ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

THE STUDY AI-POWERED BIOMETRIC IDENITIFICATION SYSTEMS AUTHENTICATION MECHANISM USING ELECTROCARDIOGRAMS INFORMATION

Arekatla Madhava Reddy, Lankala Mounika, Dr. G. Samba Siva Rao, Dr. S K Moulali

[1,2] Assistant Professor, [3,4] Professor

amreddy2008@gmail.com, lankala.mounikareddy@gmail.com

profgssrao@gmail.com, moulalishaik1275@gmail.com

Department of CSE, A.M. Reddy Memorial College of Engineering and Technology, Petlurivaripalem,

Narasaraopet, Andhra Pradesh -522601

ABSTRACT

The foundation for developing relevant Machine Learning (ML) approaches to build Electrocardiogram (ECG) based biometric authentication systems is presented in this study. The suggested framework may assist researchers and developers working on ECG-based biometric authentication methods in defining the parameters of necessary datasets and obtaining high-quality training data. Use case analysis is used to establish dataset bounds. Three separate use cases, or authentication categories, are established based on different application situations using ECG based authentication. Increasing the amount of qualified training data provided to machine learning schemes that correlate with them would raise the accuracy of machine learning-based ECG biometric identification methods. This framework uses the ECG time slicing approach with the R-peak anchoring to get high-quality ML training data. Four additional measure indicators are included in the suggested framework to assess the caliber of ML training and testing data. Additionally, a Matlab toolbox is created and made accessible to the public for additional research. It includes all suggested mechanisms, measurements, and example data with demos utilizing different ML algorithms. 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. In order to generate high-quality ML-based training and testing datasets and to leverage new measure metrics, the suggested framework remains valuable for researchers who are embracing ML approaches to create novel schemes in various study fields.

Index Terms: Neural network, regression, MATLAB, machine learning, statistical learning, identification, biomedical signal processing, electrocardiogram (ECG), 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. Recent years have seen the publication of many state-of-the-art studies on ECG-based biometrics [1-4]. 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 for Deep Learning (DL) and other Machine Learning classification techniques [5]. To address issues with ECG authentication that have been found, this paper presents a machine learning methodology. 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. The three main use cases for ECG authentication application scenarios in the proposed framework are Hospital (HOS), Security Check (SCK), and Wearable Devices (WD). Moreover, novel methods for pre-processing data are suggested, such as the baseline adjustment of ECG frequency artifacts, 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 techniques to generate machine learning-based training datasets and creates new measures for evaluating authentication accuracy. 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). An overview of the new framework model for machine learning-based ECG biometric authentication is shown in Figure 1. 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.

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

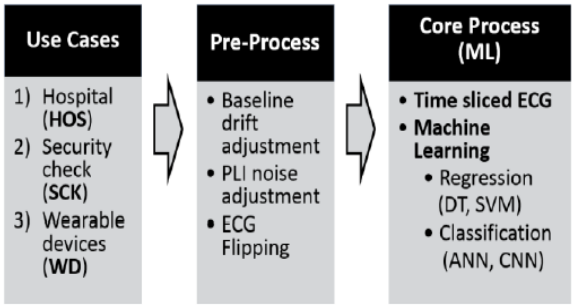

FIGURE 1. Overview of the New Framework Model for ECG based Biometric Authentication

The framework processing flow for ECG biometric authentication using machine learning approaches is shown in Figure 2. The whole suggested architecture begins with the selection of an appropriate user category for the intended system environment. Following this, the training phase begins with the acquisition of relevant ECG data from the intended users for use as the training dataset. Pre-processing procedures are used to the training data once they are received in order to provide filtered data that are of better quality (less noise). One of the chosen core process mechanisms shown in Figure 1 receives the filtered data as input, which creates an authentication assessment model. For any ML-based ECG authentication system to utilize them, the suggested core process presently supports both the trained Neural Network (NN) reference engine and the ECG reference database. After the training phase is over, the reference database, also known as the NN Engine, is formed. The freshly obtained ECG data is used to produce an ECG-based user authentication request during the testing phase. The ECG data must first apply data pre-processing methods to get filtered data with greater quality (less noise). After that, the filtered data are delivered to the validation process, which uses them as input to confirm the outcome of this user authentication request with the reference database or the NN Engine.

