Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 3, March 2024, pp. 1524~1535
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i3.pp1524-1535

An ensemble approach for electrocardiogram and lip features based biometric authentication by using grey wolf optimization

Latha Krishnamoorthy, Ammasandra Sadashivaiah Raju
Department of Bio-Medical Engineering, Sri Siddhartha Institute of Technology, SSAHE University, Tumkur, India

Article Info

Article history:

Received Oct 20, 2023
Revised Nov 17, 2023
Accepted Dec 25, 2023

Keywords:

Biometric authentication
ECG filtering
Grey wolf optimization
lip movement
SIFT
Wavelettransform

ABSTRACT

In the pursuit of fortified security measures, the convergence of multimodal biometric authentication and ensemble learning techniques have emerged as a pivotal domain of research. This study explores the integration of multimodal biometric authentication and ensemble learning techniques to enhance security. Focusing on lip movement and electrocardiogram (ECG) data, the research combines their distinct characteristics for advanced authentication. Ensemble learning merges diverse models, achieving increased accuracy and resilience in multimodal fusion. Harmonizing lip and ECG modalities establishes a robust authentication system, countering vulnerabilities in unimodal methods. This approach leverages ECG's robustness against spoofing attacks and lip's fine-grained behavioral cues for comprehensive authentication. Ensemble learning techniques, from majority voting to advanced methods, harness the strengths of individual models, improving accuracy, reliability, and generalization. Moreover, ensemble learning detects anomalies, enhancing security. The study incorporates ECG signal filtering and lip region extraction as preprocessing, uses wavelet transform for ECG features, SIFT for lip image features, and employs grey-wolf optimization for feature selection. Ultimately, a voting-based ensemble classifier is applied for classification, showcasing the potential of this integrated approach in fortified security measures.

Corresponding Author:

Latha Krishnamoorthy
Department of Bio-Medical Engineering, Sri Siddhartha Institute of Technology, SSAHE University
Karnataka, India
Email: lathak@ssit.edu.in

1. INTRODUCTION

The escalating need for heightened security measures has sparked an unparalleled surge in the exploration of automated personal authentication methods rooted in biometrics. Biometrics encompasses the array of technologies harnessing physiological or behavioral attributes to validate an individual's identity. However, unimodal systems relying solely on a single data source for identification often encounter constraints such as insufficient distinctiveness, limited universality, and susceptibility to data noise. As a consequence, these systems may fall short of meeting the performance benchmarks expected in practical scenarios. Essentially, the process of confirming and validating a user's identity can be accomplished through three primary methods: knowledge-based (requiring something the user knows), possession-based (involving something the user possesses), and biometric-based (utilizing something inherent to the user's physiology or behaviour). While the initial two methods are commonly integrated into numerous IT systems, they encounter established difficulties. During last decade, the biometrics has transformed into an immensely popular research domain. In its most basic interpretation, biometrics involves quantifying the physical and behavioral attributes of an individual [1]. This process facilitates the identification and authentication of a person by utilizing the verifiable and recognizable biometric information. This is achieved by encompassing physiological attributes which includes traits like facial features, fingerprints, iris patterns, and vocal characteristics [2]. These biometric markers encompass distinct and quantifiable attributes employed for the categorization and delineation of individuals [3]. Behavioral biometrics serve to ascertain human identity based on their singular interactions with technology and devices. These biometric systems are used in wide range of real-time applications such as banks, hospitals, and working places. Due to excessive demand of these systems, the worldwide biometric market is projected to expand from 10.60 billion USD in 2016 to an anticipated 41.39 billion USD by 2025, reflecting a compound annual growth rate of 17.06% between 2017 and 2025 [3].

Currently, the biometric systems have gained enormous attention from research, and academia industry due to its safety and convenience. The biometric systems accomplish the authentication tasks in several phases such as initially user enrols required biometric information, this information is obtained and managed by the administrator module which formulates a template of information. Finally, the new information is compared with the enrolled template to verify the authenticity of user. Zulfiqar *et al*. [4] used face data as main entity for authentication, in [5] authors used fingerprint based authentication system, in [6] authors introduced face recognition based authentication system. Similarly, several works have been developed such as iris-retina recognition [7] and voice recognition system [8]. Relying solely on a single biometric factor can lead to a decrease in the accuracy of the authentication system. This is primarily attributed to issues such as inadequate data quality, instances of identities overlapping, and constraints in resources when attempting to establish a distinct identification for each individual. In order to overcome the issues researchers have introduced multi-modal authentication system where combination of two or more modalities are used to extract the features to improve the reliability of authentication. Zhang *et al*. [9] developed a biometric system with Face and Voice data. El_Tokhy [10] used fingerprint, iris and voice features fusion to enhance the accuracy. However, these modalities are used widely in various researches and generating a linear performance. Currently, electrocardiogram (ECG) has gained attention in various biomedical systems and authentication systems [11]. The ECG signal constitutes an electrical manifestation originating from the heart's activity. Its measurement is non-intrusively achievable through sensors affixed to the human chest [12]. Recently, it has garnered attention as an authentication tool, as it aligns closely with the majority of biometric criteria. A key attribute of ECG-based authentication is its capacity to ascertain aliveness, setting it apart from other authentication methods like fingerprints and passwords, which can solely be employed by living individuals. Moreover, ECG authentication offers inclusivity by accommodating a broader demographic, encompassing amputees and individuals with disabilities who might be unable to furnish conventional biometric markers such as fingerprints, palm prints, or iris scans.

Similarly, Lip is also considered as another modality to extract the features for authentication. Cruz *et al*. [13] used combined Lip based biometric with Viola Jones and appearance based model (AAM) system. Wright and Stewart [14] presented visual lip-based biometric authentication system. Farrukh *et al*. [15] used deep learning based methods for lop print identification. Similarly, Vasavi and Abirami [16] also deep learning based UNet architecture to recognize the lip patterns. These studies shows that the lip based biometric systems can be beneficial for real-time applications. However, achieving accuracy in real-time scenarios is a challenging task. Therefore, in this work we focus on development of a novel multimodal authentication system by using ECG and Lip attributes. The main contributions of this work are as follows:

- We incorporate ECG filtering scheme to handle various types of noises and apply feature extraction approach.
- We apply data augmentation mechanism on Lip images and apply a robust feature extraction mechanism.
- The obtained features are further processed through the feature selection phase where we reduce the feature dimensionality to minimize the computational complexity.
- The obtained features are fused together through the feature fusion scheme and employ a majority voting based ensemble classifier to obtain the robust accuracy.

