

Deep transfer learning for classification of ECG signals and lip images in multimodal biometric authentication systems

Latha Krishnamoorthy, Ammasandra Sadashivaiah Raju

Department of Bio-Medical Engineering, Sri Siddhartha Institute of Technology, Sri Siddhartha Academy of Higher Education University, Tumakuru, India

Article Info

Article history:

Received Mar 26, 2024

Revised Mar 28, 2025

Accepted Jun 8, 2025

Keywords:

Biometric authentication

Classification

Deep learning

Electrocardiogram

Multimodal

ABSTRACT

Authentication plays an essential role in diverse kinds of application that requires security. Several authentication methods have been developed, but biometric authentication has gained huge attention from the research community and industries due to its reliability and robustness. This study investigates multimodal authentication techniques utilizing electrocardiogram (ECG) signals and face lip images. Leveraging transfer learning from pre-trained ResNet and VGG16 models, ECG signals and photos of the lip area of the face are used to extract characteristics. Subsequently, a convolutional neural network (CNN) classifier is employed for classification based on the extracted features. The dataset used in this study comprises ECG signals and face lip images, representing distinct biometric modalities. Through the integration of transfer learning and CNN classification, improving the reliability and precision of multimodal authentication systems is the primary objective of the study. Verification results show that the suggested method is successful in producing trustworthy authentication using multimodal biometric traits. The experimental analysis shows that the proposed deep transfer learning-based model has reported the average accuracy, F1-score, precision, and recall as 0.962, 0.970, 0.965, and 0.966, respectively.

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



Corresponding Author:

Latha Krishnamoorthy

Department of Bio-Medical Engineering, Sri Siddhartha Institute of Technology

Sri Siddhartha Academy of Higher Education University

Tumakuru, India

Email: lathak@ssit.edu.in

1. INTRODUCTION

In recent times, as internet of things (IoT) technology continues to advance, the utilization of cloud services has become prevalent. Various devices are now equipped with networking capabilities to facilitate communication between machines and humans. Consequently, ensuring information security has become paramount, particularly as user data is employed to govern a multitude of devices [1]. Confidentiality and integrity are the essential components of information security.

User authentication plays important role in this context to maintain the confidentiality and integrity of the data. Several methods have been introduced to improve the reliability of authentication such as password-based authentication, multi-factor authentication (MFA), one-time password, smart cards and tokens. These methods have been adopted widely in various applications but user's liveliness is not considered in these works. Therefore, researchers have developed biometric authentication system which uses special biometric data to identify a user, like fingerprints, face scans, iris scans, or voice recognition. The main advantages of this system are that it is difficult to spoof or replicate, offers a high level of security, and eliminates the need to remember passwords. Moreover, the biometric authentication systems utilize human characteristics for recognition

including behavioral and physical characteristics. The behavioral characteristic analysis includes voice, gait, and electrocardiogram (ECG), while physical features include the face, fingerprints, and iris. Several works have been developed based on these biometric modalities such as fingerprint [2], finger vein-based authentication system [1], face [3], and voice [4]. However, traditional biometrics faces numerous challenges, which have susceptibility to spoofing or forgery [5].

Moreover, the biometric systems rely on the unique physiological and behavioral traits of the users therefore it plays crucial role in enhancing the security and mitigating the vulnerabilities. Among the numerous biometric modalities, cardiac based biometric systems have gained huge attention in human identification and security enhancement. The intrinsic electrical activity of the human heart, as captured by ECG, photoplethysmogram (PPG), and “phonocardiogram (PCG)” signals [6], [7], offers a valuable source of information that can be utilized for secure authentication. Moreover, adopting ECG signal over other biometric modalities has several advantages over PPG and PCG signals. Therefore, ECG is considered as unique and highly individualistic biomarker in medical domain [4] because it provides intricate electrical activity pattern of heart which is beneficial in security and authentication systems.

In order to address the issues of traditional authentication systems several authors have reported the advantages of combining multimodal authentication systems. Several models have been introduced based on multimodal authentication system such as Zhang *et al.* [8] used face and voice models to develop android based authentication system. El-Rahiem *et al.* [1] used ECG and finger vein modalities. Chanukya and Thivakaran [9] used combination of fingerprint and ear modalities. However, achieving the accuracy remains challenging task due to huge variations in the multimodal. In order to solve this problem, we introduce a novel method for user authentication that takes into account ECG and lip extraction from facial pictures. Section 2 provides a brief overview of the relevant literature, section 3 describes the deep transfer learning-based model that will be used, section 4 compares and contrasts the solutions that have been explored, and section 5 concludes with suggestions.

2. LITERATURE SURVEY

A brief literature overview of current approaches for ECG, facial, and lip authentication and categorization are given in this section. To integrate many modalities in a biometric authentication system, Hammad *et al.* [10] attempted to integrate “convolutional neural networks (CNNs)” with “Q-Gaussian multi-support vector machines (QG-MSVMs)”. Several fusion levels are used by this model. To extract features for certain modalities, CNNs are employed. In this stage, we chose two CNN layers that gave us the best accuracy. Each feature description is treated as an independent layer. We then merge the feature descriptors with the proposed internal fusion approach. In addition, one of the cancellable biometric approaches is then used to further strengthen the security of the proposed system and these templates. During the authentication step, the performance is improved by using QG-MSVM as an authentication classifier.

Ahamed *et al.* [11] combined ECG and PPG signals to build a biometric system to support individualized healthcare systems. This strategy comprised time-domain and combined time-frequency domain feature extraction methods that are based on autoregressive coefficients, the Shannon entropy, and the wavelet packet transform. Using the retrieved information, a CNN-long short-term memory (LSTM) classifier is trained subsequently. Itani *et al.* [12] pointed out that biometric systems relying on facial features have problems when people are wearing masks, and authentication methods based on fingerprints have problems when users' hands get damp. In response to these concerns, the writers proposed an ear authentication method.

