International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7642
***Article Received:** 02 June 2023 **Revised:** 28 July 2023 **Accepted:** 14 August 2023*

Securing the Biometric through ECG using Machine Learning Techniques

Mr. Praveen Kumar Gupta[1], Dr. Subhash Singh Parihar[2], Dr. Ritesh Rastogi[3], Dr. Vidushi[4], Ms. Komal Salgotra[5]
[1]Assistant Professor KIET Group of Institutions, Ghaziabad
[2]Associate Professor, Pranveer Singh Institutions of Technology, Kanpur
[3]Associate Professor, Noida Institutions of Technology, Greater Noida
[4]Assistant Professor, CHRIST (Deemed to be University) Delhi-NCR
[5]Teaching Assistant, KIET Group of Institutions, Ghaziabad

Abstract: In the current era, biometrics is widely used for maintaining the security. To extract the information from the biomedical signals, biomedical signal processing is needed. One of the significant tools used for the diagnostic is electrocardiogram (ECG). The main reason behind this is the certain uniqueness in the ECG signals of the individual. In this paper, the focus will be on distinguishing the individual on the basis of ECG signals using feature extraction approaches and the machine learning algorithms. Other than preprocessing approach, the discrete cosine transform is applied to perform the extraction. The classification between the signals of the individuals is carried out using the Support Vector Machine and K-Nearest Neighbor machine learning techniques. The classification accuracy achieved through SVM is 87% and K-NN has achieved a classification accuracy of 96.6% with k=3. The work has shown how machine learning can be used to classify the ECG signal.

Keywords: auto correlation, discrete cosine transform, normalization, K-NN, SVM

I. Introduction

Verification or identification biometric information systems can be either implemented on an existing system or built from scratch[1]. The verification method strives to validate an individual's claimed identity by checking their claims against several markers of individuality. In relation to this matter, it is important to note that an identification system has the capability to ascertain the identity of an individual (among the individuals stored in an expert system) without necessitating the individual to explicitly provide their identity. The verification system employs a one-to-one search methodology, wherein a single item is searched to identify a corresponding match. Conversely, the identification system utilizes a single to Multi search approach, wherein multiple items are searched to identify potential matches. The utilization of biosignature technology has had a significant influence in a wide range of applications, facilitating enhanced security measures and enabling its involvement in diverse domains such as bank check verification, author identification, operational banking, face recognition, medicinal findings, turnout tracking, authorized file authentication, and security trials [2].

Although several biometric systems, a signature verification method remains among the most demanding and lucrative behavioural biometrics [3]. The term signature is based on signing off, which originates from the Latin root, signature, and sign. Whether it is someone's writing or someone else's, a signed specimen is used to identify a person. The signature verification system is a method of verifying the authenticity of a signer before any existing samples are used. It is one of the most desirable biometric verification systems due to its vast number of favourable characteristics such as convenience, social acceptability, and lack of legal or societal problems [4].

The recent introduction of big data and artificial intelligence technology is revolutionizing the healthcare system significantly, which could lead to significant economic change for the industry[5]. This research article is geared toward digital health security that takes advantage of machine learning for biometric data. Biometric authentication is on the rise and is becoming a more popular option for access control systems, which is why it's becoming the main control method. The utilization of ECG data for individual identification is a relatively recent development in the field of biometrics [6]. It involves the analysis of electrical conduction through the body's heat to obtain this data. Concurrently, ECG-based security systems will face hurdles in DL and ML classification methodologies like deep learning[7]. Artificial intelligence subsumes machine learning, and so computers may complete tasks without explicit instructions. Machine learning methodologies have the potential to be employed in the creation of a verification framework for authentication purposes, utilizing real-time electrocardiogram (ECG) data. The utilization of modelling techniques encompasses various applications such as prediction, classification, and pattern recognition. Machine learning has various applications, such

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7642
***Article Received:** 02 June 2023 **Revised:** 28 July 2023 **Accepted:** 14 August 2023*

as the analysis of video, image, sound data, and ECG data [8].