Rest of the manuscript is organized as follows: section 2 describes the brief literature review about existing multimodal authentication systems. Section 3 presents proposed ensemble machine learning based authentication system. Section 4 presents the comparative analysis of proposed approach and finally, section 5 presents the concluding remarks and future scope of this approach.

2. LITERATURE SURVEY

This section presents the brief overview of existing biometric based authentication system. Current advancements have reported that the multimodal biometric authentication system can be beneficial. Prabu *et al.* [17] focused on development of multimodal biometric authentication system by using hybrid feature fusion technique. These features are obtained from Hand geometry and iris images where local binary patterns and scale invariant fourier transform (SIFT) based feature extraction methods are employed. Finally, neural networks and bayes networks classifiers are applied for classification. Ahamed *et al.* [18] authors developed biometric system for personalized healthcare systems by using electrocardiogram (ECG) and PPG (Photoplethysmography) signals. This approach uses time domain and combined time-frequency domain feature extraction which includes autoregressive coefficient, Shannon entropy and wavelet packet transform techniques. The obtained features are then trained by using CNN-LSTM classifier. Itani *et al*. [19] reported that face based biometric system face difficulty when person is wearing mask, similarly, fingerprint based authentication fail when user hands are wet. To tackle this problem, authors introduced ear authentication system. Similar to this Purohit *et al.* [20] used palm, fingerprint, and ear based multimodal authentication system. This approach uses texture and shape feature extraction process by using Gabor feature for palm images, HMSB operator for fingerprint and HMSB and MRG for ear. Later, oppositional gray wolf optimization scheme is applied for efficient feature selection and finally, multi-kernel SVM classifier is used for recognition. Similar to [19], Cherifiet *et al.* [21] developed ear and arm gesture based multimodal authentication system.

For ear data, this method uses local phase quantization (LPQ) as feature extraction process whereas arm gesture features are extracted by using statistical features. Further, a weighted score fusion method is used to generate the feature vector for classification. Chanukya Thivakaran. [22] developed an authentication system by using fingerprint and ear. This process is comprised of pre-processing, feature extraction and classification. The preprocessing stage employs a median filter to facilitate image cropping, aiding in the selection of the desired position. Subsequently, texture and shape features are extracted from the preprocessed fingerprint and ear images. Ultimately, these extracted features are combined, leading to the integration of features. This integrated feature set is effectively classified using an optimal neural network (ONN). Notably, the weights of the neural network are chosen optimally through the application of the firefly algorithm (FF). Krishna *et al.* [23] used EEG and Eye tracking method for multimodal biometrics. Ganapathi *et al.* [24] reported that the ear's performance often surpasses or matches that of other biometric characteristics.

Nevertheless, recognition accuracy for 2D ear images diminishes due to posture, lighting, and scale variations. To address these issues, we suggest amalgamating 3D and 2D ear images. A keypoint detector and descriptor are constructed, leveraging a covariance matrix from texture and geometry data. The descriptor's feature vector aptly characterizes keypoints and is employed to identify the closest keypoints between a probe and gallery. This step facilitates subsequent registration. The resultant registration error serves as the matching score. Abozaid *et al.* [25] presented a proficient approach for human authentication is established through the fusion of face and voice recognition modalities, resulting in an effective multimodal biometric identification system. Voice recognition features are obtained via cepstral coefficients and statistical coefficients, and the subsequent comparison of these coefficients takes place. In parallel, facial recognition features are extracted using two distinct techniques: Eigenface and principal component analysis (PCA), followed by a comparative analysis of the outcomes. The identification process for both voice and face modalities is executed through three distinct classifiers: gaussian mixture model (GMM), artificial neural network (ANN), and support vector machine (SVM). Hammad and wang [26] introduced a robust multimodal biometric system by combining electrocardiogram (ECG) and fingerprint data through a convolutional neural network (CNN) framework. CNNs are utilized for feature extraction from each modality, resulting in the generation of biometric templates derived from these features. During the authentication phase, an innovative Q-Gaussian multi support vector machine (QG-MSVM) classifier is put forth to enhance the overall authentication performance.

3. PROPOSED METHOD

This section presents the proposed model for multimodal authentication system by using ECG and Lip data. The proposed approach comprised of following phases: pre-processing, feature extraction, feature fusion, feature selection, and ensemble classification.

- Pre-processing: This phase is performed for both modalities where for ECG, we employ filtering scheme because during signal acquisition these signals gets contaminated due to unwanted noises which may impact the quality of extracted features. On the other side, face image is taken as input and pre-processing phase considers the extraction of only lip image as lip segmentation.
- Feature extraction: In this stage we consider filtered ECG signal and segmented lip image for extracting the robust features. In this work, we consider several different types of features such as temporal features, morphological features, statistical features, frequency domain features, and heart rate variabilityfeatures.
- Feature Fusion and selection: Once the feature extraction phase is completed, we focus on fusing the extracted features and fuse them to formulate the feature vector. However, this feature vector may consist

of some redundancy therefore, we apply feature selection scheme to select the most significant attributes. Thus, the final feature vector is generated

- Ensemble classification: This process includes random forest, decision tree and neural network classifiers and employs the majority voting mechanism to obtain the final classification result. Figure 1 depicts the complete process of proposed approach where ECG and face data are given as input to the model which are process through various stages as mentioned before to obtain the final classification.

