ELECTROCARDIOGRAM-BASED BIOMETRIC AUTHENTICATION: A MACHINE LEARNING APPROACH

[1] *ZOHRA JABEEN,* [2] *KURA RAMANI*
[1]*Assistant Professor,*[2]*Student*
Department of CSE
Sree Chaitanya College of Engineering, Karimnagar

Abstract : In order to build electrocardiogram (ECG)-based biometric authentication systems, machine learning (ML) approaches must be suitably adjusted and modified. This research presents a framework for doing so. It may assist in defining the parameters of the necessary datasets and obtaining high-quality training data. The bounds of datasets are established via use case analysis. Three distinct use cases on ECG-based authentication are established based on different application contexts. Accordingly, the accuracy of ML-based ECG biometric authentication methods increases with the provision of more quality training data to related machine learning schemes. This framework uses the R-peak anchoring ECG time slicing approach to get high-quality ML training data. Four additional measure indicators are included in the suggested framework to assess the calibre of the ML training and testing data. Moreover, a MATLAB toolbox is created that includes all suggested mechanisms, measurements, and example data along with examples employing different ML approaches. The suggested framework may guide researchers in creating the appropriate ML settings, ML training datasets, and three user case scenarios in order to build ML-based ECG biometric authentication.

Index Terms: Biometric Authentication, R-peak Anchoring, Machine Learning, ECG Time Slicing, and Authentication.

I. INTRODUCTION

Users are increasingly accessing application systems by identifying themselves with their own bodies as most of them allow regular users to access the Internet. As a result, in recent years, biometric authentication has gained popularity as a study issue. Electrocardiogram authentication provides the benefit of using real user body signals during authentication, unlike other biometric authentication techniques like fingerprint scanning and face recognition. Typically, real-time ECG data is obtained from users in order to build a verification model for person identification using machine learning methods. A number of state-of-the-art publications on ECG-based biometrics have been published recently. Further research is still needed on a number of ECG biometrics-related issues, including the classification of authentication, pre-processing for improving data quality, data collecting, and selection using Deep Learning (DL) and other Machine Learning classification techniques.

In order to address the issues with ECG authentication that have been observed, this research presents a machine learning framework. Use cases must be used to determine the fundamental application scenarios in order to have a better understanding of possible application contexts for ECG authentication. Application scenarios using ECG authentication are divided into three broad use cases in the proposed framework: Hospital (HOS), Security Check (SCK), and Wearable Devices (WD). Additionally, novel methods for preprocessing data are suggested, such as the baseline adjustment of ECG frequency artefacts, the ECG data noise reduction approach for Power Line Interference (PLI), and the flipping mechanism

for the ECG signal caused by incorrect electrode placement. Furthermore, the system incorporates temporal slicing approaches to generate machine learning-based training datasets and novel measure metrics designed for authentication accuracy assessment. In the suggested framework, four new measure indicators for data quality are presented. The metrics in question include Accuracy Percentage within Ranges (APR), Accuracy per UCL (APU), Mean Absolute Error Rate (MAER), and Upper/Lower Range Control Limits (UCL/LCL).

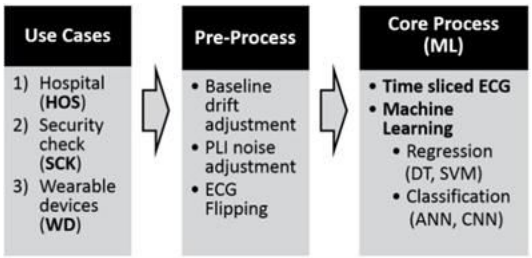


Fig: Overview of the New Framework Model for ECG based Biometric Authentication

An overview of the new framework model for machine learning-based ECG biometric authentication is shown in Figure. A number of machine learning methods are used in the core process section, including convolutional neural networks (CNN) and artificial neural networks (ANN) for classification and decision trees (DT) and support vector machines (SVM) for regression. Furthermore, a time-slicing approach for ECG data is created and linked to the main procedure.