Similarly, Purohit and Ajmera [13] proposed a multimodal authentication system established on palm, fingerprint, and ear biometrics. Gabor features are used for hand images, the human microstructure based (HMSB) administrator for fingerprints, and HMSB and multiple regular gradient (MRG) for ear biometrics in this methodology, which employed texture and form feature extraction approaches. In addition, to ensure efficient feature selection, an adversarial gray wolf optimization approach is utilized, followed by the utilization of a multi-kernel support vector machine (SVM) classifier for recognition.

The authors in [14] introduced a new method of authentication using data from facial images. The dimensions of feature vectors extracted from face images are usually high, so they have to be reduced. The biometric verification system they unveiled used digital signatures and facial recognition software. They achieve this by employing a fusion feature vector, which incorporates features retrieved from both modalities. They proposed to use a “modified context-aware (MCA)” approach to generate a feature vector and employ a “tangential discrimination analysis (TDA)” algorithm to reduce the dimensionality of the features within facial photographs. Next, they train a modified mixed sequence deep neural network (MMS-DNN) using the fusion feature vector.

Singh and Tiwari [15] developed a multimodal authentication system. The work proposed involves integration of three unimodal biometric systems to form two multimodal biometric systems. For the purpose of this study, ECG, sclera, and fingerprint are chosen as unimodal system. In the first multimodal biometric

system we adopt a sequential model approach with the “whale optimization algorithm-artificial neural network (WOA-ANN)” decision level fusion. Meanwhile, the second multimodal biometric system employs a parallel model approach, employing score-level fusion based on “salp swarm algorithm-deep belief network (SSA-DBN)”. The biometric authentication process encompasses preprocessing, feature extraction, matching, and scoring for each individual unimodal system. Matching scores and individual accuracy for each biometric attribute are encrypted independently. A fusion procedure based on matcher performance is employed for the three biometric traits, as the matchers produce varied values across these attributes.

Cherifi *et al.* [16] introduced multimodal authentication system by using arm gesture and ear shape. The ear feature extraction considers local phase quantization mechanism which is used to handle the pose and illumination variations. Similarly, for arm gesture also statistical features are extracted. Finally, the obtained features are combined on score level by considering weighted sum. Kaul *et al.* [17] presented ECG based biometric authentication system where non-fiducial feature extraction approach is introduced. This approach is constructed by the combination of discrete cosine transform and autocorrelation. Further, the obtained features are then fed into the neural network model where multilayer perceptron and radial basis functions modules are used to train the model.

Kim *et al.* [18] used electromyogram signal because these signals cannot be foraged and therefore suggested an authentication approach. According to this approach, time domain attributes are extracted from the pre-processed signal later LSTM is used to match the gesture. Finally, CNN-LSTM is used to obtain the final classification. Grace *et al.* [19] built an EEG authentication system that accomplished signal feature extraction through DWT before using feed forward neural network for training the extracted features. The complete system accuracy of this method reached 87.7%.

The multimodal biometric authentication system presented by Younis and Abuhammad [20] employs Resnet101, Resnet-Inceptionv2, Densenet201, AlexNet, and Inceptionv2 deep transfer learning model to combine oversized handcrafted features derived from Hog feature descriptor. The fusion task applies discriminant correlation analysis (DCA) and canonical correlation analysis (CCA) to complete it. This technique achieved a recognition rate of 96.6% in its results. Siam *et al.* [21] demonstrated a framework for authentication with ECG and PPG signals that provides cancelable biometrics. This work presented a template generation solution for individual users through unification techniques. The extraction of features uses Mel-frequency cepstral coefficients (MFCCs) as its framework. The performance outcomes of classification depend on the utilization of multi-layer perceptron (MLP) and logistic regression classifier. Jeong *et al.* [22] developed DemoID as a new authentication approach that uses face together with voice biometrics for authentication purposes. Aleidan *et al.* [23] developed authentication through monitoring simultaneous ECG, PPG, and PCG signals using deep learning techniques. The proposed deep learning system implements transfer learning together with LSTM architecture for feature optimization. The attributes undergo classification through an implementation of boosting mechanism. Merging face and palm print and iris biometrics constitutes the hybrid biometric authentication system described in [24]. Variations of group search optimization (MGSO) approach optimize the process of extracting features. A teacher learning based deep learning model serves to perform the last stage classification of features. Modak and Jha [25] built a multimodal authentication system by combining face and eye alongside fingerprint modals. This approach Viola-Jones approach for face segmentation, later feature extraction is performed and obtained features are optimized by using chaos-based salp swarm algorithm (CSSA). Finally, rule-based adaptive neuro-fuzzy inference system (R-ANFIS) algorithm is introduced for classification. Also, deep learning models integrated with transfer learning based recent studies [26]–[28] have shown great improvement in enhancing classification accuracy while applying on image processing tasks.

3. PROPOSED MODEL

Many applications utilize deep learning techniques because these methods offer exceptional abilities to discover intricate patterns in processing such as images and videos together with biomedical images. The transfer learning technique finds extensive use across multiple applications since it leverages deep learning models trained for initial purposes to execute or retrain them for similar-related tasks. Our research uses this model structure for conducting ECG analysis and lip image evaluation.

The classification of ECG signals becomes possible through usage of pre-trained signal processing models from distinct different processing tasks. The deep learning models trained with general time-series information can successfully perform classification duties in ECG applications. The pre-trained model enables learning of vital features containing temporal patterns and frequency characteristics. The pre-existing features undergo customization before executing specific classification procedures. The pre-trained models demonstrate functionality in the processing of face and lip images too. The pre-trained models can extract visual features related to lip movement or expression by utilizing models from ImageNet or ResNet.

The research adapts these models to extract characteristics from both data types before implementing a deep learning classification method to identify users. We have included the authenticity measure through the decision-making process.