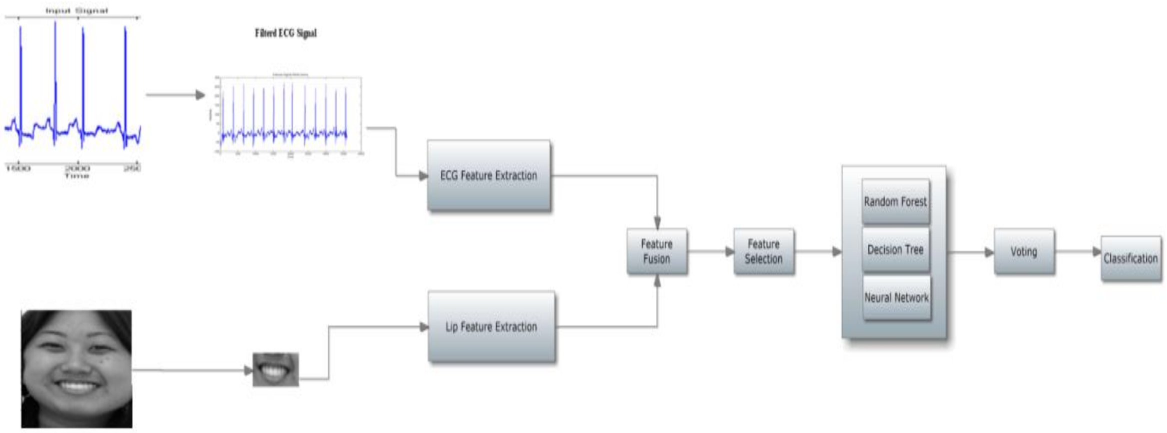


Figure 1. ECG and Lip based multimodal authentication system

3.1. Pre-processing

As discussed in Figure 1, the ECG signals gets contaminated during acquisition therefore applying filtering and removing the noise is considered as essential step. In this work we consider recursive least square (RLS) approach as base approach and introduced Adaptive recursive least square (RLS) which is obtained by modifying the notch filters based on least mean squares approach. This approach is implemented in following stages:

- Step 1: by analysing the ECG signal, we extract the time-varying frequency information of cardiac artifacts from the signal and this can be modelled as periodic signal. The frequency and phase of signal can be obtained by identifying the R peaks and sampling rate.

$$f_i = \begin{cases} \frac{F_S}{n_2 - n_1}, n \leq n_1 \\ \frac{F_S}{n_{i+1} - n_i}, n_i < n \leq n_{i+1} \\ \frac{F_S}{n_m - n_{m-1}}, n > n_m \end{cases} \text{ And } \phi(n) = \begin{cases} \frac{2\pi(n - n_1)}{n_2 - n_1}, n \leq n_1 \\ \frac{2\pi(n - n_i)}{n_{i+1} - n_i}, n_i < n \leq n_{i+1} \\ \frac{2\pi(n - n_m)}{n_m - n_{m-1}}, n > n_m \end{cases} \quad (1)$$

Where f and ϕ denotes the frequency and phase of ECG signal, and m denotes the total number of R peaks

- Step 2: based on the time-varying information, we generate the reference signal $s_{ref}(n)$ for N harmonics expressed as:

$$s_{ref}(n) = \begin{bmatrix} s_1(n) \\ s_Q(n) \end{bmatrix}$$

$$s_1(n) = \left[\cos(\phi(n)), \cos(2\phi(n)), \dots, \cos(N\phi(n))\right]^T$$

$$s_Q(n) = \left[\sin(\phi(n)), \sin(2\phi(n)), \dots, \sin(N\phi(n))\right]^T \quad (2)$$

Where N denotes the considered harmonics, $(.)^T$ is the transpose of matrix, s_1 is in-phase and s_Q is the quadrature phase.

- Step 3: the ECG artifacts can be expressed as:

$$\hat{s}_{ECG} = w(n)^T s_{ref}(n) = a(n)^T s_1(n) + b(n)^T s_Q(n)$$
$$= \sum_{k=1}^{N} a_k(n) \cos(k\phi(n)) + b_k(n) \sin(k\phi(n))$$

- Step 4: during the RR interval the ECG artifact can be expressed by using Fourier transform in trigonometricform using harmonics of time-varying in-phase and quadrature coefficients. These coefficients can be used for to track the changes in waveform. Then the clean ECG can be obtained by subtracting the artifacts from original signal and employ RLS to update the filter coefficients. This can be expressed as:

$$\hat{s}_{EEG}(n) = s_{EEG}(n) - \hat{s}_{EEG}$$

On the other hand, we apply lip extraction mechanism to extract the lip region and employ feature extraction process as mentioned in next section. In this work, we have used python Dlib library for facial landmark detection [27]. In Figure 2 depicts the representation of facial landmark analysis where Figure 2(a) depicts the facial landmarks superimposed on facial image and Figure 2(b) depicts the only facial landmark points.

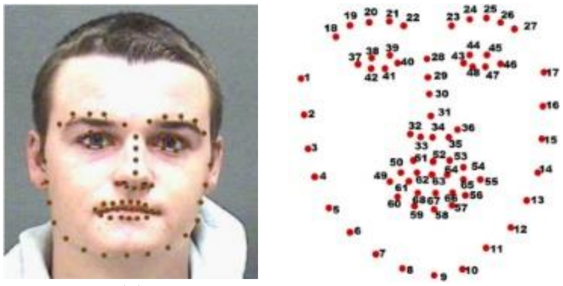


Figure 2. Sample outcome of facial landmark detection (a) facial landmark and (b) only landmark

3.2. Feature extraction

This section presents the feature extraction process for ECG and lip data. First of all, we present the feature extraction process for ECG signals. First of all, we apply Pan-Tompkins approach for ECG peak detection which can be used as basic features of these signals. Below given Figure 3 depicts the outcome of peak detection.

