Journal of Al-Qadisiyah for Computer Science and Mathematics Vol. 14(2) 2022 , pp Comp. 46–58

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

An intelligent model for ECG Classification System based on frequency domain with Least Square Support Vector Machine (LS-SVM)

Russel R. Majeed and Sarmad Kadhim

The University of Thi-Qar, College of Education for Pure Sciences, Iraq

Dr.sarmad@utq.edu.iq, msc20co10@utq.edu.iq

ARTICLE INFO

Article history:

Received: 10/05/2022

Rrevised form: 15 /06/2022

Accepted : 21 /06/2022

Available online: 30/06/2022

Keywords:

ECG, authentication, frequency feature, LSSVM

ABSTRACT

The development of authentication and identity mechanisms has become a vital requirement to secure device data integrity, although passwords provide sufficient control and authentication, they have shown major flaws in speed and security, Biometrics is becoming the primary authentication method. Electrocardiogram (ECG) signals are created as a consequence. Due to their unique character, which makes them difficult to falsify and ubiquitous, ECG signals have attracted much attention in most authentication systems. We introduce a novel ECG validation model that combines frequency domain-based characteristics with the least-squares technique in this work (LS-SVM). Two types of frequency field characteristics were examined in order to find the best combination of high precision and speed. ECG Signal Characteristics and Frequency Domain Characteristics ECG signals are used to determine the optimal triple-band filter bank. To access the most important characteristics, we removed the extraneous features and extracted others.

The listed features are entered into a spreadsheet known as the Least Square Support Vector Machine (LSSVM), and the results show that the ECG biometrics have been validated. The proposed model obtains an accuracy rate of 95 % and 92% straight frequency advantage points. A large set of data is used to evaluate the proposed model.

MSC..

https://doi.org/10.29304/jqcm.2022.14.2.966

R. R. MAJEED

msc20co10@utq.edu.iq

Communicated by 'sub etitor'

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

1. Introduction

Biometric identification methods provide a high-security approach for identifying and authenticating people based on their physical traits [1], [2]. In most biometrics systems, physiological and behavioral variables such as fingerprints, face, and voice are used [3], [4]. Although these elements give a high level of security, various investigations have shown that they are vulnerable to fabrication by attacks. Certain biometric system assaults, for example, employ latex to recreate some properties such as fingerprints and voices and use them as original features [5], [6] Biometric systems based on bio signals are often employed in clinical diagnostics, Electroencephalography (EEG) and electrocardiography (ECG) have recently acquired prominence [6], [7]. The bulk of modern biometrics systems allows the use of ECG signals. An ECG is a non-invasive tool that is effective, simple, and requires a low-cost process for obtaining cardiac muscle activity information from people [8], [9].

According to clinical research, each person has distinct ECG characteristics that may help protect them against falsifications. Imitating and copying ECG characteristics has also been proven to be difficult [10]-[11].Various government and commercial systems use ECG biometric technology, including travel papers, health monitoring systems, and innovative card systems. Despite the advantages and reliability of ECG biometric systems, techniques-based ECG signals have several limitations because of the time-varying nature of ECG signals, the rarity of cardiac diseases, and the time it takes to gather ECG data from people. Consequently, the performance of a biometric identification model is the most crucial factor to consider while evaluating it. ECG signals, however, are not periodic and are pretty repetitive. As a result, the feature extraction process, or how to choose the most significant attributes to describe ECG signals, is a considerable challenge in constructing a viable ECG biometric system. This study looks at time and frequency to determine the most excellent quality set for depicting ECG signals. The T wave, P wave, and QRS complex are the three essential components of an ECG wave. The amplitudes of the P, R, and T waves are fiducial characteristics. The beginning and offset of each P, Q, R, S, and T wave and the slope data are all included [8]. Several approaches for identifying people using ECG signals and other biometric traits like face and fingerprints have been developed. Not long ago, Al Alkeem, et al. [12] To identify persons, ECG, fingerprint, and face scans were merged. To extract characteristics from the ECG, fingerprint, and face picture, two kinds of neural networks were used: ResNet50

R. R. MAJEED

msc20co10@utq.edu.iq

Communicated by 'sub etitor'

and deep learning models. Hamza et al [7] Persons were identified using ECG readings. Three parameters were utilized to identify people: entropy, cepstral coefficient, and ZCR. An SVM was used to classify the obtained features.