Everyday attributes such as Fingered impressions, looks, hand indicators, veins, speeches, iris diaphragm, and other physical as well as behavioral characteristics remain employed to identify individuals. The usefulness of a biometric signature is shown when we look at another potential signature, such as biomedical signals. These are produced as an electrical signal when a nerve cell, muscle, or gland cell undergoes electrochemical changes, which surface electrodes may pick up in touch with the skin. Some biometric methods, like the electrocardiogram (ECG), obtained from the heart, the Electromyogram (EMG), capturing muscle activity, the Electrodermal Activity (EDA), evaluating changes in skin conductivity, and the Electroencephalogram (EEG), determining neural activity, use biological signals as signs for figuring out whether or not a person's physiological state are underneath an optimal range.

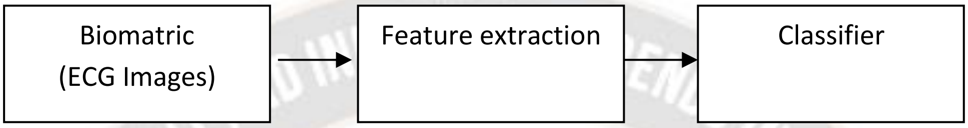

Figure 1. Biometric classification

In this work we propose the following outlines: 1st Section discusses the outline, while 2nd gives literature review of the survey. 3rd provides proposed methodology whereas 4th covers the implementation.

II. Literature Survey

Belo et.al describes that Biometrics is the study of pattern recognition[9]; it focuses on unique characteristics of a person that are recorded, registered, and compared to other records in a database. Because the effectiveness of replicating Electrocardiograms (ECG) has been a challenge, their employment has recently been increasing in the biometric industry for applications that require higher levels of security. The purpose of this effort is to improve current DNN and ECG performance by designing and implementing two new architectures.

Zahra Ebrahimi et al.[10] in their work reviews all of the latest ECG-related DL approaches for classification applications. These DL approaches include RNN ,CNN, DBN, LSTM, and Gated Recurrent Unit (GRU). According to the 75 research that were conducted in 2017 and 2018, CNN remains the maximum frequently used removal technique, through 52% of the studies finding it appropriate. GRU/LSTM, CNN, and LSTM all achieved good accuracy in the categorization of AF, SVEB, and VEB, respectively.

Kim et al.[11] intended for those interested in applying machine learning for the digital health with the use of biometric data. Forgetfulness, loss, and theft are common issues with older authentication schemes. The ease and comfort of using biometrics has greatly improved, and they have become part of everyday life. An ECG-based biometric system suited for entry verification has been adopted as a tool for building verification, and this research aims to better understand, and Using RR-interval segments of an ECG for biometric authentication. In this research, we present the OP, an innovative composite performance indicator.

Alkeem et al.[12] proposed that, Using ECG facts aimed at human documentation popular in industrial internet of things (IIoT) which may reach near-perfect accuracy when used under optimum settings. ECG signals, however, are affected by noise and interferences, which means that they are difficult to process. An upgraded biometric identification strategy could benefit from incorporating different biometric data. Using multimodal biometrics, comprising fingerprint, ECG, and facial image data, we have proposed a novel method that exclusively uses gender classification. an integrated approach Training requirements and inter-domain correlation are both reduced, allowing for improved model generalization on multiple tasks. The advantages of multitasking include help from a previously botched task in order to help correct and regulate the rest of your tasks, thereby enhancing your overall performance. A proposed solution embraces multimodality by employing a technique that combines feature-level and score-level fusion. Combining multiple multimodality approaches, multitasking, and multimodal fusion techniques, the provided ideas appear to be ground-breaking. Feature-level fusion offers an advantage over other fusion methods, however the extension to the benchmark dataset, recommended by the model, is more successful. The proposed model has undergone testing and verification using noisy and partial facts, and its validation has been based on the analysis of investigational outcomes.

Kaplan et al. [13] shows that, a useful analogy for an ECG reading is that it is an electrical tracing of the heart's electrical activity. Many other uses have been found for the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7642
***Article Received:** 02 June 2023 **Revised:** 28 July 2023 **Accepted:** 14 August 2023*

ECG signal, including monitoring heart rate, evaluating cardiac rhythms, detecting heart problems, and emotion recognition. It is critical to perform each step in order to obtain associated analyses. Also, indicators of employability and the soundness of the ECG signal database contribute to the analysis. This paper fully reviews ECG analysis, particularly since the recent decade, built on completely main points that have been described above. In the detailed explanation of each stage in ECG analysis, we also present relevant research studies that apply the techniques.