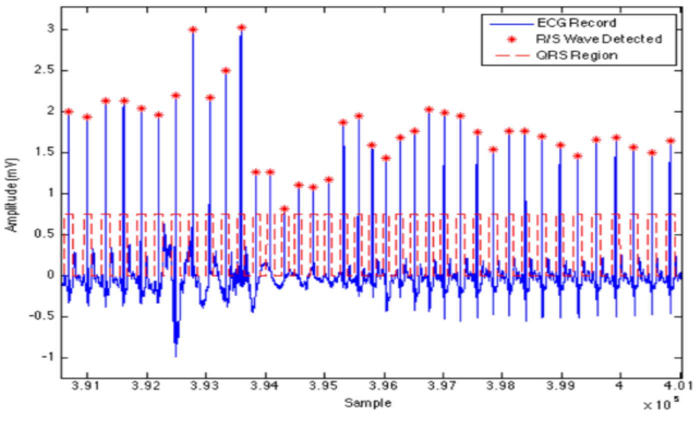


Figure 3. ECG peak detection

Further, we apply wavelet transform based approach for feature extraction. The wavelet transform approach is employed to extract the coefficient of transform of each segment. The wavelet transform of a signal $f(x)$ is expressed as:

$$W_s f(x) = f(x) * \psi_s(x) = \frac{1}{s}\int_{-\infty}^{+\infty} f(t)\psi\left(\frac{x-t}{s}\right) dt$$

where s represents the scale factor, $\psi_s(x) = \frac{1}{s}\psi\left(\frac{x}{s}\right)$ represents the dilation of wavelet. Let $s = 2^j$ then this wavelet is represented as dyadic wavelet transform. This can be expressed with the help of Mallat algorithm as:

$$S_2, f(n) = \sum_{k\in Z} h_k S_{2j-1} f(n - 2^j k)\ W_2, f(n) = \sum_{k\in Z} g_k S_{2j-1} f(n - 2^j k)$$

where S_2 represents the smoothing operator, $S_2, f(n)$ is the low frequency coefficient which denotes the approximation of original signal, $\omega_2\ , f(n)$ represents the high frequency coefficients. It is established that the wavelet transform (WT) is more suitable for examining signals that change over time, and the discrete wavelet transform adapts to the time-frequency characteristics of a given pattern. Through a multi-resolution approach, it becomes feasible to succinctly represent signal structures using only a handful of coefficients in the wavelet domain. The selection of the appropriate wavelet function and the determination of the number of decomposition levels are crucial factors when employing the WT for signal analysis. The choice of decomposition levels relies on the dominant frequency components present in the signal. These levels are decided in a manner that preserves signal portions closely associated with the frequencies necessary for signal classification within the wavelet coefficients.

On the hand, we apply SIFT feature extraction on lip images. The process of SIFT feature extraction is descried as follows:

- Scale-Space Representation: SIFT constructs a scale-space pyramid by repeatedly convolving the original image with Gaussian kernels at different scales. Mathematically, this can be represented as:

  $$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

  where$L(x, y, \sigma)$ is the scale-space image at scale σ, $G(x, y, \sigma)$ is the Gaussian kernel, and $I(x, y)$ is the original image.
  The difference-of-Gaussian (DoG) images are obtained by subtracting adjacent scales from the scale-space pyramid:

  $$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

- Keypoint Detection: SIFT detects potential keypoints as local extrema in the DoGimages across scales and image locations. This involves comparing a pixel's value with its 26 neighbors in the current and adjacent scales.
- Orientation Assignment: For each key-point, SIFT computes a dominant orientation to achieve rotation invariance. This is done by calculating gradient magnitudes and orientations in a circular neighborhood around the key-point. A histogram of gradient orientations is created, and the peak(s) indicate the dominant orientation(s).
- Descriptor Computation: The region around the keypoint is divided into smaller subregions (e.g., 4x4 or 8x8 grids). For each subregion, a histogram of gradient orientations is generated. These histograms are concatenated to form the final SIFT descriptor vector. The descriptor calculation involves the following steps for each sub-region:

Calculate gradient magnitudes (M) and orientations (θ) for the pixels within the subregion.
Distribute the gradient orientations into orientation bins in the histogram. Weight the histogram entries by the gradient magnitudes to account for gradient strength.
The resulting concatenated histogram forms the SIFT descriptor for that keypoint. Further, we estimate Statistical feature for lip images. Table 1 shows the mathematical expression to estimate feature value.

Table 1. Mathematical expression to estimate feature value

Feature	Expression
Skewness	$\frac{E(x-\mu)^3}{\sigma^3}$
Perimeter	$bwperim(BW)$
Eccentricity	$e = \left(\frac{1-b^2}{a^2}\right)\cdot\frac{1}{2}$
Standard Deviation	$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\|Ai-\mu\|^2}$

Finally, the obtained features are combined together to generate the fused vector. This is done by applying concatenating rule. In next phase, we apply grey wolf-based optimization scheme for feature selection. The feature selection is adapted.

3.3. Ensemble classification

The obtained features are further processed through the traditional machine learning classification models which include support vector machine, decision Tree and neural network. In order to obtain the final classification decision, we employ majority voting mechanism. Majority voting is a simple and effective classification technique used in ensemble learning, where multiple classifiers (or models) are combined to make a final decision. The mathematical model for majority voting classification can be described as follows: Let us consider that we have total N number of classifiers used for classification, denoted as $C_1, C_2, \ldots C_N$. Each classifier is trained on the same attributes obtained by fusing the handcrafted and deep learning features. Moreover, each classifier is capable to classify the patterns in K different classes. For any given input sample X, the majority voting classification can be employed as follows:

- Obtain the individual predictions from each classifier for the input sample X and their corresponding predicted class labels are denoted as $P_1, P_2, \ldots, P_N$ from the class $C_1, C_2, \ldots C_N$, respectively.
- Count the occurrences of each class label in the set $\{P_1, P_2, .. P_N\}$.
- Determine the class label with the highest frequency (i.e., the majority class). If there is a tie, a predefined tie-breaking rule may be used (e.g., select the class with the lowest index).
- Assign the majority class label as the final classification output for the input sample X. It can be expressed as follows: $Final\ Classification = \arg\max_k \sum_{i=1}^{N} 1(P_i = k)$.

4. RESULTS AND DISCUSSION

This section describes the outcome of proposed approach for a hybrid biometric authentication by considering ECG and Lip as the main entities for extracting the biometric information. The obtained performance is compared with other supervised classification schemes. The results are tabulated and analysed.