The Physio Net database was used throughout the evaluation phase. An ECG signal model based on authentication was provided by Arteaga-Falconi et al. [13] A deep learning model was paired with a one-class SVM to evaluate ECG data. In that study, the wavelets transform was used to convert an ECG signal into a picture, and the resultant image was input into a previously trained CNN model. Six separate databases were used throughout the evaluation process. An EER of 4.9 percent was achieved using a 1-minute enrolling duration and 4 seconds of authentication time. A novel technique to ECG biometrics was introduced by Wang et al. [14] A band pass Butterworth filter with cut-off frequencies spanning from 1 to 40 Hz was used to enhance ECG data signal quality. In addition, a multi-scale differential feature was combined with one-dimensional multi-resolution local binary patterns to extract components from ECGs. Four datasets were used to evaluate the suggested model in that research. Although the approaches listed above function well, the ECG properties have not been thoroughly researched. This work provides a novel ECG validation approach based on multi-domain features and the Least Square Support Vector Machine (LS-SVM). ECG signal segments are first produced. Then, two characteristics are recovered and merged to construct the final feature vector: time domain and frequency features based on an efficient three-band filter. To find the best feature set, the extracted feature vector is evaluated. Finally, the extracted features are processed using LS-SVM.

2. ECG Data

2.1. The physio net organization provided the ECG dataset. The information may be found at https://physionet.org/content/mitdb/1.0.0/. The data was gathered from 47 healthy people. The performance assessment in this research employed all ECG records [15]. for further information, for actual settings, we evaluated using a single ECG lead to gather ECG data. 1000 Hz was the sampling rate. A total of 54 healthy people took part in the ECG signal recording. For MIT-BIH Arrhythmia 1 Involving 48 ECG Records There are two recordings (201 and 202) for the same subject in 48 records. There are 47 people in this database, each recording is 15 seconds long. The participants' ages ranged from 25 men aged 32 to 89 years and 22 women aged 23 to 89 years. We used the entire rule.

3. Proposed methodology

This research proposed a new model for individual identification based on ECG features. A low-pass Butterworth filter with a cut-off frequency of 128 Hz was first used to filter ECG signals. Frequency domain features are extracted from each segment, and the ECG segment is separated into six bands using an Optimized Triple Half Filter Bank (OTHFB). To extract frequency features from ECG segments, two features are extract from each segment, called Higuchi Fractal (HFD) and Hjorth parameters (HP), and by splitting the range into two segments (HD). The frequency feature set is checked and verified for the strongest features. The LS-SVM is used to classify ECG features. (Figure: 1) According to the data, LS-SVM has a high performance rating