III. Proposed Methodology

In this section, flowchart (Fig. 1) is shown in which the data is preprocessed to perform the ECG Signal analysis of the three individual.

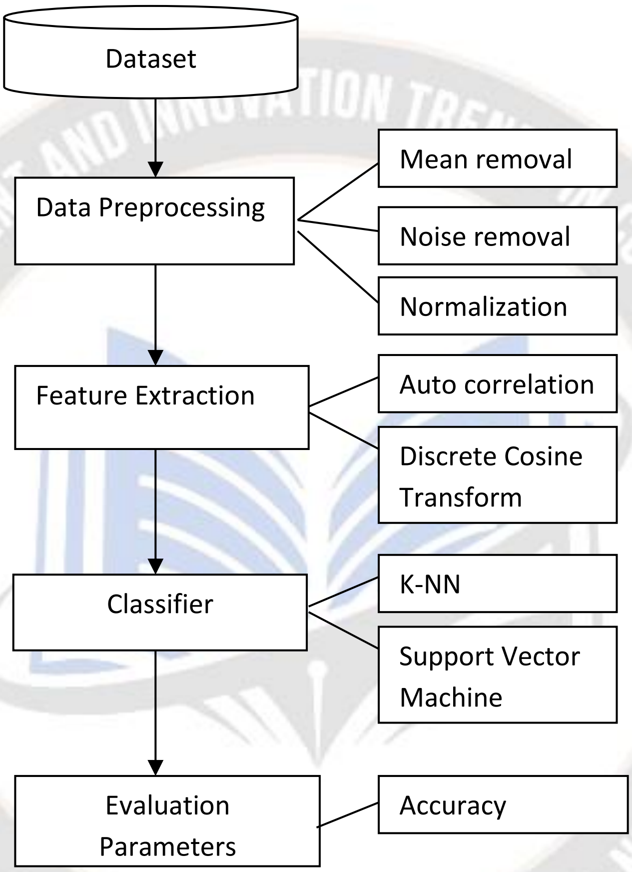

Figure 2. Flowchart of the proposed work

3.1 Datasets: The dataset contain the values of the ECG signals of the three individual i.e. S_1, S_2 and S_3 that consist of positive as well as negative values.

3.2 Data Preprocessing

The preprocessing of the data is one of the important stages before performing the classification. In this work the preprocessing of the data is done by applying mean removal method and band pass filters are used to remove the noise from the signal. Normalization is also performed the partition the ECG signal into segments.

3.3 Feature Extraction

In the preprocessed data, some of the data are still not used for the analysis purpose. So this type of data is removed by using the feature extraction approaches[14]. The auto correlation function and the discrete cosine transform will be used to extract the important features which can be used for the classification purpose.

a. Auto correlation - The statistical technique known as autocorrelation is useful in detecting the temporal link between variables[15]. There exist three distinct methodologies for analyzing time series data: employing a time series plot, utilizing a lagged scatter plot, and employing an autocorrelation function. In order to assess the presence of long-term serial correlation in ECG data, the autocorrelation function will be utilized. This section will provide a deeper understanding of autocorrelation coefficient functions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7642
***Article Received:** 02 June 2023 **Revised:** 28 July 2023 **Accepted:** 14 August 2023*

Let xi (i = 1, n) represent a time-series dataset at data point I, and let the average value of the dataset be denoted as the data's average value.

The autocorrelation function at lag k is expressed as:

$$rk = \frac{\sum_{i=1}^{m-k}(xi - a)(xi + k - a)}{\sum_{i=1}^{m}(xi - a)^2} \quad (1)$$

To show that RK stands for the correlation between two data sets separated by a lag of k, we can look at an autocorrelation function. When it's a perfect negative relationship, the score goes from a 1.0 to a -1.0. (perfect positive relation). If the relationship between the two variables is broken, the score is zero.