4.1. Dataset details

In order to construct the dataset, we have collected ECG and face images for 10 users. For each user 20 ECG signals. Generally, ECG signals get contaminated during acquisition, therefore filtering is essential for these signals. In this work, we have considered white noise, color noise, motion artifacts, electrode artifacts and baseline wander. Similarly, 20 face images are captured for each user. We have not considered occlusion scenarios because here main aim is to extract the lip region which is obtained by considering the face landmark detection by using "dlib" python library. However, we have employed data augmentation scheme to generate diverse dataset of same user. The augmentation helps to obtain robust features. Below given Figure 4 depicts the sample ECG signal where x-axis represents the "Time" and y-axis represents the "Amplitude", and Figure 5 demonstrates noise added signal (in first column) and their corresponding filtered signal (in second column). The noisy signals are depicted in Figure 5(a) white noise, Figure 5(b) color noise, Figure 5(c) motion artifact, Figure 5(d) electrode artifact, and Figure 5(e) baseline wander.

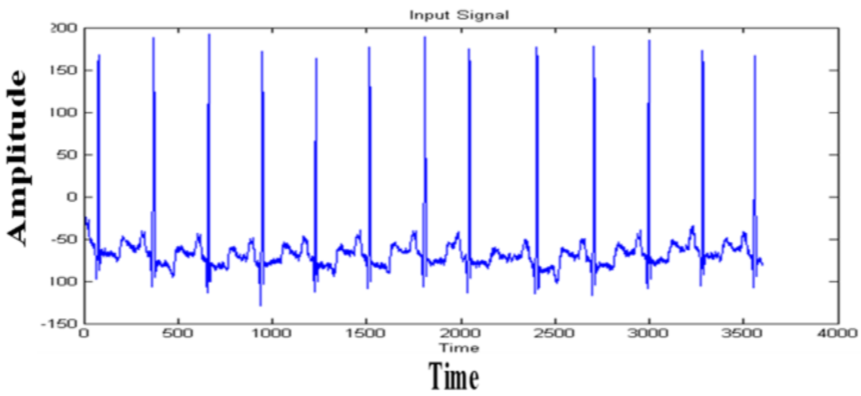
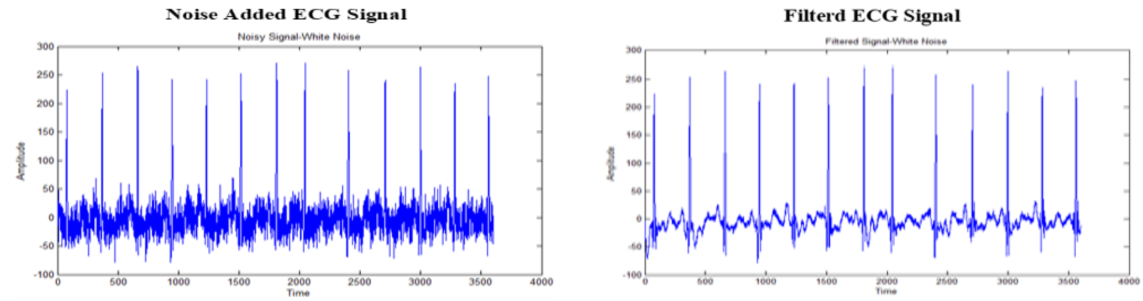
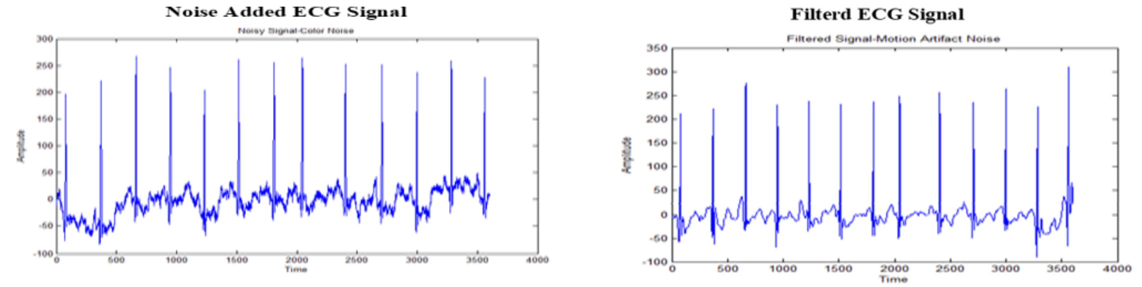


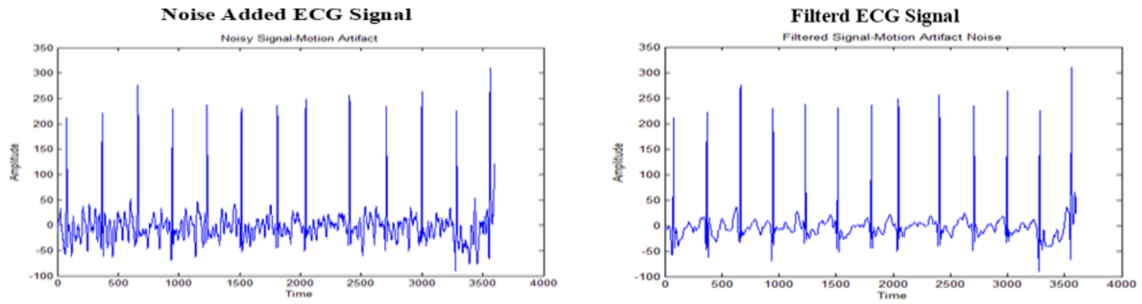
Figure 4. Sample ECG signal for input



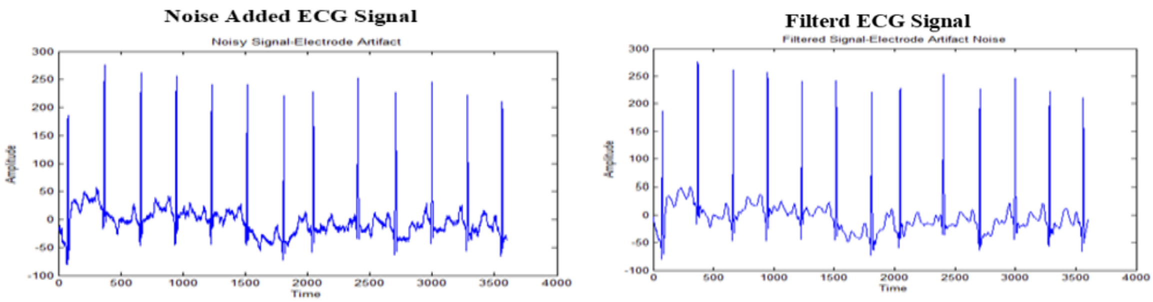
(a)