II. LITERATURE SURVEY

Heart-ID: A multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications

In the new age of smart cities, body area networks—which include smart sensors—are significantly changing health applications. A lot of emphasis is being paid to physiological signal based biometric human identification in order to fulfil the ever-increasing security and privacy requirements. This work focusses on two main obstacles: feature engineering takes a lot of effort and fits only certain datasets, and signal processing techniques are often complex and data-dependent. A unique wavelet domain multiresolution convolutional neural network is suggested to provide a highly generalisable and data-independent signal processing and feature learning procedure. In particular, it eliminates the need for the laborious process of extracting signal fiducial properties by enabling the blind selection of a physiological signal segment for identification purposes. The randomly selected signal segment is then converted to the wavelet domain, where multiresolution time-frequency representation is accomplished, in order to improve the data representation. The phase difference resulting from the blind segmentation procedure is eliminated from the converted data by using an auto-correlation technique. The person identification job is then carried out by a multiresolution 1-D-convolutional neural network (1-D-CNN), which is designed to automatically learn the inherent hierarchical features from the wavelet domain raw data without the need for data-dependent or labour-intensive feature engineering. Using eight distinct ECG datasets with varying behaviors—such as those with or without serious cardiac problems and with varying sensor implantation techniques—the efficacy of the suggested approach is comprehensively assessed. Our assessment has an average identification rate of 93.5%, and is much more comprehensive than the state-of-the-art efforts. The suggested multiresolution 1-D-CNN system may successfully detect human individuals without the need for extensive feature engineering, even from randomly chosen signal segments. It is anticipated that this work will show that blind signal processing and deep learning approaches may be used to biometric person identification with strong generalisation ability and little algorithm engineering effort.

Evolution, current challenges, and future possibilities in ECG biometrics

The most studied biometric features at the moment are the fingerprint and face, which provide trustworthy identification in a variety of applications. From cellphones to border control, commercial goods that use these characteristics for biometric identification or verification are becoming more and more common. However, the demand for qualities that are less susceptible to spoofing assaults or covert trait measurement is driven by the sophistication of methods used to counterfeit such features. This has raised interest in the electrocardiogram (ECG), which is most often used for medical diagnostics and is very resistant to assaults due to its concealed nature and ingrained liveness information. The field of ECG-based biometrics has rapidly moved in the last several years towards commercial uses, primarily in response to concerns about decreased comfort and acceptance via the development of novel wearable, off-the-person, and seamless acquisition settings. In addition, studies on the avoidance of spoofing and data security in ECG biometrics as well as the potential for deep learning techniques to improve identification robustness and accuracy have only been begun. The 93 state-of-the-art articles on their suggested methodologies, signal datasets, and publicly accessible ECG collections are thoroughly reviewed and discussed in this work. The retrieved information is used to outline the foundations and development of ECG biometrics, characterise the state-of-the-art, and make inferences about previous methods and existing issues. In addition to exploring the available prospects, our goal with this study is to stimulate and direct future research in ECG biometrics.

Learning deep off-the-person heart biometrics representations

The electrocardiogram (ECG) has been researched as a biometric characteristic for security systems and other applications from the year 2000. The collecting of ECG signals in the off-the-person category using devices like smartphones and tablets has recently made this biometric signal appropriate for practical situations. In this study, we describe the application of deep learning methods—more especially, convolutional networks—to extract meaningful representations for the identification of cardiac biometrics. In particular, we study feature representation learning for heart biometrics using two sources: the heartbeat spectrogram and the raw pulse signal. Additionally, we provide ways for augmentation of heartbeat data, which are crucial for generalisation within the framework of deep learning methodologies. We demonstrate that our approach delivers state-of-the-art outcomes in the two off-the-person publically accessible datasets using the identical experimental setup for six techniques in the literature.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

In our exist system is explain the biometric authentication system generally authentication is checks if your correct user or not so based on data system will check your credentials and give access so in this scenario every person will give their credentials after that he will get access some companies have more than 500 its check on by one so it will take so much time.

DISADVANTAGES:

- More time required
- More resources required
- Cost effective

3.2 PROPOSED SYSTEM

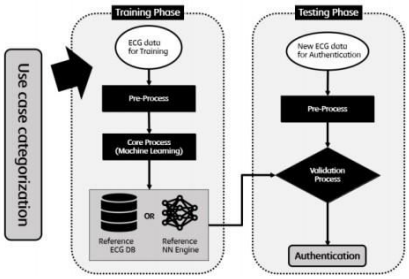
Our proposed System introduces a framework for how to appropriately adapt and adjust machine learning (ML) techniques used to construct electrocardiogram (ECG)-based biometric authentication schemes. The proposed framework can help investigators and developers on ECG-based biometric authentication mechanisms define the boundaries of required

datasets and get training data with good quality. To determine the boundaries of datasets, use case analysis is adopted. Based on various application scenarios on ECG-based authentication, three distinct use cases (or authentication categories) are developed.