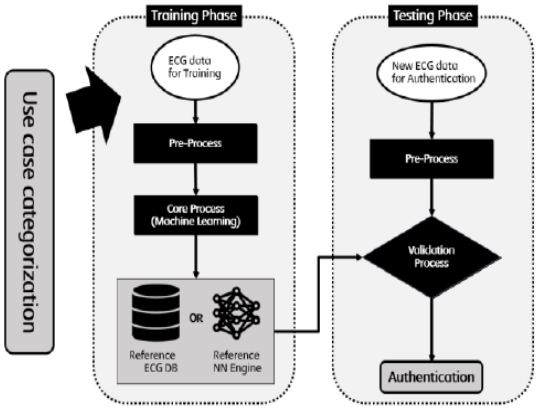

FIGURE 2. The Framework Processing Flow for

II. AUTHENTICATION CATEGORIZATION BASED ON USE CASES

Numerous studies have been conducted on biometric authentication utilizing ECG data [6–11]. Despite the fact that there is a wealth of literature on ECG-based authentication, all of them employ distinct ECG detection equipment and user contexts when using ECG-based authentication techniques. An electrocardiograph is often set up by researchers with experience in medical engineering to obtain ECG data [6–8]. On the other hand, researchers with experience in electrical engineering often put up simple ECG sensors—which are typically integrated into wearable technology—to collect ECG data [9–11]. Thus, before creating an ECG-based biometric authentication method, it is crucial to take into account and comprehend the user environment and the kind of ECG detection equipment that may be in use. A documented explanation of how users will engage with the intended system is called a use case. Use case analysis may help clarify important information for system processes and identify system needs throughout the design phase [12].

Use case analysis might be used to classify potential use cases. Three authentication categories are identified by applying the use case analysis technique to potential application scenarios for ECG-based user authentication: hospital patients, individuals undergoing identity verification at building entrances, and continuous authentication for personal use (also known as HOS, SCK, and WD, as illustrated in Figure 1). The following addresses the related user environment and assumptions for each category. Keep in mind that each category has a particular system performance requirement in terms of

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

authentication speed and accuracy rate, which varies depending on the target application systems.

A. HOSPITAL PATIENT AUTHENTICATION (HOS)

Typically, an ECG test is performed on the patient to determine if cardiac disease or a heart stress has occurred. In order to collect high-quality ECG signals for medical diagnostics, patient ECG equipment is often intricate and complex. As a result, depending on the kind of ECG test, the sampling period for obtaining ECG data may range from a few minutes to many hours, and numerous leads are utilized throughout an ECG test. Identification of hospital patients is a novel use case for ECG testing (Category 1; HOS use case). It is assumed that such patients must pre-register both their history ECG data and their identities (i.e., names or legal identification numbers). Furthermore, for both the registration (training) and verification (testing) stages of an ECG-based authentication method, it is expected that the recorded ECG signals from the same patient are sufficiently stable (that is, that the measured ECG signal values fall within a normal range). The next time the patients attend the hospital, the hospital may use an ECG-based biometric authentication technique to identify those individuals. Observe that, in contrast to asking a nurse for the patient's name and legal identity number, which could take several minutes, a well-trained ECG user authentication model (or scheme) can identify a patient by evaluating live ECG signals in less than a few seconds [13]. ECG-based user (or patient) authentication makes it simple to identify patients who are unconscious in an emergency department. Generally speaking, one of the main application contexts for ECG-based authentication techniques may be patient authentication in hospitals. The most extensively used research environment in the healthcare and medical industries is this HOS use case [14]. Numerous databases, like the PhysioBank database, are accessible to the general public and include historical ECG data [15].

B. PERSONAL IDENTIFICATION AT A BUILDING ENTRANCE (SCK)

If required, the security check (Category 2; SCK use case) for building and room admission is implemented using the second user authentication use case, which is based on user ECG data. The majority of businesses use security checkpoints to verify the identities of workers and guests. ECG-based biometric authentication systems will replace fingerprint scanning, facial recognition, voice identification, iris recognition, and retinal scanning as user authentication options for security checkpoints. This is due to the availability of portable ECG detection devices or ECG detection sensors. An ECG-based authentication system may be utilized in this SCK use case to distinguish between registered regular workers and unidentified individuals (including guests). It is assumed that legal staff members have already registered their names or legal identification numbers to this ECG authentication system, together with their history ECG data. Furthermore, it is assumed that during the registration and verification stages, the measured ECG signals from the same employee are sufficiently steady.

