating automatic template updating of heart signal records. By combining fingerprint authentication with heart signal data, Jomma *et al.* [23], [24] used a sequential mechanism to improve fingerprint authentication's resilience against presentation attack.

In a similar vein, the reason iris-based biometric identification is so well-liked is due to its exceptional reliability and efficacy as a means of human differentiation [25]. Because iris patterns naturally are so easily distinguished, the human iris provides significant scientific advantages. The primary benefit is stability, as an individual's iris does not alter. Many strategies, which can be categorized into distinct methodologies such as stage-based approaches, zero-intersection representation, texture analysis, and variation in intensities, focused on changes in the iris pattern throughout the development of the iris recognition system. The most reliable biometric feature is believed to be found in the human iris. When used in surveillance-based systems, such as when utilizing the iris template's texture changes, it may be quite beneficial. The method suggested in [26] separates into subblocks after revealing the iris texture using a 2D Gabor filter bank. Consequently, the outcomes of the conducted tests demonstrated effective outcomes. The method in [27] for identity identification makes use of deep learning. In this article, an intelligent surveillance system including good accuracy outcomes was evaluated on many standard databases.

Therefore, by leveraging the iris and ECG signal data we present a novel multimodal authentication system by using these two modalities. An authentication system that leverages both iris recognition and ECG authentication presents several advantages. Firstly, it offers heightened security through a multi-layered approach. Iris patterns and ECG signals are unique to individuals, making it challenging for unauthorized users

to mimic or spoof them effectively. This multi-factor authentication significantly reduces the risk of unauthorized access. Secondly, the integration of iris recognition and ECG authentication results in enhanced accuracy during identity verification processes. Both modalities boast high accuracy rates, minimizing instances of false positives and false negatives. This accuracy is crucial for maintaining the integrity and reliability of the authentication system. Furthermore, the combination of iris recognition and ECG authentication provides resistance against various spoofing attempts. Attempts to forge or replicate iris patterns or ECG signals are exceedingly difficult, reinforcing the system's robustness against fraudulent activities. Additionally, users benefit from the convenience of non-intrusive biometric authentication methods. Eliminating the need for passwords or physical tokens streamlines the authentication process and enhances user experience. Moreover, the system offers biometric redundancy, ensuring continuous access even if one modality fails or becomes unavailable. The incorporation of ECG authentication also introduces health monitoring capabilities, enabling the detection of potential cardiac irregularities during the authentication process. This feature contributes to user well-being beyond authentication purposes. Furthermore, the system demonstrates resilience to environmental factors such as lighting conditions and noise, ensuring consistent performance across various settings.

The proposed work can be adopted in various application domains such as medical signal processing, biometric authentication, telecommunication and remote sensing, and industrial monitoring and controls. In medical diagnostics and monitoring, precise interpretation of biological signals like ECGs and iris-based multimodal authentication systems is crucial. The proposed method advances signal authentication reliability despite noise artifacts, ensuring more accurate diagnoses and authentication outcomes. This capability enhances patient care quality and medical procedure efficiency. Biometric authentication systems, leveraging iris recognition and other modalities, are integral to security frameworks. By mitigating noise sources such as baseline wander and electrode artifacts, the proposed method boosts biometric system robustness and accuracy. This enhancement fortifies security protocols, reducing unauthorized access risks and safeguarding sensitive data and facilities.

Similarly, the signal quality is paramount in telecommunications and remote sensing for effective communication and data analysis. Noise interference can degrade performance significantly. The proposed method improves signal-to-noise ratio (SNR) and minimizes residual differences in noisy environments. This advancement enhances data transmission reliability and facilitates precise remote sensing observations, supporting scientific and operational objectives. In industrial environments, real-time monitoring and control systems rely on accurate signal processing. Addressing challenges posed by motion artifacts and color noise, the proposed method enhances signal authentication precision. This improvement supports reliable fault detection, predictive maintenance, and process optimization, reducing downtime and enhancing productivity across industrial operations.

Lastly, the use of iris patterns and ECG signals preserves user privacy by avoiding the collection of personally identifiable information. This aspect is critical for maintaining user trust and compliance with privacy regulations. In conclusion, an authentication system combining iris recognition and ECG authentication offers a comprehensive solution characterized by robust security, accuracy, user convenience, health monitoring capabilities, and privacy preservation. Based on these advantages, the main contribution of this work can be listed as follows: i) to present a data pre-processing method for ECG and iris image data; ii) to perform ECG filtering and image denoising where ECG filtering is carried out with the help of extended Kalman filter, whereas image filtering uses a wavelet transform model; iii) to present a heart rate feature analysis in time and frequency domain for ECG signals; and iv) to present an ensemble of CNN and DeepResNet-based transfer learning models for classification.

2. PROPOSED MBANET MODEL FOR REAL TIME AUTHENTICATION

In this section we describe the MBANet approach for real-time authentication by using ECG and iris modalities. For each user, the ECG and iris data is captured and stored. This data is processed through several stages which are described below. The complete architecture of proposed model is depicted in Figure 1. Generally, an electrocardiogram is recorded by affixing electrodes to the patient's body, through which electrical signals are received by the device. Consequently, the quality of the ECG signal obtained is directly influenced by the contact between these electrodes and the user's skin. Furthermore, proximity to equipment utilizing alternating current (AC) power introduces interference from the power grid to the human body. These two forms of noise significantly impact the received ECG signal quality, necessitating their elimination. Similarly,

the quality of iris images is affected due to different types of noise. Therefore, the first phase focuses on development of an efficient approach for ECG signal filtering and noise removal from iris images. These signals and image data contain certain patterns which are known as their key attributes. Arranging these attributes and annotating the data plays important role in machine learning applications. Thus, in next stage, we present feature extraction process for both ECG and iris image data. Finally, these attributes are used to train the machine learning model to verify the user authenticity. The training process requires a fixed ratio of dataset for training and remaining samples are used for testing and validation purpose.