Following that, the technique for classifying the two symptoms is described:

1) Auto-correlate ECG data of half a day time and describe the notch using lag k as outcome of correlation.

2) Using two parameters, examine the primary periodic slope segment.

b. Discrete Cosine Transform - A useful and customizable alternative for a data compression technique is the DCT (discrete cosine transform)[16]. Dynamic threshold allocation and variable sub-band coding aid in expanding the DCT's overall applicability in ECG data compression applications. In addition, the discrete cosine transform method can be modified to accommodate low-resolution, low-sampling-frequency circumstances, which are typified by low-resolution videos and higher-resolution videos. A fixed transform matrix is used, which does not respond to the input signal, and a real-time implementation is possible due to the static nature of the transform matrix. The issue of computational feasibility, which can be provided in real time, is typically more significant than the lack of adaptability. The DCT can be obtained by rearranging the elements of the input vectors to construct a Fast Fourier Transform (FFT) algorithm [17].

3.4 Classification Algorithms – In this work, two algorithms are implemented for the classification purpose. First classifier is the SVM with its two different kernels i.e. Linear and the RBF (Radial Basis Function)[18]. Next machine learning classifier is the K-Nearest Neighbor algorithm with different values of k.

3.5 KNN in ECG- K-Nearest Neighbour (KNN) is a search technique that looks for the nearest item or point in a dataset and compares the value of that point with values in a database table, and inferences are generated based on the comparison results. When a person has more than one condition, the KNN locates the most noticeable one.

3.5 SVM in ECG- In supervised machine learning, SVM is a classification method which can be used for regression problems, and a regression method which can be used for classification problems. The technique used is called the kernel trick, in which your data is first transformed and then the results are searched for an appropriate border. This data processing machine has the ability to perform sophisticated data transformations, and then learns how to divide your data into categories or outputs, based on your labels or outputs.

IV. Implementation and Results

The planned effort is realized using the Python programming language. All the files related to data preprocessing, feature extraction and classification are implemented separately. There are three files S_1, S_2 and S_3, of the values of the ECG based on individuals. The performance parameters taken for evaluating the model is accuracy. Given below table 1 has shown the comparison of the classifiers SVM and K-NN based on accuracy.

Table 1. Accuracy comparison of SVM and K-NN

Files	K-NN		SVM	
	K=3	K=5	RBF	Linear
S_1	72.1%	68%	75.1%	46.2%
S_2	16.9%	28.6%	17.5%	22%
S_3	96.6%	80.1%	85.9%	82.3%

Table 1 has shown that K-NN with k=3 having accuracy of 96.6%, improved than the other models.

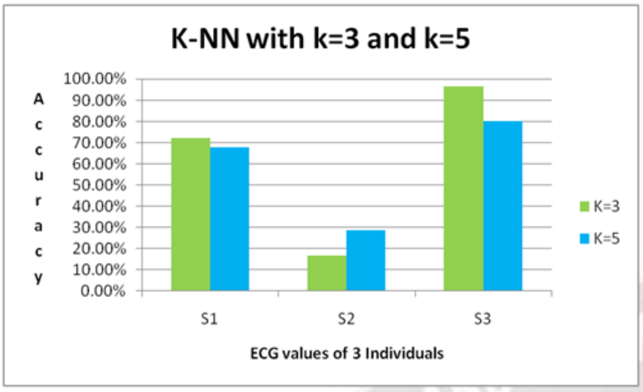

Figure 3. Accuracy comparison graph with k=3, k=5 (K-NN)

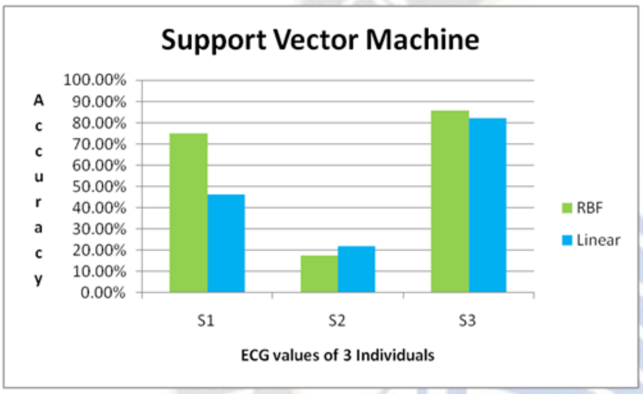

Figure 4. Accuracy comparison graph of Support vector Machine with Linear kernel and RBF kernel

Figure 3-4 has shown the comparison graph of the K-NN and the SVM respectively.