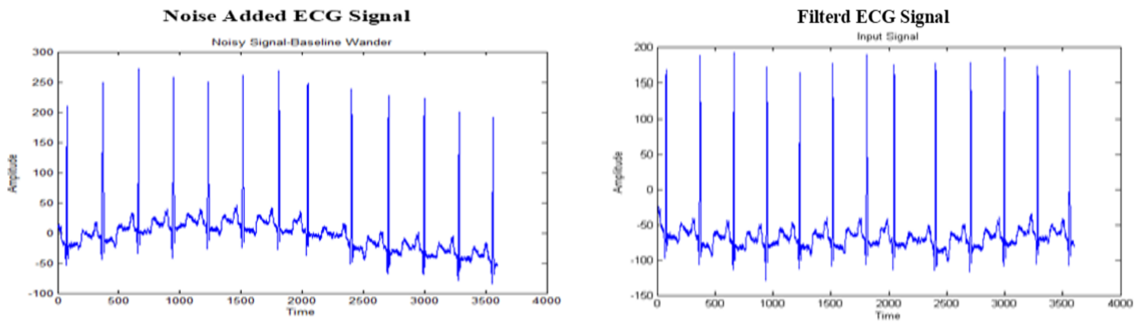
(b)



(c)



(d)



(e)

Figure 5. Noise added ECG signal and filtered signals for biometric authentication; (a) white noise, (b) color noise, (c) motion artifact, (d) electrode artifact, and (e) baseline wander

Similarly, we have considered face images and extracted the lip region by facial landmark detection. In Figure 6 depicts the face image, corresponding extracted lip and augmented images. The landmark detection process helps to obtain the coordinates of lip region which is cropped and augmented by applying image flipping horizontal and vertically.

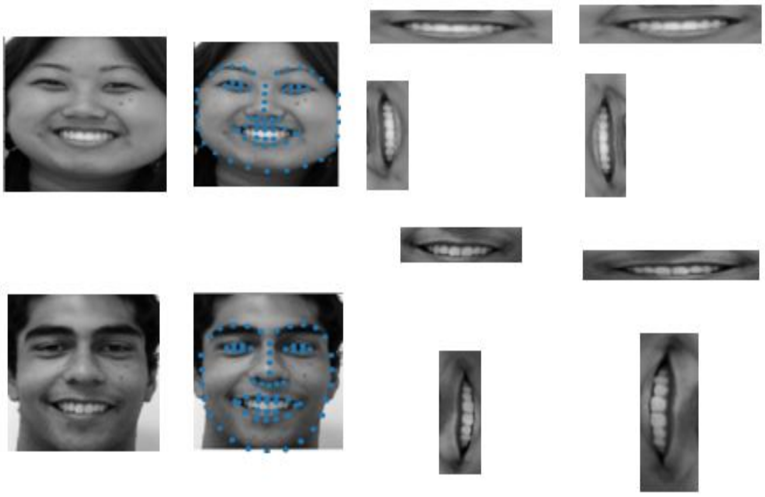


Figure 6. Face image, corresponding facial landmarks and augmentation

4.2. Performance measurement parameters

The performance of proposed approach is measured based on confusion matrix calculation. The confusion matrix generated with the help of true positive, false positive, false negative, and true negative. In Table 2 shows a sample representation of confusion matrix.

Based on this confusion matrix, we measure several statistical performance parameters such as accuracy, precision, F1-score by using proposed approach. Accuracy is the measurement of correct instance classification out of total number instances. The accuracy is measured as follows:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

then, we compute precision of the proposed approach. It is computed by taking the ratio of true positive and (true and false) positives.

$$P = \frac{TP}{TP+FP} \tag{4}$$

Finally, we compute the F-measure based on the sensitivity and precision values, which is expressed as (5).

$$F = \frac{2*P*Sensitivity}{P+Sensitivity} \tag{5}$$

4.3. Comparative analysis

Based on these performance measurement parameters, we measure the performance of proposed approach and compared the obtained performance with other supervised classification schemes. Confusion matrix generated by different algorithm solution as shown in Figure 7. Figures 7(a)-7(d) shows the confusion matrix obtained by applying support vector machine, neural network, decision tree, random forest and proposed ensemble classifier. The tabulated results can be compared.

Table 2. Confusion matrix classes

Actual class	Predicted class	
	Authentic	Authentic
Authentic	True Positive	True Positive
Imposter	False Positive	False Positive