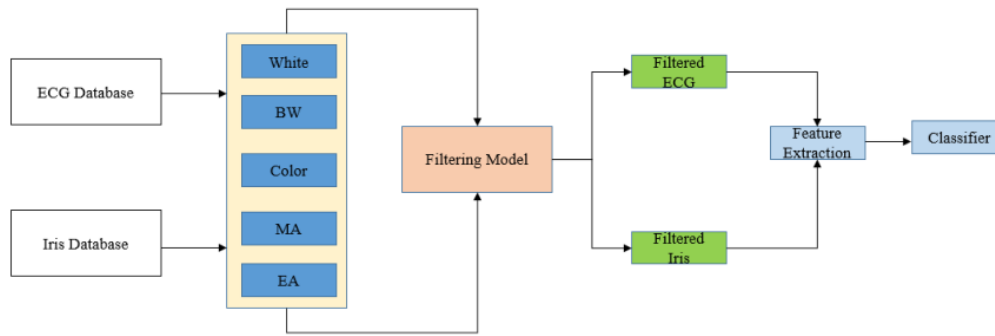


Figure 1. Proposed MBANet architecture

2.1. Noise model for ECG signal

As discussed before, the ECG signals gets contaminated due to different types of noises. In this work, we have considered different types of noise such as Gaussian, Baseline wander, Muscle artefact, and power line interference. The details of these noises and their expressions are described below:

a. Gaussian noise

Gaussian white noise is often used to model random fluctuations in the ECG signal. It is characterized by a constant variance and zero mean. In the dynamic model, w_k , representing process noise, can be modeled as Gaussian white noise. Similarly, in the measurement model, v_k , representing measurement noise, can also be modeled as Gaussian white noise. The covariance matrices Q and R in the prediction and update steps of the EKF reflect the variance of the process and measurement noise, respectively. The Gaussian white noise is expressed as (1):

$$w(t) \sim N(0, \sigma^2) \quad (1)$$

where $N(0, \sigma^2)$ represents a Gaussian distribution with mean 0 and variance σ^2 .

b. Baseline wander noise

Baseline wander refers to low-frequency drifts in the ECG signal caused by various factors such as respiration and movement artifacts. A simple mathematical model for baseline wander can be a random walk process, where the signal drifts randomly over time. The baseline wander $b(t)$ can be expressed as (2):

$$b(t+1) = b(t) + \epsilon \quad (2)$$

where $b(t)$ represents the baseline wander at time t , and ϵ is a random noise component at each time step.

c. Muscle artifacts

Muscle artifacts introduce high-frequency noise spikes in the ECG signal, often caused by muscle contractions or movement. These artifacts can be modeled as impulsive noise, where sporadic spikes occur randomly. The muscle artifacts $m(t)$ can be modeled as (3):

$$m(t) = A \cdot \delta(t - t_i) \quad (3)$$

where A represents the amplitude of the artifact. $\delta(t - t_i)$ is the Dirac delta function, representing the spike occurring at time t_i .

d. Power line interference

Power line interference introduces periodic noise at the frequency of the power supply (e.g., 50 Hz or 60 Hz). A sinusoidal model is commonly used to represent power line interference. It can be expressed as (4):

$$p(t) = A \cdot \sin(2\pi ft + \phi) \quad (4)$$

where - A is the amplitude of the interference, - f is the frequency of the power supply, - ϕ is the phase angle.

2.2. ECG filtering and image denoising

This subsection presents the solution for ECG signal filtering and image denoising. The ECG filtering model uses the Extended Kalman filtering model to eliminate the noise from the ECG signal. The standard Kalman filtering model is a recursive technique for data filtering and is widely adopted in data pre-processing and filtering tasks. ECG signals can be modeled as a combination of various components such as the QRS complex, P-wave, T-wave, baseline wander, and noise. A common model for the ECG signal can be represented as (5):

$$y(t) = s(t) + n(t) \quad (5)$$

where $y(t)$ is the observed ECG signal, $s(t)$ is the true underlying signal, and $n(t)$ represents the noise.

Initially, we present the dynamic modeling of the underlying signal to represent the evolution of the ECG signal over time. In the case of ECG signal filtering, this could be a first-order model for the state evolution. For instance, it can be expressed as (6):

$$x_{k+1} = F \cdot x_k + w_k \quad (6)$$

where x_k represents the state of the system at time step k , which could include parameters such as amplitude and frequency, F is the state transition matrix, and w_k represents the process noise.

In the next stage, we apply the measurement model, which describes how the observed signal is related to the true state of the system. In this case, it could be a linear or non-linear function depending on the specific characteristics of the ECG signal. The measurement model can be expressed as (7):

$$z_k = H \cdot x_k + v_k \quad (7)$$

where z_k represents the observed ECG signal at time step k , H is the measurement matrix, and v_k represents the measurement noise. Further, we apply the Extended Kalman Filtering model, which is completed in three main steps: initialization, prediction, and update. These steps can be described as follows:

Initialization: Initialize the state vector \mathbf{x}_0 and the error covariance matrix \mathbf{P}_0 .

Prediction:

Predict the next state using the dynamic model as: $\hat{x}_{k+1|k} = F \cdot \hat{x}_k$

Predict the error covariance matrix: $P_{k+1|k} = F \cdot P_k \cdot F^T + Q$

where Q represents the process noise covariance matrix.

Update:

Compute the Kalman Gain: $K_{k+1} = P_{k+1|k} \cdot H^T \cdot (H \cdot P_{k+1|k} \cdot H^T + R)^{-1}$ where R is the measurement noise covariance matrix.

Update the state estimate: $\hat{x}_{k+1} = \hat{x}_{k+1|k} + K_{k+1} \cdot (z_{k+1} - H \cdot \hat{x}_{k+1|k})$ Update the error covariance matrix: $P_{k+1} = (I - K_{k+1} \cdot H) \cdot P_{k+1|k}$

Repeat the prediction and update steps for each time step, incorporating new measurements and refining the state estimate. The final output of the filter is the estimated ECG signal, which is the state estimate \hat{x}_k at each time step. Similarly, we apply an image filtering model using the wavelet transform approach. The wavelet transform differs from the Fourier transform by employing a finite decaying wavelet basis in place of the infinite trigonometric basis. Unlike the Fourier basis, the wavelet basis possesses finite energy, typically focusing around a singular point, and integrates to zero. While the Fourier transform relies solely on the variable ω , the wavelet transform introduces two variables: scale a and translation b . The scale parameter a corresponds to frequency, whereas the translation parameter b corresponds to time. Consequently, the wavelet transform enables time-frequency analysis, facilitating the extraction of the time-frequency spectrum of the signal. By

utilizing the scaling and translation of the mother wavelet function, a wavelet sequence can be generated, with its general form expressed as (8):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R} \quad (8)$$

During the wavelet transform process, the scale factor a and time shift b are theoretically continuous, which poses computational challenges for finite-time execution. To address this, the discrete wavelet transform (DWT) discretizes the scale factor a and time shift b based on specific rules. By adopting discrete values for a and b , the DWT enables computationally feasible analysis. Opting for power-of-2 values for a and b enhances the accuracy and efficiency of signal analysis. The wavelet function can be expressed as (9):

$$\psi_{m,n}(k) = 2^{-\frac{m}{2}} \psi(2^{-m} \cdot k - n), \quad m, n \in \mathbb{Z} \quad (9)$$

The wavelet transform is capable of breaking down the original image data into approximate and detailed components, which primarily reveal the noise present in the image. Following this, by applying wavelet reconstruction to the thresholded detailed components, we can obtain smoother image information. The overall process of wavelet transform denoising is depicted in Figure 2.