3.1. Data augmentation

The data augmentation plays important role in deep learning and machine learning based tasks. It is widely adopted mechanism in order to widen training data and expand dataset size by applying different mechanisms. In the context of ECG, data augmentation can be particularly useful for improving model generalization and robustness, especially when dealing with limited data. The data augmentation schemes are as follows:

- Adding the noise: adding random noise to ECG signals can simulate noise present in real-world recordings, making the model more robust to noise. Various kinds of noise, such as Gaussian noise, white noise, or even specific types of interference noise, can be added to the ECG signals.
- Temporal variations: the ECG signals have huge impact of temporal variations therefore we incorporate. This helps the model learn to be invariant to slight temporal shifts in the signal.
- Amplitude scaling: it can simulate variations in signal strength or electrode placement thus helps to generalize better to variations in signal intensity
- Baseline wander: introducing baseline wander by adding a low-frequency sinusoidal component to the ECG signal can simulate variations in baseline drift. This helps the model learn to detect and classify ECG features accurately in the presence of baseline drift.

Similarly, we apply data augmentation on lip image dataset. The augmentation on image data includes several stages such as: rotation, flip, brightness, contrast adjustment, noise addition, and cropping. A brief discussion about these augmentations is given as follows:

- Rotation: it performs rotation on the original image to introduce the variability which helps to improve the robustness to different variations.
- Image flipping: flip the face images horizontally to generate left-right mirror images, which can help the model generalize better to faces with different orientations.
- Brightness and contrast adjustment: this helps to consider the varying lightning and illumination conditions.
- Noise addition: introduce random noise to the face/lip images to simulate noise in image acquisition devices or environmental conditions.
- Crop and zoom: crop and zoom the face images to focus on different regions of interest, such as the lips, to help the model learn invariant representations.

3.2. Feature extraction

We use the augmented ECG and lip image data to extract the features. The first subsection presents the process of feature extraction from ECG data and second stage presents the outcome of lip image feature extraction. Here, ResNet50 pre-trained model is used to extract the deep features from scalogram images.

3.2.1. ECG feature extraction

This section describes the suggested model for ECG feature extraction. In this work, we have considered wavelet transform base method to extract the features from input ECG signal. Wavelet transforms are widely adopted as the extension of conventional Fourier transform model. However, the Fourier transform operates on a singular frequency or scale, whereas wavelets operate across multiple scales of frequencies. Wavelet analysis involves decomposing any signal into various versions with different shifts and scales from the original wavelets. In our proposed methodology, we primarily use the “continuous wavelet transform (CWT)”. We map the signal onto a time scale domain, where each time scale indexes a specific subset of the frequency domain. The CWT of any given signal $E(t)$ is obtained based on an integral of $E(t)$ as follows:

$$CWT(x, y) = \frac{1}{\sqrt{x}} \int_{-\infty}^{\infty} E(t) * \varphi_{x,y} \left(\frac{t-x}{y} \right) dt \quad (1)$$

Where φ characterises the mother wavelet. The shifting and scaling operation on mother wavelet produce daughter wavelet which is represented as $\varphi_{x,y}$ where x and y represents the scaling and shifting factors, respectively. The CWT generates several wavelet coefficients C . The CWT module is the pre-processed filtered signal. Specific CWT wavelet coefficients are derived using the CWT. Then, the series of continuous wavelet filter banks are applied on these coefficients. A two-dimensional image of the CWT coefficients for each ECG record is produced as the outcome. Figure 1 illustrates the scalogram of authentic user and imposter where Figure 1(a) illustrates the scalogram of authentic user and Figure 1(b) shows the scalogram of imposter.

Transfer learning-based models provide the extraction of robust features by processing the data through deep learning models. The model obtains deep features from scalogram images by employing

ResNet50 as its pre-trained architecture. The network structure named ResNet-50 functions as a deep convolutional neural network (DCNN) architecture. The ResNet architecture provides training capabilities for deep neural networks and this is one of its variant versions. ResNet-50 represents a 50-layer network structure that operates through blocks containing convolutional operations although ResNet-110 demonstrates the least variation between this model and the standard DenseNet implementation. The main breakthrough of ResNet serves as the foundation for introducing residual connections which extend deep network training capabilities by solving the vanishing gradient problem. The residual links connect straight to the convolutional layers for adding their extra information to the output features that exist in the convolutional layers and the original network input. The network training mechanism learns residual functions because residual connections enable this learning method instead of performing direct mapping. The technique enables the training of deep networks by helping optimization processes. The ResNet50 model houses 50 sequential layers for its structure. The complete model structure appears in Figure 2.

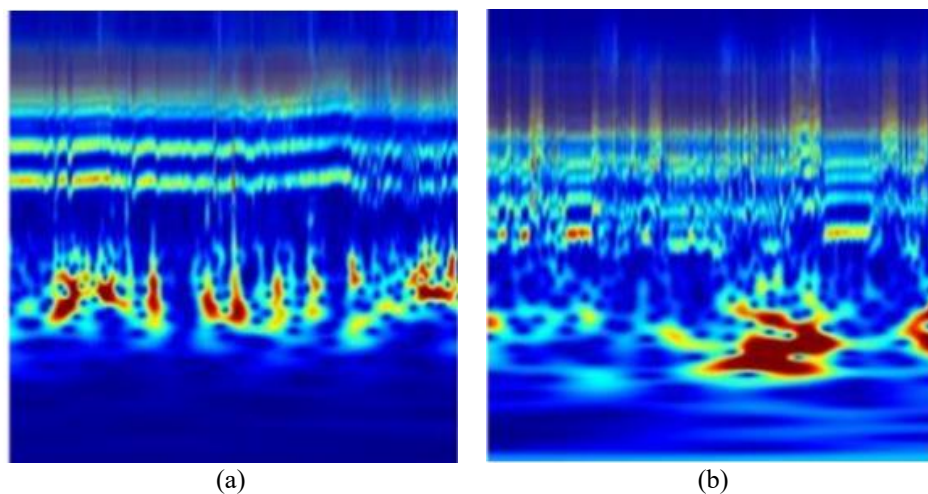


Figure 1. Scalogram representations of ECG signals: (a) authentic ECG and (b) imposter ECG