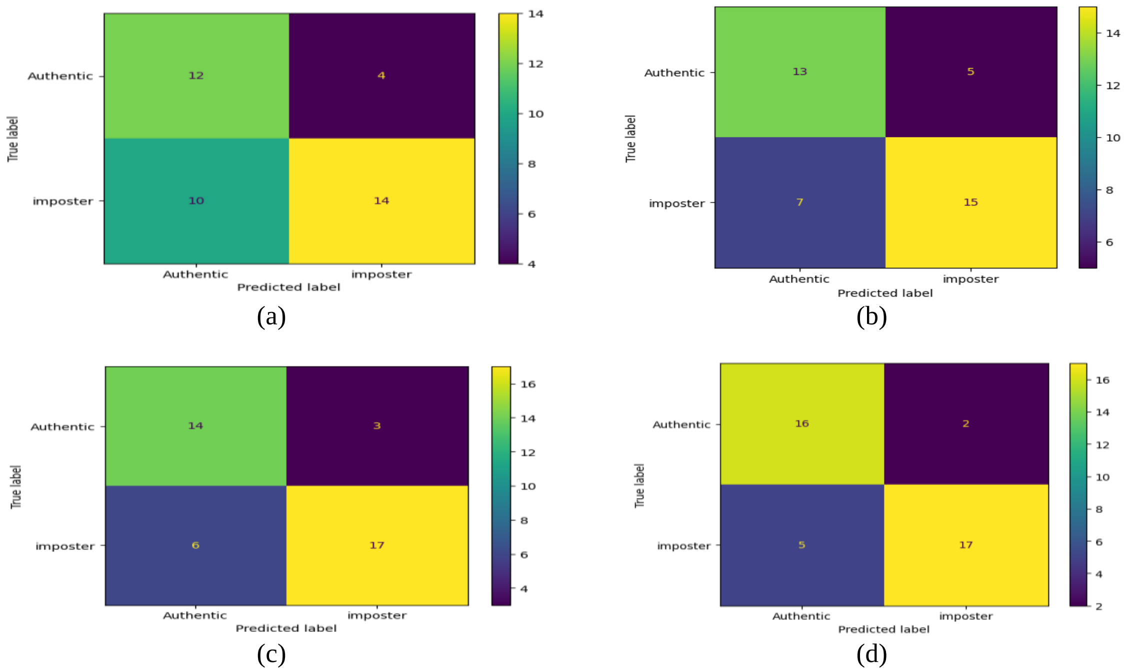


Figure 7. Confusion matrix generated by different algorithm solution; (a) SVM, (b) neural network, (c) random forest, and (d) decision tree

Based on these confusion matrices, we measure the performance of SVM and neural network as depicted in Table 3. This analysis is done by identifying the authentic and imposter classes by employing NN and SVM. Similarly, we measure the performant of random forest and decision tree classifier. Below given Table 4 shows the outcome of these two classifiers. Finally, proposed ensemble classifier is applied. Below given figure demonstrates the obtained confusion matrix. Based on this confusion matrix, we measure other performance measurement parameters as presented in Table 5. Figure 8 shows the confusion matrix.

Table 3. Results generated by SVM and neural network algorithm

	SVM			Neural network		
Parameter	Authentic	Imposter class	Average	Authentic	Imposter class	Average
Accuracy	0.65	0.65	0.65	0.70	0.70	0.70
F1-score	0.63	0.66	0.65	0.68	0.714	0.70
Precision	0.54	0.77	0.65	0.65	0.75	0.70
Recall	0.75	0.58	0.65	0.72	0.68	0.70
Sensitivity	0.75	0.58	0.65	0.72	0.68	0.70
Specificity	0.58	0.75	0.65	0.68	0.72	0.70
True negative rate	0.58	0.75	0.65	0.68	0.72	0.70
True positive rate	0.75	0.58	0.65	0.72	0.68	0.70
False negative rate	0.25	0.416	0.35	0.27	0.31	0..3
False positive rate	0.416	0.25	0.35	0.31	0.27	0.3

Table 4. Results generated by random forest and decision tree algorithm

	Random forest			Decision tree		
Parameter	Authentic	Imposter class	Average	Authentic	Imposter class	Average
Accuracy	0.77	0.77	0.775	0.825	0.825	0.825
F1-score	0.75	0.79	0.775	0.8205	0.8292	0.825
Precision	0.70	0.85	0.775	0.761	0.894	0.825
Recall	0.8235	0.739	0.775	0.888	0.7727	0.825
Sensitivity	0.823	0.739	0.775	0.888	0.7727	0.825
Specificity	0.739	0.823	0.775	0.7727	0.888	0.825
True negative rate	0.7391	0.8235	0.775	0.7727	0.888	0.825
True positive rate	0.8235	0.7391	0.775	0.888	0.727	0.825
False negative rate	0.176	0. 2608	0.225	0.111	0.227	0.175
False positive rate	0.260	0.1764	0.225	0.227	0.111	0.175

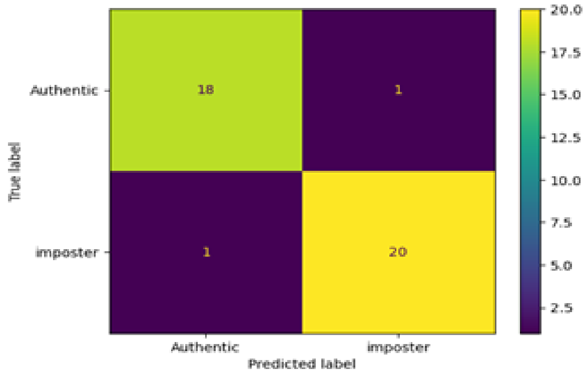


Figure 8. Obtained confusion matrix

Table 5. Results generated by proposed mechanism

	Proposed ensemble classifier		
Parameter	Authentic	Imposter	Average
Accuracy	0.95	0.95	0.95
F1-score	0.947	0.952	0.95
Precision	0.947	0.9523	0.95
Recall	0.947	0.952	0.95
Sensitivity	0.947	0.952	0.95
Specificity	0.952	0.9473	0.95
True negative rate	0.952	0.9473	0.95
True positive rate	0.947	0.9523	0.95
False negative rate	0.052	0.047	0.05
False positive rate	0.0476	0.052	0.05

5. CONCLUSION

In the realm of biometric authentication, the fusion of multiple modalities has emerged as a powerful strategy to address the limitations of single-modal systems. Lip and ECG data, representing distinct physiological and behavioural characteristics, offer complementary information that can significantly enhance the accuracy and security of authentication processes. The ensemble learning paradigm, which amalgamates the outputs of multiple models, further elevates the authentication performance by mitigating individual model weaknesses and biases. The integration of lip and ECG modalities into an ensemble learning framework is a noteworthy advancement in the field. Lips, being a visually distinctive feature, can capture unique behavioral traits. On the other hand, ECG data captures the electrical activity of the heart, providing a physiological dimension that is both difficult to replicate and less prone to external influence. Moreover, ensemble classification approach helps to obtain the better accuracy when compared with the other traditional machine learning algorithms.

REFERENCES

[1] A. F. Abate, L. Cimmino, I. Cuomo, M. Di Nardo, and T. Murino, "On the impact of multimodal and multisensor biometrics in smart factories," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 9092–9100, Dec. 2022, doi: 10.1109/TII.2022.3178376.