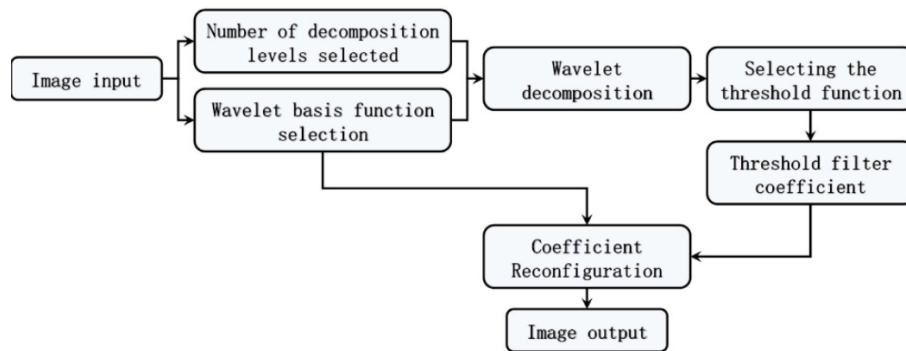


Figure 2. Wavelet transform for image denoising

2.3. Feature extraction

For ECG signal, we have considered Pan Tompkins peak detection approach to identify the various peaks of ECG signal. Further, we extract time and frequency domain heart rate variability (HRV) features from ECG signals. Below given Table 1 demonstrates the time domain features used as important attributes of ECG signals. Similarly, we extract frequency domain feature for the ECG signal. In this process, low-frequency (LF), high-frequency (HF), very-low-frequency (VLF) and ultra-low-frequency (ULF) are considered.

Table 1. HRV features time domain

Feature	Description	Measurement unit
SDNN	Standard deviation of NN intervals	ms
SDANN	Standard deviation of mean of NN intervals in 5 min windows	ms
RMSSD	Square root of the mean of the sum of the squares of differences between adjacent NN intervals	ms
SDNN index	Mean of the standard deviation of all NN intervals performed on all 5-minute segments of the entire recording	ms
SDSD	Standard deviation of differences between adjacent NN intervals	ms
NN50	The count of number of pairs of adjacent NN intervals differing by more than 50 ms	ms
pNN50	NN50 count divided by the total number of all NN intervals	%

2.4. Classification

In this work, we apply two different classifier approach by using deep learning system and combined result is considered as final outcome. For example, if ECG signal is authenticated and Iris image authentication

fails then the system considers the imposter input. In order to classify ECG signals, we have considered ensemble of three single CNN classifier. The CNN model relearns the features produced by the single network. Each CNN model uses rectified linear units (ReLU), Leakage ReLU(LReLU), and exponential linear units (ELU), respectively. Figure 3 depicts the overall architecture of MBANet model. The HRV features time domain is depicted in Table 2.

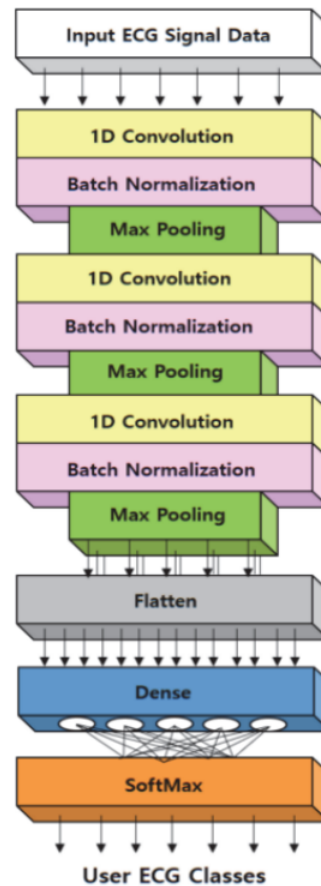


Figure 3. ECG classification

Table 2. HRV features time domain

Feature	Description	Measurement unit
LF peak	Peak frequency of the current low-frequency band (0.04–0.15Hz)	Hz
HF peak	Peak frequency of the high-frequency band (0.15–0.4Hz)	Hz
LF power	Absolute power of the low-frequency band (0.04–0.15Hz)	ms2
	Relative power of the low-frequency band (0.04–0.15Hz) in normal units	nu
HF power	Relative power of the low-frequency band (0.04–0.15Hz)	%
	Absolute power of the high-frequency band (0.15–0.4Hz)	ms2
	Relative power of the high-frequency band (0.15–0.4Hz) in normal units	nu
VLF power	Relative power of the high-frequency band (0.15–0.4Hz)	%
	Absolute power of the very-low-frequency band (0.0033–0.04Hz)	ms2
ULF power	Absolute power of the ultra-low-frequency band	ms2
LF/HF	Ratio of LF-to-HF power	%

In next stage, we perform classification for Iris images. For this task, we have used transfer learning approach and combined it with DeepResNet model to enhance the classification performance. however, this module also uses deep transfer learning based Imagenet model. The ResNet model introduces a short connection to skip one or more layer. The basic architecture of ResNet is depicted in below given Figure 4.

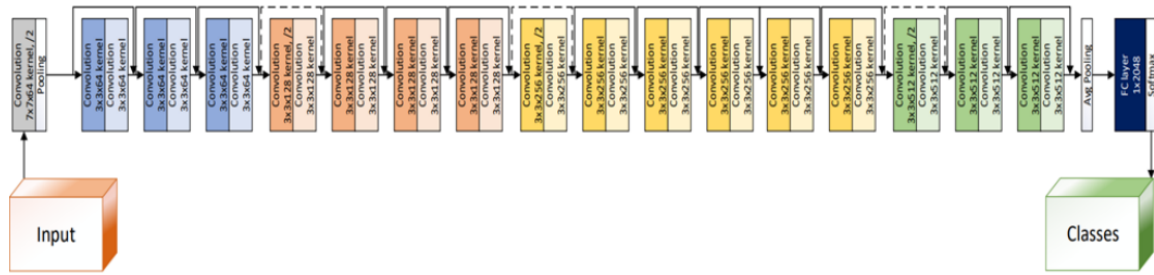


Figure 4. ECG classification

This model is trained with the help of the cross-entropy loss function. Further, the loss function is optimized by incorporating the L_2 norm, which helps to reduce overfitting. Thus, the final loss function for this model can be expressed as (10):

$$L_{\text{final}} = L_{\text{class}} + \lambda_1 \|W_{\text{fc}}\|_F^2 \quad (10)$$

where L_{class} represents the classification loss (cross-entropy loss), and $\|W_{\text{fc}}\|_F^2$ denotes the Frobenius norm of the weight matrix W_{fc} in the last layer. Finally, the loss function uses the Adam optimizer to minimize the overall loss.