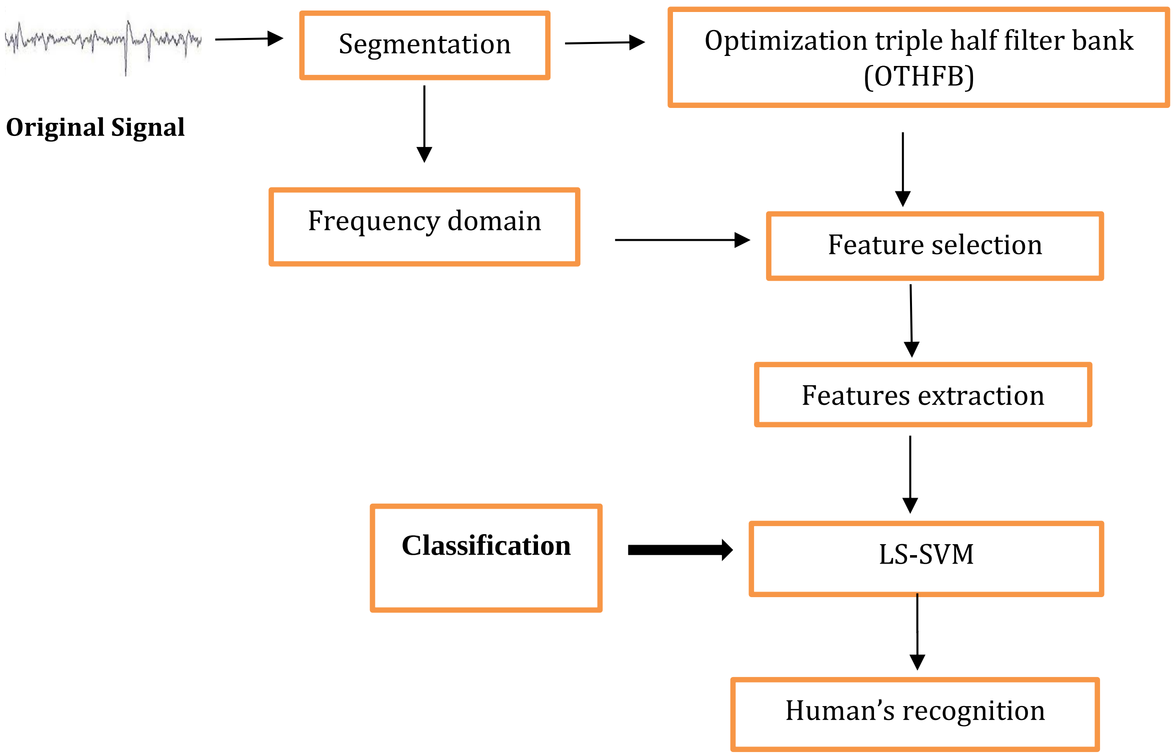

Fig. 1 – proposed system based on frequency domain

4. ECG Segment

Each ECG signal was divided into m intervals in this study. To segment the ECG signals, we used fiducial points as a reference. The start and end of the heartbeat are determined by the credit points LP and TP. The R peak must be found before the LP and TP can be calculated. Figure 2 of an ECG segment shows the LP and TP positions. We used a differentiation detection method from [6] in this investigation. This method is advantageous since it is rapid, does need a minimum, and has a low mean time error. For example, consider the following ECG signal, which has m data points: X = x 1, x 2, x 3, and so on, up to x m. Using LP and TP, the signal X was divided into n windows, each containing k data points, and stored in a vector as X = n 1, n 2, n 3...., n m for the next step. As a result, each pulse is divided across the ECG signal by the LP and TP credit points.

5. feature extraction based on frequency domain

5.1. Optimization triple half filter bank (OTHFB)

To evaluate ECG data, a triplet half-band filter bank is used. The THFB is used to extract frequency information from each segment [11]. The THFB stands for

H0(z)G0(z)+ H0(-z) G0(-Z)=2 (1)

H0(z) and G0(z) are the analysis and synthesis low pass filters' transfer functions, respectively (LPF). These filters were made using the THFB structure. The low pass filter (LPF) H0(z) and high pass filter (HPF) H1(z) for the THFB analysis are computed as follows:

$$H0(z) = \frac{1+P}{2} + \frac{1}{2} L1\ (Z)\big(1 - pL0(z)\big) \qquad (2)$$

L0(z); L1(z); and L2(z) are the THFB kernel functions produced from the half band polynomial Pi(z) of order K, Pi(z) = 1 + Li(z) for I = 0; 1; 2. These half-band filters and kernels are non-causal filters that meet the PR requirement. The shape parameters of p0, p1, and p2 are computed as follows: p0 = p1 =((1+P))/2 and p2=((1-p))/((1+p)) where p's range is between 0 and 1