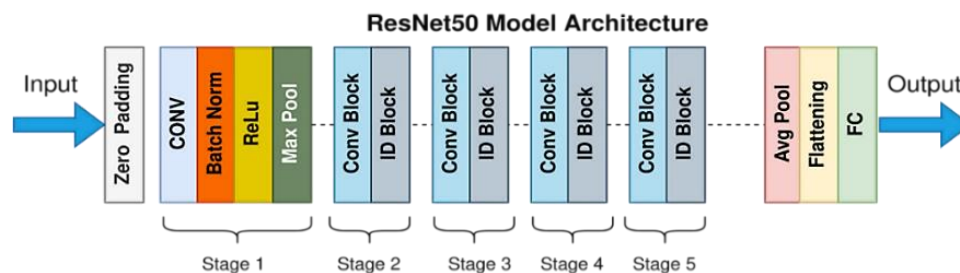


Figure 2. ResNet model

The detailed discussion is presented as follows:

- a) Input layer: the input layer accepts the input image data. In ResNet, input images are typically resized to a fixed size, often 224×224 pixels, and normalized.
- b) Convolutional layers (conv blocks): the first layer is a CNN layer with 64 filters of size 7×7 , followed by a max-pooling layer. This layer serves as the initial feature extractor.
- c) Residual blocks: ResNet-50 consists of 16 residual blocks, organized into different stages. Each residual block contains several convolutional layers and shortcut connections.
 - Stage 1 (Conv2_x): the first stage contains one residual block with two convolutional layers. The output of this stage is passed through activation and batch normalization layers.
 - Stage 2 (Conv3_x): the second stage contains three residual blocks, each with several convolutional layers. The output of each block is passed through activation and batch normalization layers.

- Stage 3 (Conv4_x): the third stage contains four residual blocks, each with several convolutional layers. Similar to the previous stages, the output of each block is passed through activation and batch normalization layers.
- Stage 4 (Conv5_x): the fourth stage contains six residual blocks, each with several convolutional layers. The output of each block is passed through activation and batch normalization layers.
- d) Global average pooling: the output feature map is passed through a global average pooling layer after the last residual block. It crowds the spatial dimension of the feature map down to 1×1 while keeping the same number of channels.
- e) Fully connected layer (output layer): then, the final output is produced by adding a fully connected layer with activation. When dealing with image classification, the output is the probabilities that the image is of a class.

3.2.2. Lip image extraction

This subsection describes the proposed transfer learning-based model for lip image feature extraction for authentication. For this task, we have used VGG16 feature extraction. Many computer vision tasks have made extensive use of VGG16, such as feature extraction, object detection, and image categorization. It has also served as a base model for transfer learning in many applications, where the pre-trained VGG16 weights are fine-tuned on specific datasets for certain tasks. With numerous convolutional layers followed by max pooling layers, VGG16's sixteen layers are grouped into five groups. Fully connected categorization layers make up the last set of layers. VGG16 is characterized by its repetitive convolutional blocks, each containing multiple 3×3 convolutional layers followed by a max pooling layer. Specifically, the architecture consists:

- Convolutional layers: short 3×3 filters with a 1 stride and padding are used by the convolutional layers to maintain the spatial dimensions of the input feature maps. These layers are responsible for learning spatial hierarchies of features in the input images.
- Rectified linear unit (ReLU) activation: to make the network non-linear, ReLU activation features are added after every convolutional layer.
- Max pooling layers: max pooling layers follow each group of convolutional layers to decrease computational complexity by downsampling feature maps' spatial dimensions and controlling overfitting.

After the convolutional blocks, VGG16 includes three fully connected layers followed by a SoftMax output layer. The architecture of this model is depicted in Figure 3. These layers serve as the classifier for the network, mapping the extracted features to class probabilities.

- Flatten layer: the feature maps are first flattened into a 1D vector before being fed into the dense layers, which are located preceding the fully connected layers.
- Dense layers: the fully connected layers contain a large number of neurons, enabling the network to acquire knowledge of general characteristics and generate forecasts using the retrieved representations.
- SoftMax activation: the SoftMax activation function in the output layer converts the raw output of the network into class probabilities, enabling multi-class classification.

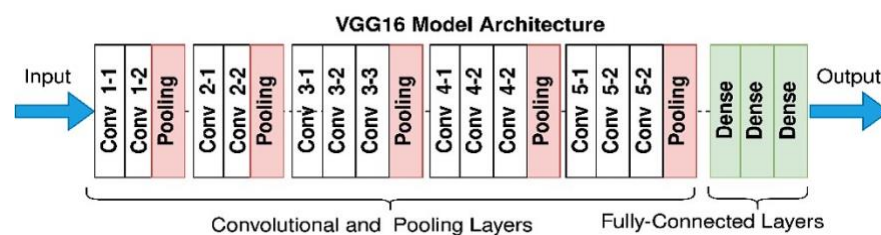


Figure 3. VGG16 architecture

3.2.3. Classification

This section shows the classification model used for both ECG and lip image data. This method uses CNN based classification model. This classification model input layer, convolution layer, max pooling layer, flatten layer, dense layer, drop out layer and output layer. Figure 4 shows the architecture of CNN classifier. Architecture: input (image)-convolutional layer (e.g., 3×3 filters, ReLU activation)-max pooling layer (e.g., 2×2 pool size)- convolutional layer-max pooling layer-flatten layer-dense layer (e.g., 128 neurons, ReLU activation), dropout layer (optional, for regularization), output layer (e.g., SoftMax activation for multi-class classification).