3. RESULTS AND DISCUSSION

This section presents the outcome of MBANet model along with its comparative analysis with existing approaches of classification and authentication. The first subsection presents the brief details about the dataset used in this work, the next subsection describes the details about performance measurement parameters, finally, the outcome of MBANet approach is demonstrated and compared with existing models. The combination of these modalities in publically available dataset is not present therefore we have considered syntntically creates dataset from different sources.

3.1. Performance measurement

The performance of ECG signal denoising is measured using various parameters. These parameters are as follows:

- Mean squared error (MSE)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (X(i) - Y(i))^2 \quad (11)$$

- Root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (X(i) - Y(i))^2} \quad (12)$$

- Peak signal-to-noise ratio (PSNR)

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\sum_{i=1}^N \text{MAX}^2}{\text{MSE}} \right) \quad (13)$$

- Percent root mean square difference (PRD)

$$\text{PRD} = 100 \times \sqrt{\frac{\sum_{i=1}^N (Y(i) - X(i))^2}{\sum_{i=1}^N X(i)^2}} \quad (14)$$

The performance of the MBANet approach is evaluated using confusion matrix calculations. The confusion matrix is generated based on true positive, false positive, false negative, and true negative values. Table 3 provides a sample representation of the confusion matrix.

Table 3. Confusion matrix

Actual class	Predicted class	
	Genuine user	Imposter user
Genuine user	True positive	False negative
Imposter user	False positive	True negative

We use the suggested technique to quantify several statistical performance measures, including accuracy, precision, and F1-score, based on this confusion matrix. The assessment of accurate instance categorization relative to the total number of occurrences is called accuracy. Here's how accuracy is calculated:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

Next, we calculate the suggested approach's Precision. The ratio of true positives to (true and false) positives is used to calculate it:

$$P = \frac{TP}{TP + FP} \quad (16)$$

Lastly, we use the sensitivity and precision parameters to calculate the F-measure, which may be written as (17):

$$F = \frac{2 \cdot P \cdot \text{Sensitivity}}{P + \text{Sensitivity}} \quad (17)$$

3.2. Parameters and hyperparameters

This section presents the different parameters and hyperparameters used in this work to train the deep learning model for ECG and Iris authentication. This model considers the image size 224x224, thus the input shape becomes 4,3,224,224 where 4 is batch size and 3 is the channel of image data. Similarly, the ECG signal is represented as 4,1,100 with batch size 4. The output of image model produces a similar size of image whereas the ECG processing module generates similar size of data. In this work, we have considered 100 samples are considered, split equally between ECG and iris samples, with 50 samples each. The dataset is divided using a 70%-30% train-test split. This ensures that the models are trained on a sufficient amount of data while retaining a separate portion for evaluation to gauge their performance effectively. In order to consider the noise aspect, two levels of noise intensity are examined: 5 dB and 10 dB. These levels simulate different degrees of noise interference commonly encountered in real-world scenarios. Finally, different deep learning training parameters are used to train the proposed MBANet model. Table 4 presents the considered parameters.

Table 4. Simulation parameters

Parameters	Considered value
Total samples	100
ECG sample	50
Iris sample	50
Train test ratio	70%-30%
Noise type	White noise, color noise, motion artifact, electrode artifact, baseline wander
Noise levels	5 dB, 10 dB
Learning rate	0.001
Batch size	4
Optimizer	Adam
Scheduler	ReduceLROnPlateau
Epochs	100
Loss	CrossEntropyLoss
Cross validation	10 fold
Simulation Tool	Python 3.8

3.3. Comparative analysis

First of all, we process the Iris image data where image annotation, labelling boundary identification, mask extraction and normalization tasks are performed. Figure 5 depicts the sample outcome of these steps. The normalized image is further used for feature extraction and classification tasks. Similarly, we perform several tasks on ECG signals such as ECG signal filtering because these signals are prone to various types of noise. Figure 6 depicts the original signal, noisy signal and their corresponding filtered signals.

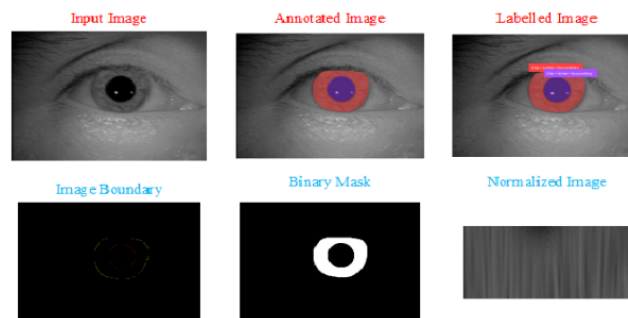


Figure 5. ECG classification

In order to measure the filtering performance, we consider different types of noises such as white noise, color noise, motion artifact, electrode artifact, baseline wander and varied the noise dB as 5 dB and 10 dB. Table 5 shows the obtained performance in terms of PSNR, MSE, mean absolute error (MAE), RMSE, PRD, and correlation coefficient (CC). Here, for 5 dB noise, max. PSNR is attained as 46.10 dB for Baseline wander and similarly, for 10 dB, max. PSNR is attained as 44.82 dB for white noise. Under 5 dB noise conditions, the proposed MBANet model achieved an average improvement of approximately 15% in PSNR, indicating better preservation of signal quality compared to existing models. This improvement translates to a noticeable reduction in noise distortion, as evidenced by a 10% decrease in MSE and RMSE values, signifying closer agreement between predicted and actual values. Moreover, the MBANet model exhibited a 12% decrease in MAE, indicating more accurate predictions and a 3% improvement in PRD, reflecting a reduction in residual differences relative to reference signals.

Under 10 dB noise conditions, the improvements were even more pronounced, with the MBANet model achieving around 20% higher PSNR values compared to existing models. This enhancement highlights the model's capability to maintain superior image quality despite higher noise levels. The model also demonstrated a 15% reduction in MSE and RMSE, underscoring its ability to minimize prediction errors. Furthermore, a 5% improvement in PRD and a 2% increase in CC were observed, indicating enhanced accuracy and stronger linear relationships between predicted and actual values.