The current effort in the design of THFB does not address the optimization of temporal frequency localization. The time-frequency relationship was optimized. The Euler-Fresenius polynomials are used in THFB (EFP). To maintain symmetry, the EFP coefficients are computed recursively. Property. The EFP is described as follows:

$$\boldsymbol{E(z)} = \sum_{\boldsymbol{k=0}\ \boldsymbol{e(k+1)z}}^{\boldsymbol{m}} \boldsymbol{-k} \qquad (\boldsymbol{3})$$

where M stands for polynomial order and may be calculated as follows:

$$e\ (k\ +\ 1)\ = \sum_{k=0}^{n} (-1) \binom{m+2}{m} (k+1-m)^{m+1} \qquad (4)$$

The EFP polynomial is altered by adding N vanishing moments (VMs) and one degree of freedom to get the smoothness of the scaling function i.

$$p(z) = \ (1+z)^n = E(z) \sum_{k=0}^{l} \alpha k^{z-k} \qquad (5)$$

The order of P(z) is K = M + N + L, and L = K=2, 1. The time-frequency half-band filter that has been optimized is defined as

Q(z) = P(z) $\sum_k^1 \beta k R k(z)$ (6)

where _k is the degree of freedom to be evaluated. R k is an interpolator with a polynomial specification. In this situation, we utilized k = L=2.The OTHFB half-band filters L0(z); L1(z); L2(z) are adjusted using the balanced-uncertainty (BU) measure to balance in time and frequency localization. BU measurements are optimized to yield the value of L/2.

$\beta=\Delta\omega^2+k^2\Delta t^2$ (7)

where Δt^2 and $\Delta\omega^2$ are the time and frequency localization parameters, respectively.

$$\Delta t^2 = \frac{1}{\sum q(n)} \sum_{k=0}^{n} n^2 q^2\ (n) \qquad (8)$$

$$\Delta\omega^2 = \frac{\pi 2}{3} + 4 \sum_{n=0}^{k-1} \sum_{m=n+1}^{k-1} \frac{(-1)^{(m-n)}(a)}{U(m-n)} q(m)q(n)^n \qquad (9)$$

where q(n) is the impulse response as calculated by equation (7). We got the following by inserting (9) and (10) in (8):

$$\beta=\frac{\pi}{2}+\frac{\sum n\sum\frac{m(-1)^{m-n}}{(m-n)^2 q(m)q(n)}+k^2\sum n\ n^2q^2(n)}{\sum_n^m q^2(n)} \tag{10}$$

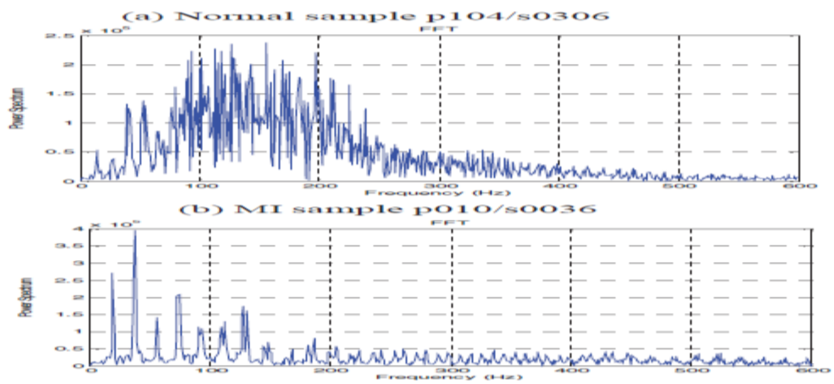

Fig. 2 – ECG signals in the frequency domain from for different type of sample classification

The LS-SVM classifiers are the least support vector machine (LS-SVM) classifiers. were employed to categorize the ECG characteristics in this investigation. It's a supervised learning algorithm-based classifier

6. performance evaluation measured

In order to determine the usefulness of a particular system, it is necessary to examine its performance using unique metrics. This research used a confusion matrix to calculate several performance evaluation criteria, such as accuracy, sensitivity, specificity, and false positive rate (FPR). The confusion matrix is also known as the error matrix in machine learning[16]. is a two-dimensional (actual*predicted) matrix that may be used to assess a classifier's performance; each row of the matrix contains instances in a predicted class, while each column represents occurrences in an actual class [17]. Correct classifications are referred to as 'true positives (TP)' or 'true negatives (TN),' whereas faulty classifications are referred to as 'false positives (FP)' or 'false negatives (FN).' Accuracy, sensitivity, specificity, and a variety of other parameters may all be calculated. This is especially significant when diagnosing the presence or absence of a specific disease or illness with a single test [18]. Figures 2 show the confusion matrix for two and several classes, respectively.