[2] M. Babaei, J. Shirzad, K. K. M. Pham, P. F. Fard, and A. Eftekhari, "Challenges of using biometric evidence in identification," *Journal of Police Medicine*, vol. 11, no. 1, 2022, doi: 10.30505/11.1.29.

[3] J. Mason, R. Dave, P. Chatterjee, I. Graham-Allen, A. Esterline, and K. Roy, "An investigation of biometric authentication in the healthcare environment," *Array*, vol. 8, p. 100042, Dec. 2020, doi: 10.1016/j.array.2020.100042.

[4] M. Zulfiqar, F. Syed, M. J. Khan, and K. Khurshid, "Deep face recognition for biometric authentication," in *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, IEEE, Jul. 2019, pp. 1–6. doi: 10.1109/ICECCE47252.2019.8940725.

[5] W. Yang, S. Wang, J. J. Kang, M. N. Johnstone, and A. Bedari, "A linear convolution-based cancelable fingerprint biometric authentication system," *Computers & Security*, vol. 114, p. 102583, Mar. 2022, doi: 10.1016/j.cose.2021.102583.

[6] E. Fourati, W. Elloumi, and A. Chetouani, "Anti-spoofing in face recognition-based biometric authentication using image quality assessment," *Multimedia Tools and Applications*, vol. 79, no. 1–2, pp. 865–889, Jan. 2020, doi: 10.1007/s11042-019-08115-w.

[7] V. Conti, L. Rundo, C. Militello, V. M. Salerno, S. Vitabile, and S. M. Siniscalchi, "A multimodal retina-iris biometric system using the Levenshtein distance for spatial feature comparison," *IET Biometrics*, vol. 10, no. 1, pp. 44–64, Jan. 2021, doi: 10.1049/bme2.12001.

[8] S. Duraibi, F. T. Sheldon, and W. Alhamdani, "Voice biometric identity authentication model for IoT devices," *International Journal of Security, Privacy and Trust Management*, vol. 9, no. 2, pp. 1–10, May 2020, doi: 10.5121/ijsptm.2020.9201.

[9] 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.
[10] M. S. El_Tokhy, "Robust multimodal biometric authentication algorithms using fingerprint, iris and voice features fusion," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 1, pp. 647–672, Jan. 2021, doi: 10.3233/JIFS-200425.
[11] A. N. Uwaechia and D. A. Ramli, "A comprehensive survey on ECG signals as new biometric modality for human authentication: recent advances and future challenges," *IEEE Access*, vol. 9, pp. 97760–97802, 2021, doi: 10.1109/ACCESS.2021.3095248.
[12] H. Bin Hwang, H. Kwon, B. Chung, J. Lee, and I. Y. Kim, "ECG authentication based on non-linear normalization under various physiological conditions," *Sensors*, vol. 21, no. 21, p. 6966, Oct. 2021, doi: 10.3390/s21216966.
[13] J. C. D. Cruz, R. G. Garcia, S. M. M. Go, J. L. Regala, and L. J. T. Yano, "Lip biometric authentication using Viola-Jones and appearance based model (AAM) system," in *ITC-CSCC 2020 - 35th International Technical Conference on Circuits/Systems, Computers and Communications*, 2020, pp. 372–377.
[14] C. Wright and D. W. Stewart, "Understanding visual lip-based biometric authentication for mobile devices," *EURASIP Journal on Information Security*, vol. 2020, no. 1, p. 3, Dec. 2020, doi: 10.1186/s13635-020-0102-6.
[15] 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.
[16] J. Vasavi and M. S. Abirami, "Pattern recognition of grooves in human lips for improved authentication in cyber-physical systems using U-Net architecture," *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 1, pp. 1683–1693, 2023, doi: 10.3233/JIFS-223488.
[17] S. Prabu, M. Lakshmanan, and V. N. Mohammed, "A multimodal authentication for biometric recognition system using intelligent hybrid fusion techniques," *Journal of Medical Systems*, vol. 43, no. 8, p. 249, Aug. 2019, doi: 10.1007/s10916-019-1391-5.
[18] 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, p. 222, Jul. 2022, doi: 10.3390/fi14080222.
[19] 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, Feb. 2022, doi: 10.1109/TCE.2021.3137474.
[20] 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, p. 24, Jan. 2021, doi: 10.1007/s00138-020-01146-6.
[21] 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, Apr. 2021, doi: 10.1007/s11042-021-10524-9.
[22] 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, Jan. 2020, doi: 10.1007/s11042-019-08123-w.
[23] V. Krishna, Y. Ding, A. Xu, and T. Höllerer, "Multimodal biometric authentication for VR/AR using EEG and eye tracking," in *Adjunct of the 2019 International Conference on Multimodal Interaction*, New York, NY, USA: ACM, Oct. 2019, pp. 1–5. doi: 10.1145/3351529.3360655.
[24] I. I. Ganapathi, S. S. Ali, U. Sharma, P. Tomar, M. Owais, and N. Werghi, "RHEMAT: robust human ear based multimodal authentication technique," *Computers & Security*, vol. 132, p. 103356, Sep. 2023, doi: 10.1016/j.cose.2023.103356.
[25] A. Abozaid, A. Haggag, H. Kasban, and M. Eltokhy, "Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16345–16361, Jun. 2019, doi: 10.1007/s11042-018-7012-3.
[26] M. Hammad and K. Wang, "Parallel score fusion of ECG and fingerprint for human authentication based on convolution neural network," *Computers & Security*, vol. 81, pp. 107–122, Mar. 2019, doi: 10.1016/j.cose.2018.11.003.
[27] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019, doi: 10.1109/ACCESS.2019.2906757.

BIOGRAPHIES OF AUTHORS

Latha Krishnamoorthy received the M. Tech degree in Electronics and Communication Engineering from Sri Siddhartha institute of Technology, Visweswaraiaha Technological University, Belgaum Karnataka, India in 2009.Currently she is working as Asst.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 fromVTU, Belgaum. Currently he is serving as Professor Head of the department of Bio-Medical Engineeringat 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.