ADVANTAGES:

- Easy to maintain
- Fast tracking
- Less effective

IV. SYSTEM ARCHITECTURE



V. SYSTEM IMPLEMENTATION MODULES

5.1 PRE-PROCESS FOR ECG DATA QUALITY ENHANCEMENT

The pre-process is about adjusting data before starting the core process (i.e., machine learning process). The signal processing techniques have been widely applied into adjusting ECG data since ECG data could be considered as signals. Many applications in the signal processing including the filter designs and the Fourier Transforms are applied for enhancing the ECG recognition. Although many pre-processes are adopted for enhancing signals, three processes are recommended for enhancing ECG data before starting the machine learning process.

5.2 TIME SLICING AND MACHINE LEARNING

The time slicing technique is considered for building up the dataset of the machine learning training. This approach is especially applicable for building up the machine learning training data. The ECG data are sliced based on a slice (window) time (typically known as a sliding window) with the R-peak anchoring. This method could generate enough data samples and each sliced data becomes a sample input for the machine learning training. The time sliced ECG dataset is very flexible not only to mix with other training inputs but also to apply various ML training methods.

5.3 DATA QUALITY MEASURES

Evaluating your machine learning algorithm is an essential part of any ML projects and delivering the good quality of samples is vital for evaluating the ML system. Some performance measures (evaluation metrics) are the values of evaluation results of sample data qualities and regression approaches based authentication systems.

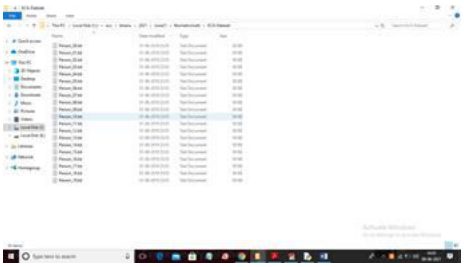
5.4 MATLAB TOOLBOX

The proposed Toolbox for demonstrating the proposed processes and techniques in each section is actually implemented as the functions of the Matlab. The Amgecg Toolbox (Amang ECG Toolbox) which is the set of the Matlab functions which researchers could use for their own ECG authentication projects and this section introduces some of functions in the Toolbox.

VI. SCREEN SHOTS

In this paper author is using machine learning algorithm to authenticate user via Electrocardiogram (ECG) dataset and this authentication can be used to authenticate users in Hospitals, Security Check and Wearable devices. In hospitals then can collect ECG data from patients and assign ID to each ECG record and then train this ID and ECG with machine learning algorithms such as SVM, decision tree, ANN and CNN algorithms. After training model then this trained model can be used to authenticate patient by giving ECG input.

To implement this project we have used ECG dataset from 20 persons and below screen shots showing all 20 person ECG records and each record contains 5000 ECG readings,



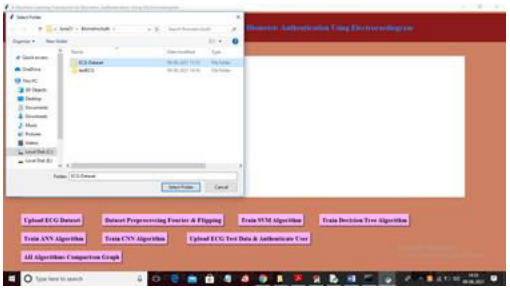
In above screen open any file to see that person ECG data and to implement this project we are developing following modules

1) Upload ECG Dataset: using this module we will upload dataset to application
2) Dataset Preprocessing Fourier & Flipping: using this module we will preprocess dataset to remove missing values and then apply flipping and Fourier transform algorithm to select important attributes from dataset
3) Train SVM Algorithm: using this module we will train SVM with ECG dataset and then calculate accuracy, mean squared error and mean absolute error on test data
4) Train Decision Tree Algorithm: using this module we will train Decision Tree with ECG dataset and then calculate accuracy, mean squared error and mean absolute error on test data
5) Train ANN Algorithm: using this module we will train ANN with ECG dataset and then calculate accuracy, mean squared error and mean absolute error on test data
6) Train CNN Algorithm: using this module we will train CNN with ECG dataset and then calculate accuracy, mean squared error and mean absolute error on test data
7) Upload ECG Test Data & Authenticate User: using this module we will upload test ECG data and then machine learning algorithm will identify user id from that test ECG data
8) All Algorithms Comparison Graph: using this module plot accuracy comparison graph between all algorithms