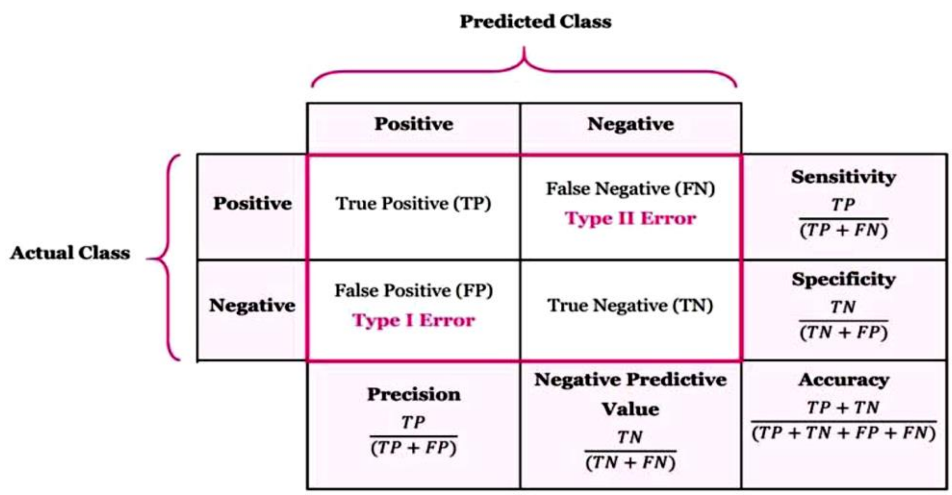

Fig. 3 – confusion Matrix [1].

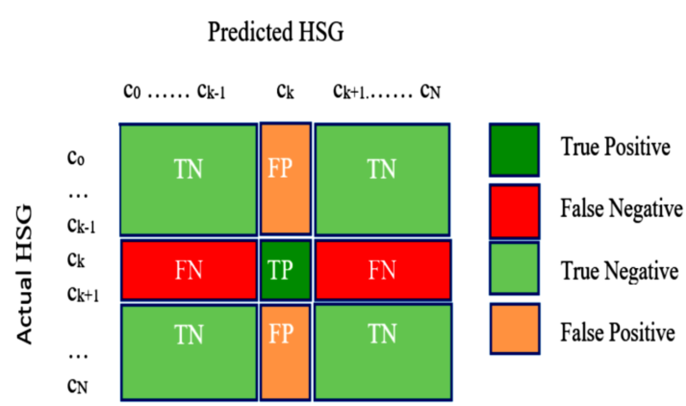

Fig. 4 –confusion matrix ,show with totals for positive and negative tuples [2]

To calculate measures that used in this work, the following equations used[19].

- $Accuracy\ \frac{TP+TN}{TP+FN+TN+FP}$ (11)

- $Sensitivity, recall, hit\ rate, or\ true\ positive\ rate\ (TPR)$

$$Sensitivity = \frac{TP}{TP + FN} \quad (12)$$

- $Specificity, selectivity\ or\ true\ negative\ rate\ (TNR)$

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

- $False\ Positive\ Rate\ (FPR)\ or\ fall - out$

$$FP = \frac{FP}{FP + TN} \tag{14}$$

6. Experimental results

The performance of the suggested model is examined on many scales in this section. The extracted features were put to the test to find the optimum feature sets for frequency domain extraction. The research aims to identify a set of feature extractors that can be used to validate ECG biometrics and evaluate the classification algorithm's effectiveness. An electrocardiogram (ECG) is a test that measures the heart's electrical activity (cut off). We gathered ten outcomes from each subject's ECG recording to create a validation model. The database determines the number of EKG beats during the verification process. In our testing, we employed the LSSVM technique to extract features and rate the proposed model's performance in terms of accuracy, sensitivity, and specificity. Many aspects of the ECG signal were investigated in the frequency domain.