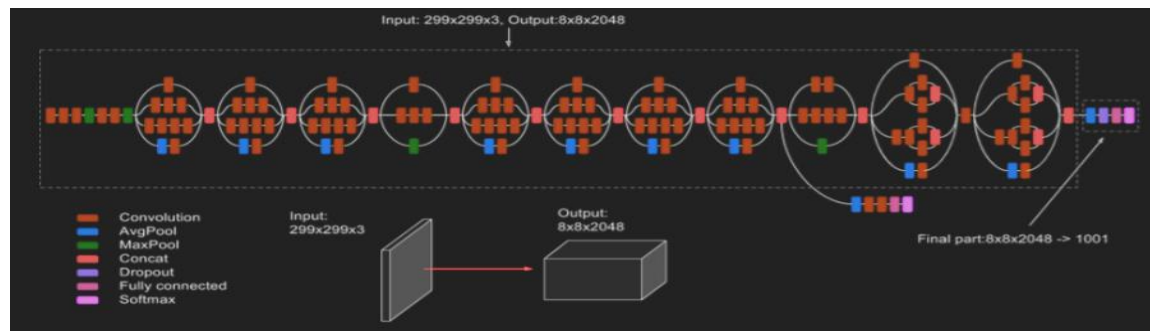


Figure 4. Architecture of CNN classifier

4. RESULTS AND DISCUSSION

This section consists of different subsection where subsection 4.1 presents the dataset details, section 4.2 presents the details of performance measurement parameters, section 4.3 presents the result of proposed model and compares it performance with current classification schemes.

4.1. Dataset details

The dataset development included collecting 20 ECG signals and 20 face images from each of the 10 participants in the study. The acquisition of ECG signals requires a filtering process because the signals tend to become contaminated. The research investigates multiple forms of ECG signal interference which consist of white noise and color noise along with motion artifacts electrode artifacts and baseline wander. To extract the face lip region from images the "dlib" Python library detects facial landmarks. This work excluded scenario analysis for obstructed regions because the main research objective was the extraction of lip areas. The team applied data augmentation methods to generate various datasets for each user because it improved the reliability of extracted features. The ECG signal sample in Figure 5 displays two dimensions where time flows across the horizontal axis and amplitude rises on the vertical axis.

The unfiltered signals and their filtered counterparts are shown in Figure 6. White noise is shown in Figure 6(a), color noise in Figure 6(b), motion artifact in Figure 6(c), electrode artifact in Figure 6(d), and baseline wander in Figure 6(e). The illustrations of each kind of noise and filtered signal shown in Figure 6. The input signal, white noise, color noise, artifacts caused by motion, artifacts caused by electrodes, and baseline wander, as well as the filtered signals used for biometric authentication were shown in Figure 6. We have also utilized facial landmark detection to extract the lip region from face photos in a similar vein. The face image, extracted lip image and enhanced image are given in Figure 7. Landmark identification procedure is used to get coordinates of the reduced and enhanced lip region and apply horizontal and vertical image flipping.

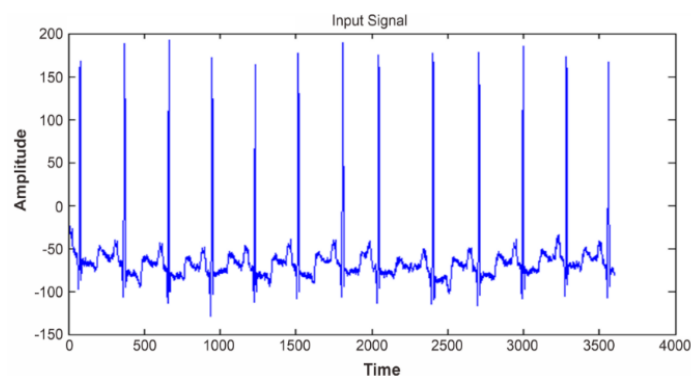


Figure 5. Sample ECG signal for input

4.2. Performance measurement parameters

Evaluation of the proposed method depends on confusion matrix calculations. The confusion matrix creation depends on true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Table 1 illustrates this classes.