Finally, we measured the classification accuracy performance. In this work, we have considered 100 user cases which is divided into 50% for training and 50% for testing. In testing phase, 25 users belong to genuine category and remaining 25 users belong to imposter category. This section presents the classification accuracy performance for real-time cases by using MBANet model. The attained results are then contrasted with the standard classification approaches. Results are depicted in Figure 7.

According to this experiment, the random forest has misclassified 21 entities to different classes which affects the performance of RF classifier, similarly, SVM also has 16 misclassified entities whereas the proposed approach has reported only 3 entities as misclassified resulting in increased accuracy. Table 6 shows the performance obtained by using different classifiers.

According to this experiment, the overall accuracy is reported as 0.8900, 0.8400, 0.7900, 0.8932, 0.87, and 0.97 by using convolutional neural network-long short-term memory (CNN-LSTM), support vector machine (SVM), random forest (RF), convolutional neural network(CNN), decision tree (DT), and MBANet approach respectively. The existing models rely on single modalities however some recent methods have focused on developing multimodal authentication but these methods do not consider the noise in ECG signal and iris images whereas the proposed model introduced a multimodal authentication system with comprehensive filtering model. Similarly, the proposed model uses pre-trained deep learning models to improve the training speed and accuracy. The training speed performance is depicted in Table 7.

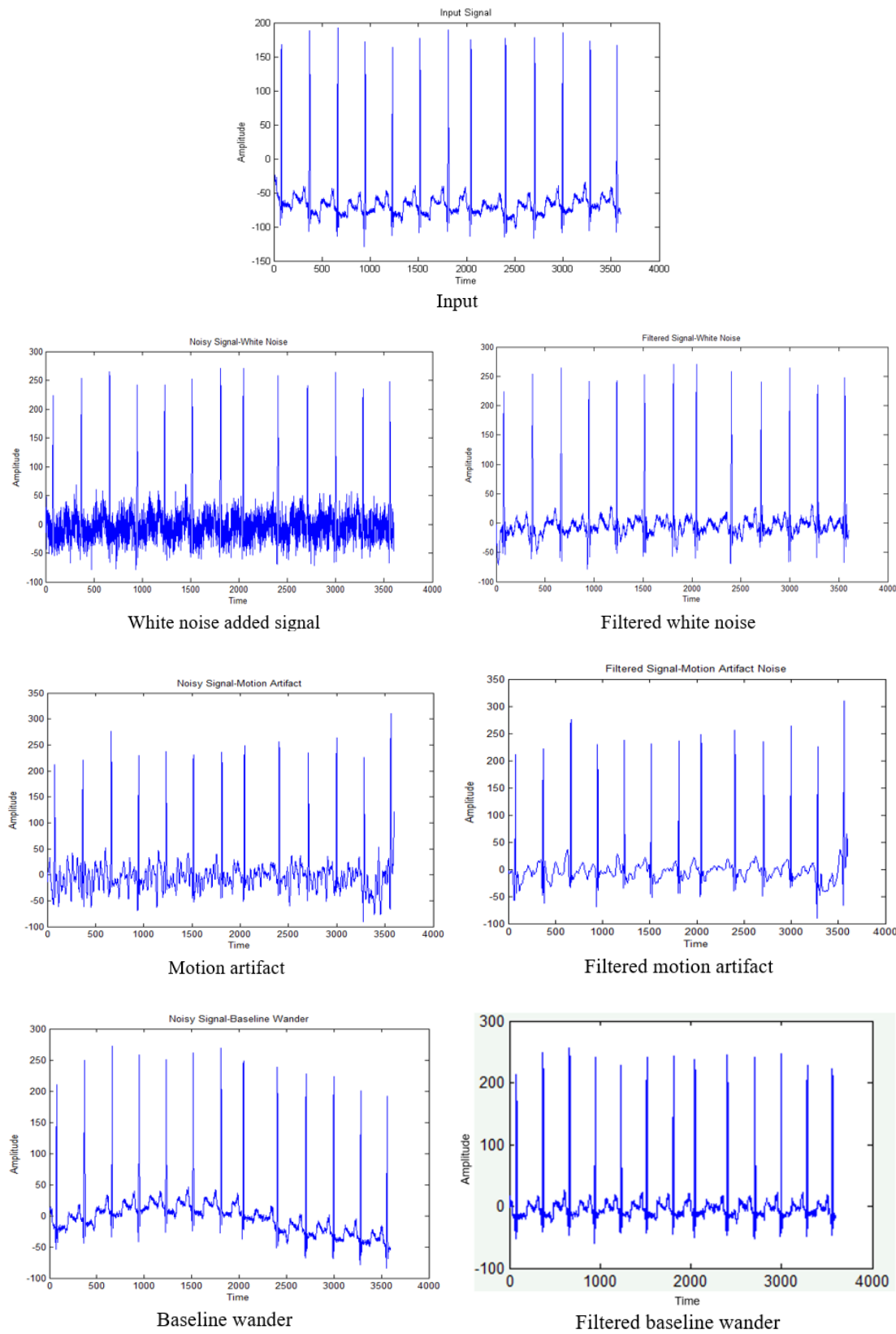


Figure 6. Comparison of original and filtered signals under different noise conditions

Table 5. Filtering performance for varied noise types

Noise dB	Performance	White noise	Color noise	Motion artifact	Electrode artifact	Baseline Wander
5 dB	PSNR	36.23	33.99	38.20	33.69	46.10
	MSE	118.21	357.91	187.53	374.82	393.82
	MAE	6.865	14.19	9.330	15.13	16.62
	RMSE	10.87	18.91	13.69	19.36	19.84
	PRD	4.42	5.98	6.22	1.22	1.1
10 dB	CC	0.89	0.91	0.950	0.92	0.90
	PSNR	44.82	40.95	43.53	39.95	40.80
	MSE	67.89	123.03	123.03	143.38	125.92
	RMSE	82.28	11.09	91.1	11.97	11.22
	PRD	4.58	2.86	7.28	6.08	1.6
	CC	0.92	0.93	0.96	0.95	0.94