C. ONGOING PERSONAL USE AUTHENTICATION (WD)

The third use case for ECG-based user authentication is personal wearables (like smart watches) that are equipped with ECG sensors to continually check whether the person using them is indeed the owner (Category 3; WD use case). Generally speaking, all a wearable gadget needs to do is constantly verify that its owner is the wearer. A new framing technique may be needed to normalize the received ECG signals by filtering out potential signal noise caused by the user's body status, as the heart beat period (R-R peak period) and the amplitude of the signals can change dramatically when the person is under different body statuses, such as walking, running, and sleeping. A wearable gadget may also provide its user with second factor authentication capabilities if it has an integrated ECG sensor and an ECG authentication module. A wearable device user with WD may increase security control by using a two factor authentication mechanism (e.g., authentication using both user password and user ECG data).

Table I: Boundaries of the dataset classified by authentication use cases

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

Cat No.	Cat. Name	Known ID classification	Unknown ID	Personal Status
1	Hospital (HOS)	O	X	X
2	Security Check (SCK)	O	O	X
3	Wearable Devices (WD)	X*	O	O

** There is only one known ID in the WD case*

Depending on the use case classifications, researchers could take other aspects into account. For example, a security checkpoint's ECG sample time ought to be less than a hospital's patient's ECG sample time. Because the goals for using the gathered ECG data are different in the WD situation, the sampling frequency for ECG signals should be substantially lower than the ECG sampling frequency created on ECG signals by typical medical measuring equipment. Additionally, because different kinds of ECG equipment are used in the HOS instance and the WD case, human operating mistakes such lead misplacement will not be taken into account in the WD category.

III. PRE-PROCESS FOR ECG DATA QUALITY ENHANCEMENT

Preparing the data for the core process (i.e., machine learning process) is the goal of the pre-process. Since ECG data may be seen as signals, signal processing methods have been frequently used to alter the data. To improve ECG detection, a variety of signal processing techniques are used, such as filter designs [16–18] and Fourier transforms [19–20]. Three procedures are advised for improving ECG data before beginning the machine learning process, despite the fact that several pre-processes are used for signal enhancement.

A. ADAPTATION OF BASELINE

An ECG low frequency artifact caused by breathing, electrically charged electrodes is known as the baseline drift (or baseline wander) [21]. The method of baseline correction eliminates baseline drift. Generally, the high-pass filter's cut-off frequency must be set higher than the signal's lowest frequency in order to completely remove baseline drift. A common feature of most baseline wander reduction methods is their cancellation of the signal's low frequency components. Typically, the baseline wander high-pass filter's frequency is adjusted to be little around 0.5 Hz [22]. The right frequency for the baseline drift reduction should be decided in advance, despite the fact that these methods are extensively researched and used [23]. Beyond low frequency noise, an applicant's movements may potentially cause baseline drift during the collection of ECG data. Thus, these filtering methods could not be helpful if we don't know the right cut-off frequency indication or if we anticipate a certain applicant movement (HOS example).

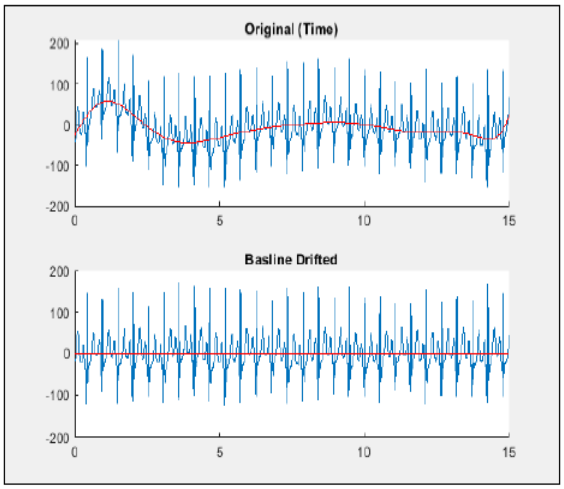

FIGURE 3. Baseline adjustment by using the polynomial curve fitting [15]

This baseline correction method combined with polynomial curve fitting is often helpful when low frequency noise is causing issues beyond merely baseline wander. Furthermore, the baseline is automatically adjusted to zero (0) to prevent any movement of the ECG amplitude. Atypical causes of baseline wander include an applicant's specific movements during the collection of ECG data, apart from low frequency noise.