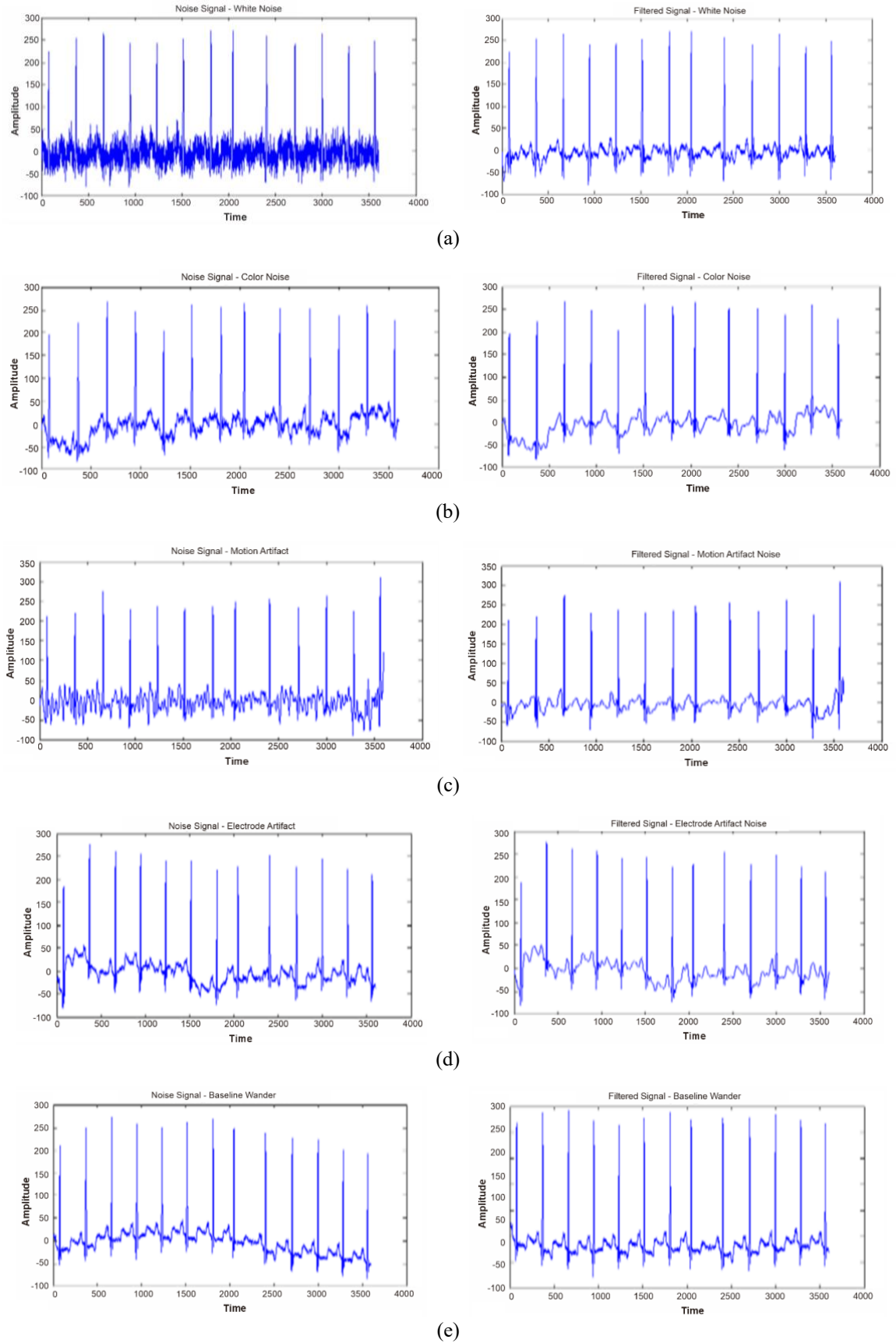


Figure 6. Noise added ECG signal and filtered signals for biometric authentication (a) white noise, (b) color noise, (c) motion artifacts, (d) electrode artifacts, and (e) baseline wander

Table 1. Confusion matrix classes

Actual class	Predicted class	
	Authentic	Imposter
Authentic	TP	FN
Imposter	FP	TN

We calculate accuracy and precision and F1-score by using our proposed approach based on the confusion matrix data. Out of all total number of instances accuracy represents the proportion of properly classified cases. The accuracy determination follows this method.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision evaluation takes place for the produced method. The precision evaluation includes a calculation using TP instances and their combination with true and false instances.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Lastly, the computation of F-measure depends on the sensitivity and precision values:

$$F = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (4)$$

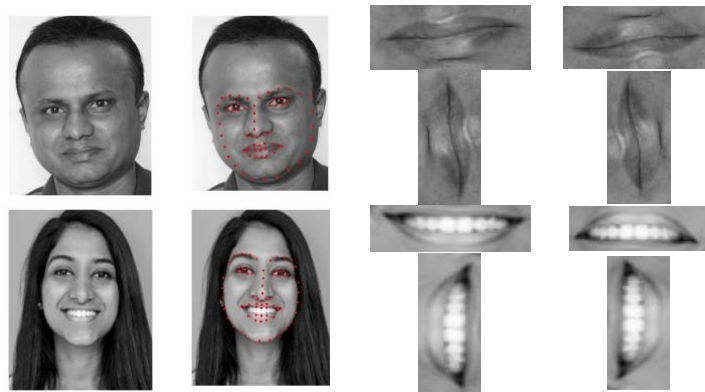


Figure 7. Face image, corresponding facial landmarks and augmentation

4.3. Comparative analysis

The proposed model measures its performance using accuracy, precision alongside F1-score metrics which are assessed against current benchmark classification models. The current output performance is assessed against all existing machine learning and deep learning classifiers. The machine learning classification model performance results appear in Table 2.

Table 2. Comparative analyses with machine learning techniques

Parameter	SVM	Neural network	RF	DT	Proposed
Accuracy	0.651	0.712	0.781	0.835	0.962
F1-score	0.655	0.698	0.772	0.820	0.970
Precision	0.648	0.711	0.795	0.815	0.965
Recall	0.658	0.728	0.735	0.805	0.966
Sensitivity	0.655	0.711	0.733	0.842	0.977
Specificity	0.661	0.715	0.785	0.835	0.985

According to this experiment, the proposed deep transfer learning model reported higher classification accuracy as 96.2% whereas the SVM classifier reported the lowest accuracy as 65.10%. The SVM model

considers simple texture, color and shape features whereas proposed model uses transfer learning model where ECG signal is converted in to image and then deep features are extracted. Figure 8 depicts the obtained performance analysis.

Further, we compare the obtained performance with different deep learning-based classification models. The obtained comparative analysis is depicted in below given Table 3. According to this experiment, the proposed model has reported highest accuracy whereas the other deep learning classifiers have reported 0.843, 0.855, 0.858, and 0.901 by using CNN, LSTM, gated recurrent unit (GRU), and CNN-LSTM classifiers, respectively. The comparative analysis depicted in Figure 9.

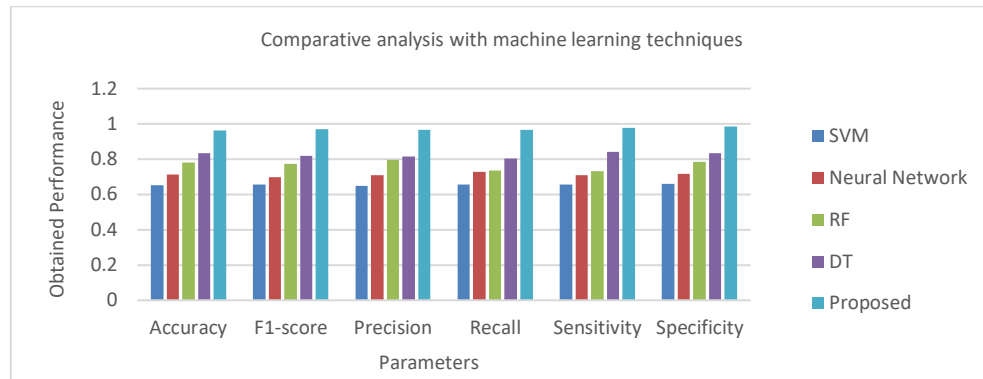


Figure 8. Comparative analysis with different machine learning techniques

Table 3. Comparative analyses with deep learning techniques

Parameter	CNN	LSTM	GRU	CNN-LSTM	Proposed
Accuracy	0.843	0.855	0.858	0.901	0.962
F1-score	0.855	0.850	0.875	0.911	0.970
Precision	0.835	0.856	0.855	0.902	0.965
Recall	0.865	0.842	0.835	0.908	0.966
Sensitivity	0.860	0.855	0.833	0.911	0.977
Specificity	0.866	0.861	0.855	0.912	0.985