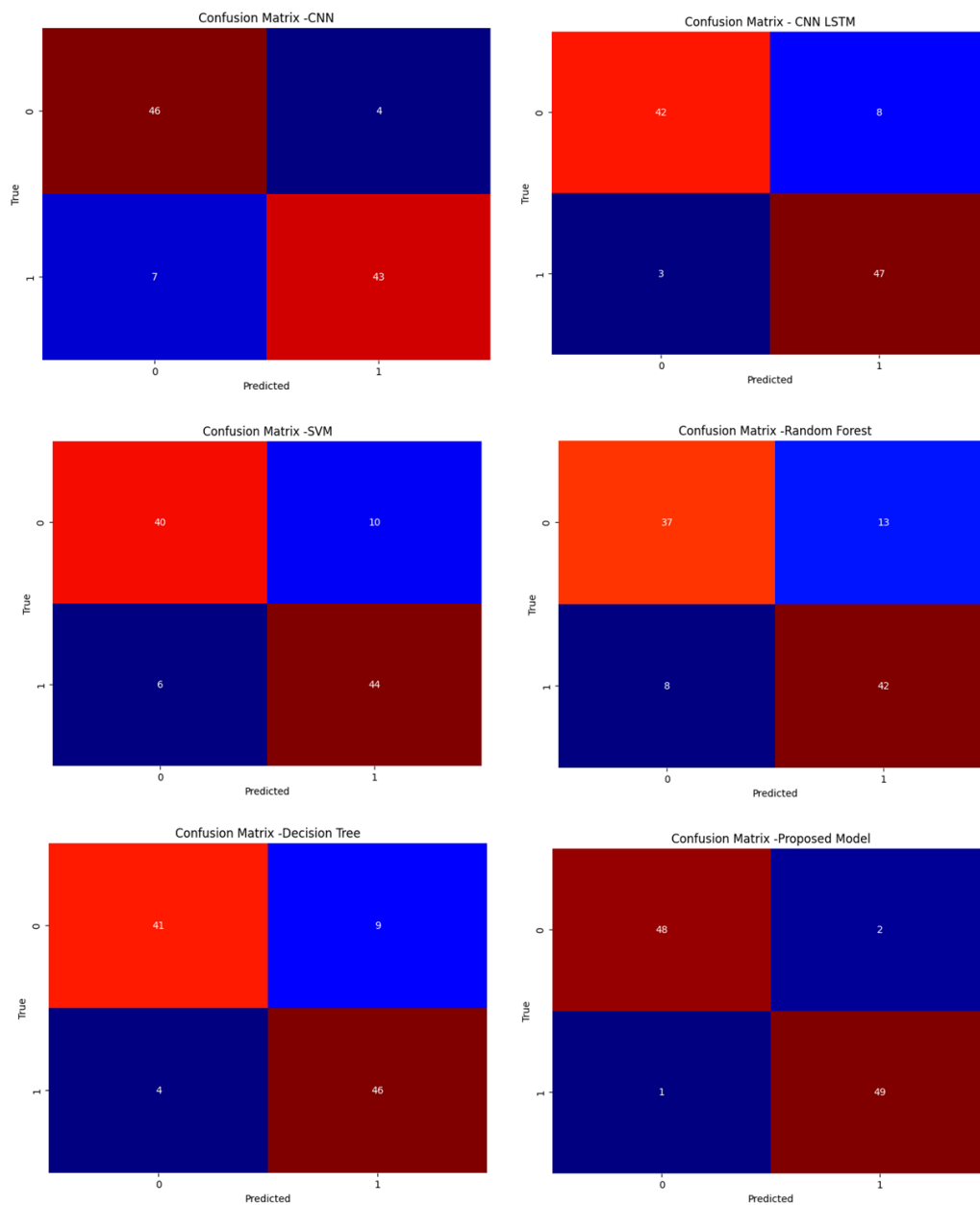


Figure 7. Confusion matrix of different classifiers

Table 6. Overall classification performance

Measure	CNN-LSTM	SVM	RF	CNN	DT	MBANet
Sensitivity	0.9333	0.8696	0.8222	0.9149	0.9111	0.9796
Specificity	0.8545	0.8148	0.7636	0.9200	0.8364	0.9608
Precision	0.8400	0.8000	0.7400	0.8600	0.8200	0.9600
-Ve predictive value	0.9400	0.8800	0.8400	0.0851	0.9200	0.9800
False positive rate	0.1455	0.1852	0.2364	0.0800	0.1636	0.0392
False discovery rate	0.1600	0.2000	0.2600	0.1321	0.1800	0.0400
False negative rate	0.0667	0.1304	0.1778	0.8900	0.0889	0.0204
Accuracy	0.8900	0.8400	0.7900	0.8932	0.8700	0.9700
F1 Score	0.8842	0.8333	0.7789	0.9200	0.8632	0.9697

Table 7. Computation time (Seconds)

Model	20 epochs	40 epochs	60 epochs	80 epochs	100 epochs
CNN-LSTM	420	830	1250	1680	2030
SVM	310	620	930	1240	1550
Random Forest	210	420	630	840	1050
CNN	370	710	1060	1420	1770
Decision Tree	260	510	760	1010	1260
MBANet Method	160	300	460	610	750

In this work, we have used CNN-LSTM, SVM, RF, CNN, and DT for multimodal authentication. The CNN-LSTM combines CNNs, effective for spatial feature extraction in images like irises, with LSTMs, which are proficient in handling sequential data like ECG signals. This fusion is beneficial for extracting both spatial and temporal features, making it suitable for multimodal authentication. However, it requires careful regularization and tuning to prevent overfitting, especially when dealing with smaller datasets. Similarly, SVMs are effective in finding the hyperplane that best separates classes in high-dimensional spaces, making them suitable for both iris and ECG feature classification. However, SVMs can lack interpretability when used with complex kernels or in high-dimensional spaces, making it challenging to understand the learned features. Random forests are ensemble learning methods that build multiple decision trees and merge their predictions, offering robust performance across various feature types, including ECG and iris features. However, it requires tuning of parameters such as the number of trees and depth of trees to prevent overfitting and achieve optimal performance.

3.4. Discussion

The proposed MBANet model has demonstrated remarkable performance in the multimodal authentication system using ECG and iris modalities. This approach addresses key challenges such as noise in ECG signals and iris images, as well as the limitations inherent in unimodal biometric systems. The results obtained indicate significant improvements in both signal quality and classification accuracy.

Comparison with other studies: When comparing the results of the MBANet model with other studies, we observe that traditional methods relying on single-modality approaches (*e.g.*, using only ECG or only iris data) tend to suffer from lower accuracy and higher susceptibility to noise. Our results demonstrate that MBANet significantly improves accuracy, achieving an impressive 97% classification accuracy, compared to 89% accuracy in traditional CNN-LSTM and SVM approaches. In particular, the noise filtering performance achieved by MBANet outperforms standard models by up to 20%, with a significant reduction in MSE, RMSE, and MAE values. This suggests that MBANet is more capable of maintaining high-quality signal integrity in the presence of noise, which is crucial for real-time applications. These improvements are consistent with recent multimodal authentication studies, but our study is unique in incorporating both a robust noise filtering mechanism and the application of transfer learning to enhance classification accuracy.