B. REMOVAL OF POWER LINE INTERFERENCE NOISE

Baseline wander is one of the many different kinds of noise signals, and the noise from the PLI connected to signal carrying cables is especially problematic. cables transporting signals from the examination room to the monitoring equipment are susceptible to electromagnetic interference (EMI) of frequency by continuous supply lines, which often occurs in

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

medical equipment in hospitals (HOS case) [26]. A typical source of noise in the ECG is electromagnetic fields from power lines, which are characterized by 50 or 60 Hz sinusoidal interference, perhaps with many harmonics. Due to the introduction of spurious and unreliable waveforms, such narrow band noise complicates the interpretation of low amplitude waves [27]. To eliminate PLI noise, the Infinite Impulse Response (IIR) notch filter is often used [28]. A notch filter allows signals above and below the stop band frequency range to flow through while rejecting or attenuating frequencies in that particular band [29]. These kinds of filters might be used to eliminate baseline wander, but first it's important to choose an appropriate target frequency. On the other hand, comparable results may be achieved by using a notch filter to eliminate aberrant peak points in the frequency domain without first identifying the target frequency. With PLI's high peaks in the frequency domain, Figure 4's PLI noise might be eliminated using this method.

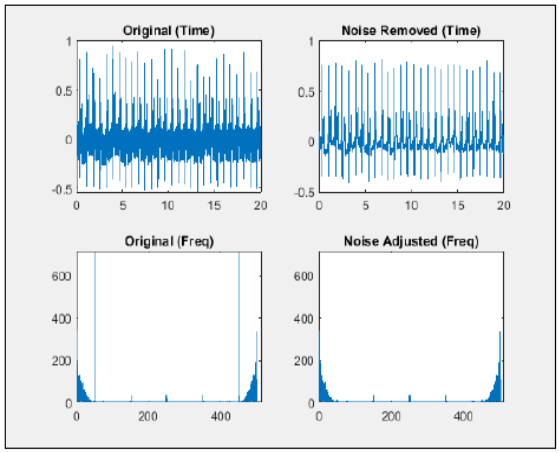

FIGURE 4. Noise adjustment for removing PLI frequency [30]

IV. TIME SLICING AND MACHINE LEARNING

When constructing the dataset for the machine learning training, the temporal slicing approach is taken into consideration. This method works particularly well for accumulating training data for machine learning. The R-peak anchoring is used to slice the ECG data according to a slice (window) time, sometimes called a sliding window. This technique might provide a sufficient number of data samples, and each slice of data serves as an input sample for machine learning training. The time-sliced ECG dataset is very adaptable; it may be used with other ML training techniques, as detailed in subsection B. A., as well as combined with additional training inputs. ECG DATA SLICED TIME-WISE FOR MACHINE LEARNING

The center and most noticeable portion of the tracing is often the QRS complex, which is the conjunction of three graphical deflections shown on a standard ECG [32]. the maximum of a QRS complex, or an R-peak. It represents a single heartbeat, and the R-peak moment is often used to optimize the sliding window duration and serve as the anchor of the QRS complex, along with R-peak detection [33]. To cut the ECG signal from an R-peak moment to the sliding window period and layer these parts depending on the R-peak moment (i.e., R-peak anchoring), time slicing, which is essentially slicing ECG data in the time domain, is the goal. Sample inputs for the machine learning training are generated from each slice based on the R-peak anchoring shown in Figure 6. In this research, a slicing time (i.e., sliding window) of 0.6 seconds, or 100 bps pulse rate, is selected as the average minimum of a heartbeat interval from an unusual heart rate [34]. The goal of ECG-based projects determines the optimal sliced window period, and certain machine learning performance metrics are partially dependent on the ECG slicing time. While improving window time for biometrics is an intriguing ECG-based security research issue, this optimization challenge is not being studied at this time.

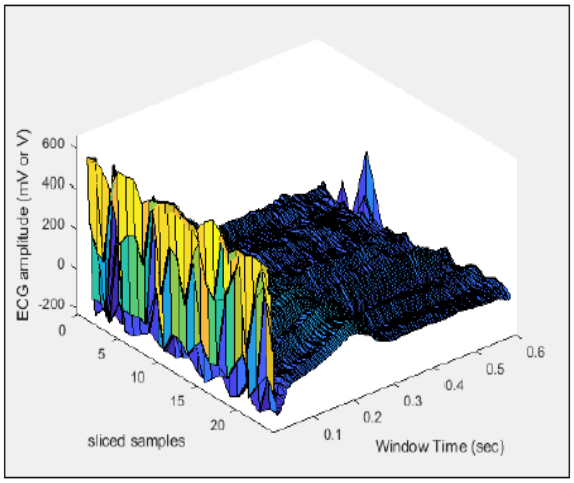

FIGURE 6. ECG Time Slicing with the R-peak anchoring [15]