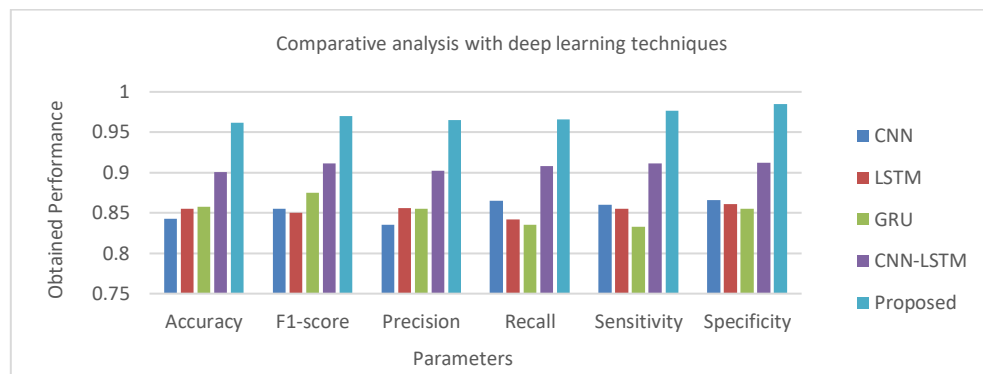


Figure 9. Comparative analysis with different deep learning methods

5. CONCLUSION

This study explored the effectiveness of multimodal authentication leveraging ResNet and VGG16 transfer learning for feature extraction, followed by CNN for classification. The real-time dataset comprised ECG signals and face lip images from individuals, representing distinct biometric modalities. Through the integration of ResNet and VGG16 transfer learning techniques, rich and discriminative features were extracted from the ECG signals and face lip images. Leveraging the pre-trained models allowed for the utilization of deep hierarchical representations, enhancing the robustness and discriminative power of the extracted features. Subsequently, a CNN classifier was employed to classify the extracted features and authenticate users based

on their biometric traits. The CNN model was trained on the fused feature vectors obtained from both modalities, allowing for comprehensive and accurate authentication. Overall, the integration of multimodal authentication utilizing ECG signals and face lip images, coupled with transfer learning and CNN classification, presents a promising approach for enhancing security in authentication systems. Future research could explore further optimization and refinement of the proposed methodology, as well as its applicability in real-world scenarios.

ACKNOWLEDGMENTS

Authors acknowledge the support from SSAHE University for the facilities provided and thanks to reviewers for their valuable suggestions.

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
Latha Krishnamoorthy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Ammasandra	✓	✓		✓	✓	✓		✓		✓	✓	✓		✓
Sadashivaiah Raju														

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing-Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing-Review & Editing	

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest related to this work.

DATA AVAILABILITY

The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] B. A. El-Rahiem, F. E. A. El-Samie, and M. Amin, "Multimodal biometric authentication based on deep fusion of electrocardiogram (ECG) and finger vein," *Multimedia Systems*, vol. 28, no. 4, pp. 1325–1337, 2022, doi: 10.1007/s00530-021-00810-9.

[2] W. Yang, S. Wang, J. J. Kang, M. N. Johnstone, and A. Bedari, "A linear convolution-based cancelable fingerprint biometric authentication system," *Computers and Security*, vol. 114, 2022, doi: 10.1016/j.cose.2021.102583.

[3] V. Wati, K. Kusrini, H. Al Fatta, and N. Kapoor, "Security of facial biometric authentication for attendance system," *Multimedia Tools and Applications*, vol. 80, no. 15, pp. 23625–23646, 2021, doi: 10.1007/s11042-020-10246-4.

[4] A. Musa, K. Vishi, and B. Rexha, "Attack analysis of face recognition authentication systems using fast gradient sign method," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 1346–1360, 2021, doi: 10.1080/08839514.2021.1978149.

[5] A. Wells and A. B. Usman, "Trust and voice biometrics authentication for internet of things," *International Journal of Information Security and Privacy*, vol. 17, no. 1, 2023, doi: 10.4018/IJISP.322102.

[6] D. Marzorati, D. Bovio, C. Salito, L. Mainardi, and P. Cerveri, "Chest wearable apparatus for cuffless continuous blood pressure measurements based on PPG and PCG signals," *IEEE Access*, vol. 8, pp. 55424–55437, 2020, doi: 10.1109/ACCESS.2020.2981300.

[7] C. Yaacoubi, R. Besrour, and Z. Lachiri, "A multimodal biometric identification system based on ECG and PPG signals," in *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, New York, United States: ACM, 2020, pp. 1–6, doi: 10.1145/3423603.3424053.




[8] X. Zhang, D. Cheng, P. Jia, Y. Dai, and X. Xu, "An efficient android-based multimodal biometric authentication system with face and voice," *IEEE Access*, vol. 8, pp. 102757–102772, 2020, doi: 10.1109/ACCESS.2020.2999115.

[9] P. S. V. V. N. Chanukya and T. K. Thivakaran, "Multimodal biometric cryptosystem for human authentication using fingerprint and ear," *Multimedia Tools and Applications*, vol. 79, no. 1–2, pp. 659–673, 2020, doi: 10.1007/s11042-019-08123-w.