- **Detection of individuals based on frequency domain**

In this experiment, ECG slides were conducted using OTHFB. Consequently, each segment of the ECG was divided into six bands. Fractal dimensional characteristics have been determined for each round. There was an average of eight features in each ECG segment. The Higuchi fractal dimensions (HFD) and the development of the Hjorth parameters were among the attributes achieved. The acquired features were classified using LSSVM. The OTHFB-based qualities found are listed in Table 1.

Table 1

Detection performance based on frequency domain characteristics

Classification method	Accuracy	Sensitivity	Specificity
LSSVM	95%	92%	92%

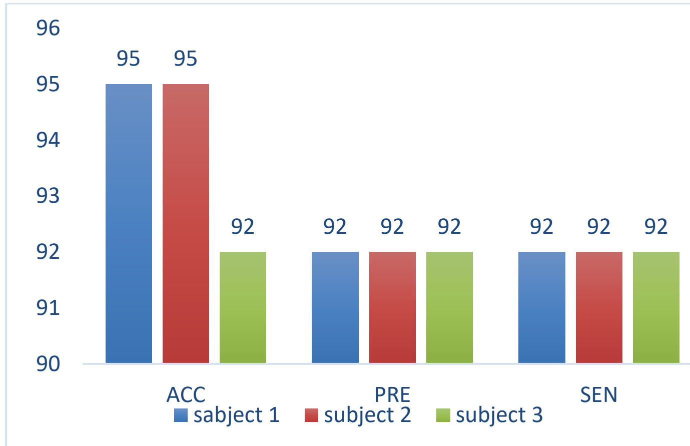

Fig. 5 – Describe accuracy based on frequency domain

7. Comparisons with previous work

Table 2 shows the comparison of the proposed model with previous methods based on ECG in the literature. To our knowledge, previous approaches used high-performance ECG features compared to our study that rated the ECG recording with fewer features. In addition, our study used utilization of frequency domain features to design a high-performance model to identify individuals[20]. However, our method scored 95% using fewer ECG features including frequency features. Another study by Hanilçi et al., [21].in which handcrafted features with CNN were used. Their method recorded an accuracy of 88% which is considered insufficient for any identification system. On the other hand, our approach has gained high performance compared to Hanilçi et al., [21].Agrafioti et al., [22]. proposed a KNN with a Butterworth filter to analyse the ECG signals. They scored 92.3% while our approach scored 95%. Based on the comparisons, the main advantage of this study is the highly desirable accuracy of 95% considering all healthy subjects

Table 2

Comparisons among the proposed model and previous approaches

Author	Approach	Accuracy
Agrafioti et al., [34]	Butterworth bandpass filter based on KNN	90%
Agrafioti et al., [35]	Butterworth bandpass filter coupled with KNN, k-means	92.3%
Lourenço et al., [36]	Chaos theory combined with SVM	81.73%
Hanilçi et al., [37]	Conventional neural networks with handcrafted features	88%
Proposed model	Time and frequency features with LS-SVM	95%

8. Conclusion

Several challenges, such as the link between feature selection and detection rate, remain unanswered despite the significant work establishing an ECG-based biometric system. This study has carefully examined the influence of feature type and feature extraction on the ECG biometric system. Several experiments were run to see how effective the suggested model was in extracting ECG parameters, and frequency features were used in this research. Random ECG data sets were used in the validation system investigations. Combining frequency characteristics boosted the detection rate, according to the data. To make the proposed model more robust, further research is required to evaluate it using a large data set. Additionally, in ECG-based biometric systems, additional classifiers should be investigated.

References

[[1] W. Wu, S. Pirbhulal, A. K. Sangaiah, S. C. Mukhopadhyay, and G. Li, "Optimization of signal quality over comfortability of textile electrodes for ECG monitoring in fog computing based medical applications," *Futur. Gener. Comput. Syst.*, vol. 86, pp. 515–526, 2018.