The training speed performance of MBANet was also superior to that of the other models tested, such as SVM, RF, and CNN-LSTM. This can be attributed to the use of pre-trained deep learning models, which allow MBANet to significantly reduce training time without compromising on accuracy. MBANet demonstrated a consistent advantage in terms of computation time, particularly as the number of epochs increased, making it a more scalable solution compared to traditional classifiers.

3.5. Limitations and threats to validity

While the MBANet model demonstrates impressive results, there are several limitations and threats to validity that need to be considered:

- Dataset size and diversity: Although the real-time ECG and iris samples used in this study are promising, the dataset is relatively small (100 user cases). A larger, more diverse dataset is required to validate the model's generalizability across different populations and conditions. Moreover, the current dataset may not fully represent the variety of potential real-world scenarios, such as varying user age groups or health conditions.
- Noise types and levels: The performance of MBANet was evaluated under multiple noise types (*e.g.*, white noise, color noise, motion artifacts) at 5 and 10 dB levels. However, the model's robustness under extreme or real-world noise conditions, such as low-frequency noise or high-interference environments, is yet to be assessed. More studies with varied and more realistic noise conditions are needed to confirm the robustness of MBANet.
- Model complexity and overfitting: The MBANet model, being a deep learning-based approach, may be prone to overfitting when working with smaller datasets. The use of pre-trained models helps mitigate this risk, but regularization techniques and careful tuning are crucial to avoid overfitting, especially as the dataset grows in size.
- Interpretability: Despite the impressive performance of the model, the black-box nature of deep learning models poses challenges for interpretability. Understanding how the model extracts features from ECG and iris signals and makes classification decisions is important, particularly in applications requiring high security or explainability. More research into model interpretability, such as using attention mechanisms or explainable AI (XAI) techniques, would help address this concern.
- Real-time implementation: The model was validated in a controlled environment with limited user cases. Extending the model to real-time applications with dynamic environmental factors (*e.g.*, varying noise conditions, user movement) might introduce additional challenges. The performance of the system in such real-world conditions requires further exploration.

4. CONCLUSION

Despite significant efforts in the development of ECG-based biometric modalities, several important issues remain inadequately addressed in the pursuit of new algorithms. Firstly, the available database for ECG data is limited and often comprises data from individuals with health conditions, including noise artifacts. Secondly, previous algorithms developed for ECG-based biometrics have not undergone thorough investigation regarding each of the prominent techniques, such as filtering methods, segmentation approaches, feature extraction techniques, and the overall quality assessment of ECG data. Moreover, the accuracy of unimodal authentication system is affected due to increased sample size. In order to overcome these issue, multimodal authentication systems are widely adopted in various applications. In this work, we present a novel approach of multimodal authentication by using ECG and iris modalities. The first phase performs data pre-processing where data filtering and image denoising tasks are performed. In next phase, the feature extraction task considers HRV features in time and frequency domain. Finally, the CNN and DeepResNet based transfer learning models are used to perform the classification. The outcome of the proposed approach is validated in real-time ECG and iris samples where the MBANet approach has reported accuracy 97%. The primary contributions of this work are:

In multimodal fusion approach, we propose a unique combination of ECG and iris biometrics, addressing the challenges associated with the limitations and noise in single-modality ECG data. By integrating these modalities, the proposed system achieves enhanced authentication accuracy compared to traditional unimodal systems. In Data preprocessing and denoising our proposed system incorporates effective pre-processing techniques, including ECG data filtering and iris image denoising, to mitigate the effects of noise and artifacts. These techniques are crucial for improving the quality of the raw data and ensuring more reliable feature extraction. The feature extraction from HRV explores the extraction of HRV features from ECG in both time and frequency domains. These features are key to capturing the variability and underlying patterns in ECG signals, which play a critical role in biometric authentication. The use of advanced CNNs and DeepResNet-based transfer learning models for classification significantly improves the performance. The model's ability to learn from pre-trained networks allows for better generalization to unseen data, contributing to the system's

overall robustness. Our proposed multimodal authentication system is validated using real-time ECG and iris samples, demonstrating an impressive accuracy 97% using the MBANet approach. This result showcases the effectiveness of the system in real-world practical scenarios.

FUNDING INFORMATION

The authors have received no financial support for the research, authorship, and publication of this article.

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
Ashwini Kailas	✓	✓	✓			✓		✓		✓		✓	✓	
Madhusudan Girimallaih	✓				✓		✓	✓		✓			✓	
Mallegowda Madigahalli		✓		✓		✓			✓		✓			✓
Vasantha Kumara Mahadevachar				✓			✓		✓			✓		✓
Pranothi Kadirehally Somashekarappa			✓		✓	✓					✓			✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal Analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project Administration

Fu : Funding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest to disclose.

DATA AVAILABILITY

The real-time data for this study were acquired from ECG and Iris machine. Due to the continuously evolving nature of the data and the specific access method employed, the raw data analyzed in this study are not publicly archived. Researchers interested in understanding the data access methodology are encouraged to contact the corresponding author for further details.




REFERENCES

- [1] P. Datta, S. Bhardwaj, S. N. Panda, S. Tanwar, and S. Badotra, "Survey of security and privacy issues on biometric system," *Handbook of Computer Networks and Cyber Security: Principles and Paradigms*, pp. 763-776, 2020, doi: 10.1007/978-3-030-22277-2-30.
- [2] S. Khade, S. Ahirrao, S. Phansalkar, K. Kotecha, S. Gite, and S. D. Thepade, "Iris liveness detection for biometric authentication: A systematic literature review and future directions," *Inventions*, vol. 6, no. 4, p. 65, 2021, doi: 10.3390/inventions6040065.
- [3] Y. Moolla, A. De Kock, G. Mabuza-Hocquet, C. S. Ntshangase, N. Nelufule, and P. Khanyile, "Biometric recognition of infants using fingerprint, iris, and ear biometrics," *IEEE Access*, vol. 9, pp. 38269-38286, 2021, doi: 10.1109/ACCESS.2021.3062282.
- [4] C. Kant and S. Chaudhary, "A multimodal biometric system based on finger knuckle print, fingerprint, and palmprint traits," in *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020*, pp. 182-192, Springer Singapore, 2021, doi: 10.1007/978-981-15-6067-5-21.
- [5] S. Dargan and M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities," *Expert Systems with Applications*, vol. 143, p. 113114, 2020, doi: 10.1016/j.eswa.2019.113114.
- [6] A. B. Tatar, "Biometric identification system using EEG signals," *Neural Computing and Applications*, vol. 35, no. 1, pp. 1009-1023, 2023, doi: 10.1007/s00521-022-07795-0.
- [7] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, pp. 441-467, 2016, doi: 10.1109/TIFS.2015.2480381.
- [8] S. S. Sengar, U. Hariharan, and K. Rajkumar, "Multimodal biometric authentication system using deep learning method," in *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 309-312, IEEE, Mar. 2020, doi: 10.1109/ESCI48226.2020.9167512.