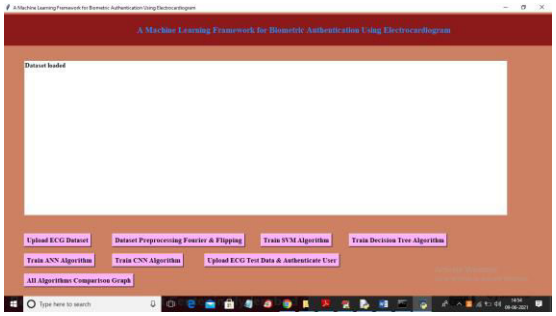
SCREEN SHOTS

To run project double click on 'run.bat' file to get below screen

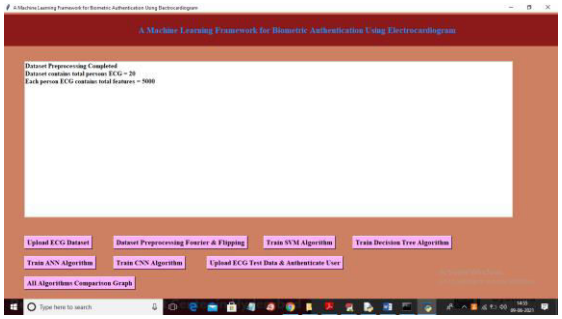
In above screen click on 'Upload ECG Dataset' button to upload dataset and to get below screen



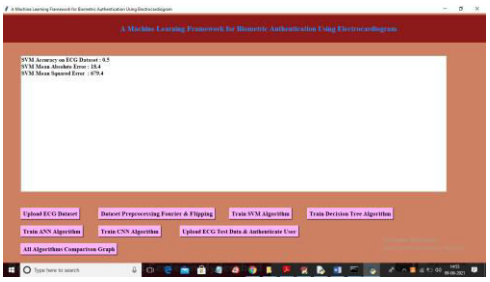
In above screen selecting and uploading 'ECG-Dataset' folder and then click on 'Select Folder' button to load dataset and to get below screen



In above screen dataset loaded and now click on 'Dataset Preprocessing Fourier & Flipping' button to read and clean dataset



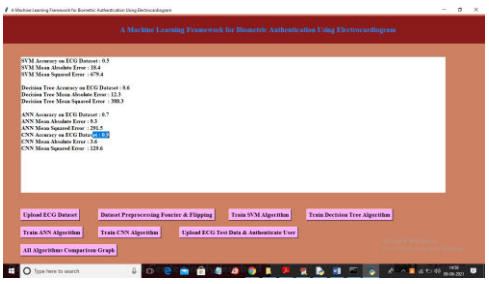
In above screen dataset contains 20 person ECG records and each ECG record contains 5000 values and now click on 'Train SVM Algorithm' button to train SVM with above dataset



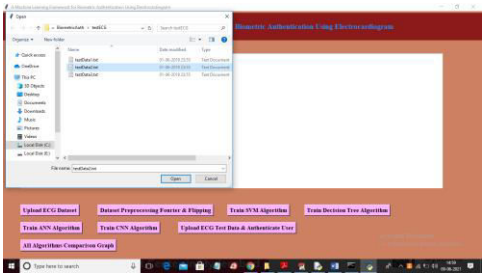
In above screen SVM is trained and we got accuracy as 0.50% and now click on 'Train Decision Tree Algorithm' button to get decision tree model

In above screen with decision tree we got 0.60% accuracy and now click on 'Train ANN Algorithm' button to train ANN with above dataset

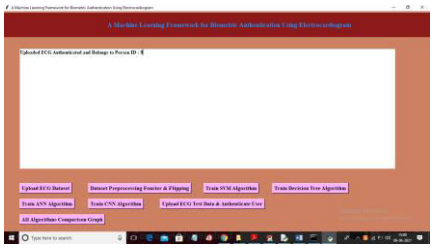
In above screen with ANN we got accuracy as 0.70% and we got MSE and MAE for each algorithm and now click on 'Train CNN Algorithm' button to train SVM with above dataset