V. Conclusion

In this work, the biometric signal in ECG is analyzed through the machine learning algorithms SVM and K-NN. The work will ensure that the alteration of the health record, which is in the form of ECG, should not be altered. To perform this, accurate classification of ECG should be done; otherwise, it is very difficult to recognize the correct and the fake ECG. The implemented model K-NN with k=3 has achieved an accuracy of 96.6 %, which can be considered a good classifier for the prediction purpose. In future work, deep learning algorithms with more patients will be proposed and implemented.

References

[1] D. Kim, D. Choi, and J. Jin, "Method for detecting core malware sites related to biomedical information systems," Comput. Math. Methods Med., vol. 2015, 2015.

[2] B. J. Rauscher, E. R. Canavan, S. H. Moseley, J. E. Sadleir, and T. Stevenson, "Detectors and cooling technology for direct spectroscopic biosignature characterization," J. Astron. Telesc. Instruments, Syst., vol. 2, no. 4, 2016.

[3] I. Stylios, S. Kokolakis, O. Thanou, and S. Chatzis, "Behavioral biometrics & continuous user authentication on mobile devices: A survey," Inf. Fusion, vol. 66, 2021.

[4] A. N. Azmi, D. Nasien, and F. S. Omar, "Biometric signature verification system based on freeman chain code and k-nearest neighbor," Multimed. Tools Appl., vol. 76, no. 14, 2017.

[5] R. Nair and A. Bhagat, "A Life Cycle on Processing Large Dataset - LCPL Rajit Nair," vol. 179, no. 53, pp. 27–34, 2018.

[6] R. N. Vargas and A. C. P. Veiga, "Electrocardiogram signal denoising by clustering and soft thresholding," IET Signal Process., vol. 12, no. 9, 2018.

[7] R. Nair and A. Bhagat, "Genes expression classification using improved deep learning method," Int. J. Emerg. Technol., 2019.

[8] T. Wang, C. Lu, and G. Shen, "Detection of Sleep Apnea from Single-Lead ECG Signal Using a Time Window Artificial Neural Network," Biomed Res. Int., 2019.

[9] D. Belo, N. Bento, H. Silva, A. Fred, and H. Gamboa, "Ecg biometrics using deep learning and relative score threshold classification," Sensors (Switzerland), vol. 20, no. 15, 2020.

[10] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," Expert Systems with Applications: X, vol. 7. 2020.

[11] S. K. Kim, C. Y. Yeun, E. Damiani, and N. W. Lo, "A machine learning framework for biometric authentication using electrocardiogram," IEEE Access, vol. 7, 2019.

[12] E. Al Alkeem et al., "Robust Deep Identification using ECG and Multimodal Biometrics for Industrial Internet of Things," Ad Hoc Networks, vol. 121, 2021.

[13] S. Kaplan Berkaya, A. K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," Biomedical Signal Processing and Control, vol. 43. 2018.

[14] R. Nair and A. Bhagat, "Feature selection method to improve the accuracy of classification algorithm," Int. J. Innov. Technol. Explor. Eng., 2019.

[15] M. S. Bin Sinal and E. Kamioka, "An Efficient Arrhythmia Detection Using Autocorrelation and Statistical Approach," J. Comput. Commun., vol. 06, no. 10, pp. 63–81, 2018.

[16] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Recent advances and emerging challenges of feature selection in the context of big data," Knowledge-Based Syst., 2015.

[17] M. Szymkowski, P. Jasiński, and K. Saeed, "Iris-based human identity recognition with machine learning methods and discrete fast Fourier transform," Innov. Syst. Softw. Eng., 2021.

[18] I. B. Aydilek, "Examining Effects of the Support Vector Machines Kernel Types on Biomedical Data

Classification," in 2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018, 2019.