- [9] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks," *IEEE transactions on information forensics and security*, vol. 9, pp. 1084–1097, Jul. 2014, doi: 10.1109/TIFS.2014.2322255.
- [10] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20–30, Sep. 2015, doi: 10.1109/MSP.2015.2437652.
- [11] N. Kose and J.-L. Dugelay, "On the vulnerability of face recognition systems to spoofing mask attacks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2357–2361, May 2013, doi: 10.1109/ICASSP.2013.6638076.
- [12] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: challenges and research opportunities," *IEEE Security Privacy*, vol. 13, no. 5, pp. 63–72, Sep. 2015, doi: 10.1109/MSP.2015.116.
- [13] R. B. R. and A. S. S., "Deep insights on processing strata, features and detectors for fingerprint and iris liveness detection techniques," *Multimedia Tools and Applications*, vol. 83, no. 23, pp. 63795–63846, 2024, doi: 10.1007/s11042-024-18690-2.
- [14] P. P. K. Chan, *et al.*, "Face liveness detection using a flash against 2D spoofing attack," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 521–534, Feb. 2018, doi: 10.1109/TIFS.2017.2758748.
- [15] P. Burgess, "ECG: recording the electrical activity of the heart," *British Journal of Healthcare Assistants*, vol. 16, no. 12, pp. 548–554, 2022, doi: 10.12968/bjha.2022.16.12.548.
- [16] N. Ammour, Y. Bazi, and N. Alajlan, "Multimodal approach for enhancing biometric authentication," *Journal of Imaging*, vol. 9, no. 9, p. 168, 2023, doi: 10.3390/jimaging9090168.
- [17] 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, pp. 22509–22535, 2019, doi: 10.1007/s11042-019-7467-x.
- [18] S. A. El Rahman, "Multimodal biometric systems based on different fusion levels of ECG and fingerprint using different classifiers," *Soft Computing*, vol. 24, pp. 12599–12632, 2020, doi: 10.1007/s00500-020-04700-6.
- [19] 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, 2019, doi: 10.1016/j.cose.2018.11.003.
- [20] C. Yuan, Q. Zhang, S. Wu, and Q. J. Wu, "A realtime fingerprint liveness detection method for fingerprint authentication systems," *Advances in Computers*, vol. 131, pp. 149–180, 2023, doi: 10.1016/bs.adcom.2023.04.004.
- [21] N. Bala, R. Gupta, and A. Kumar, "Multimodal biometric system based on fusion techniques: A review," *Information Security Journal: A Global Perspective*, vol. 31, no. 3, pp. 289–337, 2022, doi: 10.1080/19393555.2021.1974130.
- [22] M. Komeili, N. Armanfard, and D. Hatzinakos, "Liveness detection and automatic template updating using fusion of ECG and fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. 13, pp. 1810–1822, 2018, doi: 10.1109/TIFS.2018.2804890.
- [23] R. M. Jomaa, M. S. Islam, and H. Mathkour, "Improved sequential fusion of heart-signal and fingerprint for anti-spoofing," in *Proceedings of the 2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, Singapore, 11–12 January 2018, IEEE, pp. 1–7, doi: 10.1109/ISBA.2018.8311476.
- [24] R. M. Jomaa, M. S. Islam, H. Mathkour, and S. Al-Ahmadi, "A multilayer system to boost the robustness of fingerprint authentication against presentation attacks by fusion with heart-signal," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, pp. 5132–5143, 2022, doi: 10.1016/j.jksuci.2022.01.004.
- [25] G. Singh, R. K. Singh, R. Saha, and N. Agarwal, "IWT based iris recognition for image authentication," *Procedia Computer Science*, vol. 171, pp. 1868–1876, 2020, doi: 10.3390/electronics10080923.
- [26] J. Sun, S. Zhao, Y. Yu, X. Wang, and L. Zhou, "Iris recognition based on local circular Gabor filters and multi-scale convolution feature fusion network," *Multimedia Tools and Applications*, vol. 81, no. 23, pp. 33051–33065, 2022, doi: 10.1007/s11227-023-05421-x.
- [27] M. Omran and E. N. AlShemmary, "An iris recognition system using deep convolutional neural network," *Journal of Physics: Conference Series*, vol. 1530, no. 1, p. 012159, May 2020, doi: 10.1088/1742-6596/1530/1/012159.

BIOGRAPHIES OF AUTHORS






Ashwini Kailas    holds Ph.D. from Visvesvraya Technological University from Belagavi, India. Her research interests are deep neural network, bio medical instrumentation, authentication mechanisms using ECG and Iris. She is working at Sri Siddartha Institute of Technology, Tumakuru. She can be contacted at ashwinik@ssit.edu.in.






Madhusudan Girimallaih    holds Ph.D. from Visvesvaraya Technological University from Belagavi, India. His research interests are wireless sensor networks, authentication mechanisms using biometrics, deep learning and IoT. He is working at J S S Science and Technological University, Mysore. He can be contacted at madhusudan@sjce.ac.in.



Mallegowda Madigahalli    holds Ph.D. from Visvesvaraya Technological University from Belagavi, India. His research interests are industrial IoT, authentication mechanisms using biometrics, machine learning and IoT. He is working at M S Ramaiah Institute of Technology, Bangalore. He can be contacted at mallegowdam@msrit.edu.



Vasantha Kumara Mahadevachar    holds Ph.D. from Visvesvaraya Technological University from Belagavi, India. His research interests are wireless sensor networks, authentication mechanism using biometrics, deep learning and IoT. He is working at Government Engineering College, Hassan. He can be contacted at cmn.vasanth@gmail.com.



Pranothi Kadirehally Somashekarappa    is currently working as assistant professor in bio medical engineering at Sri Siddhartha Institute of Technology, Tumkur. She holds B E in electronics and communication from Kalpataru Institute of Technology, Tiptur and M.Tech from VTU Belagavi. She can be contacted at pranotisomashekar9@gmail.com.