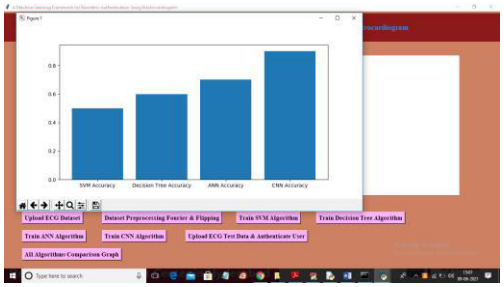
In above screen with CNN we got 0.90% accuracy and now click on 'Upload ECG Test Data & Authenticate User' button to upload test ECG and then application identify person ID from test data



In above screen selecting and uploading 'testData2.txt' file and then click on 'Open' button to get below authentication result



In above screen person ID is Authenticated or Identified as 'Person ID 9' and similarly you can upload any other test data and authenticate user. Now click on 'All Algorithms Comparison Graph' button to get accuracy graph of all algorithm



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from all algorithms CNN is giving better result.

VII. CONCLUSION

ECG based biometric authentication will be used on large application systems worldwide in the near future as new ECG detection devices become lightweight, portable, embeddable with smartphones and wearables, and wired to distant servers. ML approaches are often used to create

a more reliable assessment model for ECG-based biometric authentication in order to achieve high accuracy on user authentication. This study presents a generalised machine learning framework for biometric authentication based on electrocardiograms. To make it easier for researchers to create and assess an ML-based ECG user authentication scheme, a suggested framework explains the overall data processing flow of an ML-based ECG authentication mechanism together with a number of function characteristics. Those features include four new data quality measures, three new general authentication categories for ECG user authentication, three new data pre-processing methods, a temporal slicing method to produce high-quality ECG datasets, and a publically accessible Matlab Toolbox (also known as amgecg Toolbox). Several data pre-processing techniques and newly defined measure metrics offered by the proposed framework are still useful for people using ML technologies to study other topics instead of ECG based biometric authentication, and they can help researchers expedite the development of their ML-based schemes.

BIBLIOGRAPHY

1. Q. Zhang, D. Zhou and X. Zeng, "HeartID: A Multiresolution Convolution Neural Network for ECG-Based Biometrics Human Identification in Smart Health Applications", IEEE Access, vol. 5, pp. 11805-11816, 2017.
2. J.R. Pinto, J.S. Cardoso and A. Lourenco, "Evolution, Current Challenges, and Future Possibilities in ECG Biometrics", IEEE Access, vol. 6, pp. 34746-34776, 2018.
3. E.J.S. Luz, G.J.P. Moreira, L.S. Oliveira, W.R. Schwartz and D. Menotti, "Learning Deep Off-the-Person Heart Biometrics Representations", IEEE Transaction on Information Forensics and Security, vol. 13, no. 5, pp. 1258-1270, 2018.
4. H. Kim and S.Y. Chun, "Cancelable ECG Biometrics Using Compressive Sensing-Generalized Likelihood Ratio Test", IEEE Access, vol. 7, pp. 9232-9242, 2019
5. Y. Xin, L. Kong, Z. Liu, Y.Chen, Y. Li, H. Zhu, M. Cao, H. Hou and C. Wang, "Machine Learning and Deep Learning Methods for Cybersecurity", IEEE Access, vol. 6, pp. 35365-35381, 2018.
6. H. J. Kim and J. S. Lim, "Study on a Biometric Authentication Model based on ECG using a Fuzzy Neural Network", 2018 IOP Conf. Ser.: Mater. Sci. Eng. 317, 10 pages, 2018.
7. J. R. Pinto, J. S. Cardoso and et al., "Towards a Continuous Biometric System Based on ECG Signals Acquired on the Steering Wheel", Sensors 2017, vol. 17 no. 10, 14 pages, 2017.
8. M. Sansone, R. Fusco and et al., "Electrocardiogram Pattern Recognition and Analysis Based on Artificial Neural Networks and Support Vector Machines: A Review", Journal of Healthcare Engineering, vol. 4, no. 4, pp. 465-504, 2013.