- [10] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 25527–25542, 2019, doi: 10.1109/ACCESS.2018.2886573.
- [11] F. Ahamed, F. Farid, B. Suleiman, Z. Jan, L. A. Wahsheh, and S. Shahrestani, "An intelligent multimodal biometric authentication model for personalised healthcare services," *Future Internet*, vol. 14, no. 8, 2022, doi: 10.3390/fi14080222.
- [12] S. Itani, S. Kita, and Y. Kajikawa, "Multimodal personal ear authentication using acoustic ear feature for smartphone security," *IEEE Transactions on Consumer Electronics*, vol. 68, no. 1, pp. 77–84, 2022, doi: 10.1109/TCE.2021.3137474.
- [13] H. Purohit and P. K. Ajmera, "Optimal feature level fusion for secured human authentication in multimodal biometric system," *Machine Vision and Applications*, vol. 32, no. 1, 2021, doi: 10.1007/s00138-020-01146-6.
- [14] M. Singhal and K. Shinghal, "Secure deep multimodal biometric authentication using online signature and face features fusion," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 30981–31000, 2024, doi: 10.1007/s11042-023-16683-1.
- [15] S. P. Singh and S. Tiwari, "A dual multimodal biometric authentication system based on WOA-ANN and SSA-DBN techniques," *Sci*, vol. 5, no. 1, 2023, doi: 10.3390/sci5010010.
- [16] F. Cherifi, K. Amroun, and M. Omar, "Robust multimodal biometric authentication on IoT device through ear shape and arm gesture," *Multimedia Tools and Applications*, vol. 80, no. 10, pp. 14807–14827, 2021, doi: 10.1007/s11042-021-10524-9.
- [17] A. Kaul, A. S. Arora, and S. Chauhan, "AI-based approach for person identification using ECG biometric," in *AI and Deep Learning in Biometric Security*, CRC Press, 2021, pp. 133–153. doi: 10.1201/9781003003489-6.
- [18] J. S. Kim, M. G. Kim, and S. B. Pan, "Two-step biometrics using electromyogram signal based on convolutional neural network-long short-term memory networks," *Applied Sciences*, vol. 11, no. 15, 2021, doi: 10.3390/app11156824.
- [19] R. K. Grace, S. G. Devasena, and R. Manimegalai, "BABW: Biometric-based authentication using DWT and FFNN," in *Tele-Healthcare: Applications of Artificial Intelligence and Soft Computing Techniques*, Wiley, 2022, pp. 201–220. doi: 10.1002/9781119841937.ch9.
- [20] M. C. Younis and H. Abuhammad, "A hybrid fusion framework to multi-modal bio metric identification," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 25799–25822, 2021, doi: 10.1007/s11042-021-10818-y.
- [21] A. I. Siam *et al.*, "Enhanced user verification in IoT applications: a fusion-based multimodal cancelable biometric system with ECG and PPG signals," *Neural Computing and Applications*, vol. 36, no. 12, pp. 6575–6595, 2024, doi: 10.1007/s00521-023-09394-z.
- [22] D. Jeong, E. Choi, H. Ahn, E. Martinez-Martin, E. Park, and A. P. D. Pobil, "Multi-modal authentication model for occluded faces in a challenging environment," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 5, pp. 3463–3473, 2024, doi: 10.1109/TETCI.2024.3390058.
- [23] A. A. Aleidan *et al.*, "Biometric-based human identification using ensemble-based technique and ECG signals," *Applied Sciences*, vol. 13, no. 16, 2023, doi: 10.3390/app13169454.
- [24] S. B. Jadhav, N. K. Deshmukh, and V. T. Humbe, "HDL-PI: hybrid DeepLearning technique for person identification using multimodal finger print, iris and face biometric features," *Multimedia Tools and Applications*, vol. 82, no. 19, pp. 30039–30064, 2023, doi: 10.1007/s11042-022-14241-9.
- [25] S. K. S. Modak and V. K. Jha, "A novel multimodal biometric authentication framework using rule-based ANFIS based on hybrid level fusion," *Wireless Personal Communications*, vol. 128, no. 1, pp. 187–207, 2023, doi: 10.1007/s11277-022-09949-8.
- [26] E. F. Zohra and E. K. Hamada, "Facial emotion recognition based on upper features and transfer learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 37, no. 1, pp. 530–539, 2025, doi: 10.11591/ijeecs.v37.i1.pp530-539.
- [27] W. Farrukh and D. van der Haar, "Lip print-based identification using traditional and deep learning," *IET Biometrics*, vol. 12, no. 1, pp. 1–12, Jan. 2023, doi: 10.1049/bme2.12073.
- [28] N. N. Zanje, A. M. Bongale, and D. Dharrao, "Detecting facial image forgeries with transfer learning techniques," *International Journal of Advances in Applied Sciences*, vol. 13, no. 1, pp. 93–105, 2024, doi: 10.11591/ijaas.v13.i1.pp93-105.

BIOGRAPHIES OF AUTHORS



Latha Krishnamoorthy    received the M.Tech. degree in electronics and communication engineering from Sri Siddhartha institute of Technology, Visvesvaraya Technological University, Belgaum, Karnataka, India in 2009. Currently she is working as assistant professor in the Department of Bio-Medical Engineering at Sri Siddhartha Institute of Technology, SSAHE University, Karnataka, India. Her research interests include pattern recognition, bio-medical signal processing, and machine learning. She can be contacted at email: lathak@ssit.edu.in.



Dr. Ammasandra Sadashivaiah Raju    received M.Tech. degree in bio-medical instrumentation from SJCE, Mysore from VTU, Belgaum in 2004 and a Ph.D. in 2018 from VTU, Belgaum. Currently he is serving as professor head of the Department of Bio-Medical Engineering at Sri Siddhartha Institute of Technology, SSAHE University, Karnataka, India. His area of research includes biomedical signal and image processing, pattern recognition, computer vision. He can be contacted at email: rajuas@ssit.edu.in.