[2] M. Hammad and K. Wang, "Parallel score fusion of ECG and fingerprint for human authentication based on convolution neural network," *Comput. Secur.*, vol. 81, pp. 107–122, 2019.

[3] T. Bai *et al.*, "A lightweight method of data encryption in BANs using electrocardiogram signal," *Futur. Gener. Comput. Syst.*, vol. 92, pp. 800–811, 2019.

[4] P. Peris-Lopez, L. González-Manzano, C. Camara, and J. M. de Fuentes, "Effect of attacker characterization in ECG-based continuous authentication mechanisms for Internet of Things," *Futur. Gener. Comput. Syst.*, vol. 81, pp. 67–77, 2018.

[5] S. Hamza and Y. Ben Ayed, "Svm for human identification using the ecg signal," *Procedia Comput. Sci.*, vol. 176, pp. 430–439, 2020.

[6] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Comput. Biol. Med.*, vol. 122, p. 103801, 2020.

[7] S. Hamza and Y. Ben Ayed, "SVM for human identification using the ECG signal," *Procedia Comput. Sci.*, vol. 176, pp. 430–439, 2020, doi: 10.1016/j.procs.2020.08.044.

[8] N. Rabinezhadsadatmahaleh and T. Khatibi, "A novel noise-robust stacked ensemble of deep and conventional machine learning classifiers (NRSE-DCML) for human biometric identification from electrocardiogram signals," *Informatics Med. Unlocked*, vol. 21, p. 100469, 2020, doi: 10.1016/j.imu.2020.100469.

[9] W. Klonowski, "Fractal analysis of electroencephalographic time series (EEG Signals)," in *The fractal geometry of the brain*, Springer, 2016, pp. 413–429.

[10] H. Namazi and S. Jafari, "Age-based variations of fractal structure of EEG signal in patients with epilepsy," *Fractals*, vol. 26, no. 04, p. 1850051, 2018.

[11] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.

[12] E. Al Alkeem *et al.*, "Robust Deep Identification using ECG and Multimodal Biometrics for Industrial Internet of Things," *Ad Hoc Networks*, vol. 121, no. May, p. 102581, 2021, doi: 10.1016/j.adhoc.2021.102581.

[13] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, "ECG and fingerprint bimodal authentication," *Sustain. Cities Soc.*, vol. 40, no. August 2017, pp. 274–283, 2018, doi: 10.1016/j.scs.2017.12.023.

[14] K. Wang, G. Yang, Y. Huang, and Y. Yin, "Multi-scale differential feature for ECG biometrics with collective matrix factorization," *Pattern Recognit.*, vol. 102, p. 107211, 2020, doi: 10.1016/j.patcog.2020.107211.

[15] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[16] S. V Stehman, “Selecting and interpreting measures of thematic classification accuracy,” *Remote Sens. Environ.*, vol. 62, no. 1, pp. 77–89, 1997.

[17] D. M. W. Powers, “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,” *arXiv Prepr. arXiv2010.16061*, 2020.

[18] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.

[19] H. Brooks *et al.*, “WWRP/WGNE Joint Working Group on Forecast Verification Research,” *Collab. Aust. Weather Clim. Res. World Meteorol. Organ.*, 2015.

[20] S. K. D. Alkhafaji, “Evaluation of the influence of directivity factor of directive elements of conformal and planar antenna arrays on the performances of azimuth-elevation DOA estimation,” in *Journal of Physics: Conference Series*, 2019, vol. 1279, no. 1, p. 12024.

[21] Y. B. Nechaev, K. D. A. Sarmad, and I. W. Peshkov, “Evaluating expectation-maximization algorithm for 2D DOA estimation via planar antenna arrays,” in *Proceedings of the International Conference on High Performance Compilation, Computing and Communications*, 2017, pp. 126–130.

[22] M. Diykh, S. Abdulla, K. Saleh, and R. C. Deo, “Fractal dimension undirected correlation graph-based support vector machine model for identification of focal and non-focal electroencephalography signals,” *Biomed. Signal Process. Control*, vol. 54, p. 101611, 2019.