V. DATA QUALITY MEASURES

Any ML project must include the assessment of machine learning algorithms [51], and providing high-quality samples is crucial for the evaluation of ML methods. Regression-based authentication

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

systems and the values of assessment findings of sample data characteristics are examples of performance measurements (evaluation metrics) [52]. The following are typical performance and data quality metrics for a regression approach:

- Sum of Squares Error
- Sum of Squares Total
- Mean Squared Error
- Mean Absolute Error

The study presents new measuring criteria in addition to standard quality metrics. These quality metrics are used for both the criterion of validating samples and the assessment of regression-assisted machine learning systems. The following are the new metrics for data quality:

• Range Control Limits, Upper and Lower; • Mean Absolute Error Rate

• Percentage of accuracy within ranges

• Precision according to Upper Control Limit

The next subsections provide further information on each of the aforementioned new data quality metrics.

A. The MAER, or mean absolute error rate

When assessing mistakes in a machine learning model, the Mean Square Error (MSE) and the root MSE (RMSE) are helpful metrics to utilize. An average squared difference between the two estimated datasets is known as the MSE [53]. Comparably, the RMSE is the MSE squared, and both metrics are often used as metrics for evaluating machine learning. Even though the majority of regression-based machine learning model assessments include both (MSE and RMSE) metrics, it may be difficult to tell by just looking at them whether a model is excellent or terrible. This study introduces a new metric to assess machine learning engines or training datasets. The following is the definition of the Mean Absolute Error Rate (MAER):

$$MAER = \frac{1}{N}\sum_{n=1}^{N}\frac{|Y_n - \mu_n|}{\mu_n + \epsilon}, \epsilon \sim 0 \qquad (1)$$

B. UPPER/LOWER RANGE CONTROL LIMITS (UCL, LCL)

The application of statistical methods to regulate a process production method in quality engineering is known as statistical process control, or SPC for short. SPC and statistical quality control are often used interchangeably [54]. The two primary uses of SPC in quality control are control charts and acceptance sampling [55]. There are two kinds of control charts: X-charts, which show the data mean, and R-charts, which show the data range, also known as a variance or standard deviation. On a control chart, control limits, also known as the Upper Control Limit and Lower Control Limit, are horizontal lines that are often drawn at a distance of ±3 or ±6 standard deviations from the data mean of the plotted statistic. It was possible to construct the R-charts both with and without the reference values. The data quality is shown in both charts, and the control values may be used to exclude outliers before training the data. The range values R's control limits (UCL and LCL) might be found in the manner described below:

$$UCL = \bar{R}_Y + \sigma(b)\bar{R}_Y, \qquad (2)$$

$$LCL = Max\left(0, \bar{R}_Y - \sigma(b)\bar{R}_Y\right), \qquad (3)$$

C. ACCURACY PERCENTAGE WITHIN RANGES (APR)

The percentage of the ECG data that falls into the ranges between 0 and UCL is known as the Accuracy Percentage within Ranges (APR). Out of the total number of sliced ECG data samples, it is the counting of numbers inside the ranges.

$$APR = \frac{n(\{R \le UCL\})}{n(\Omega_R)} \qquad (4)$$

VI. MATLAB TOOLBOX

The primary procedure and essential elements for implementing machine learning methods in ECG-based biometric authentication systems are provided in the preceding parts I–IV. The functions of Matlab are really used to create the suggested Toolbox, which serves as a demonstration of the suggested procedures and methods in each section. This section covers some of the Matlab functions included in the Amgecg Toolbox, also known as the Amang ECG

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 3, 2021

Toolbox, which researchers may utilize for their own ECG authentication studies.

A. Pre-processing and time-slicing tools

As the pre-process of the Machine Learning adaptations, three new processes have been implemented, namely the baseline drift adjustment in Figure 3, the noise adjustments in Figure 4, and the flipping of ECG data in Figure 5. The Matlab functions in the new Toolbox are as follows:

- *baselinedrift*
- *enhancednoiseadjust*
- *pseudoflip*

The way of using each function could be found by using "help" function of the Matlab. Users can earn the detailed information for how to use each function via the Matlab command window as shown in Figure 8.

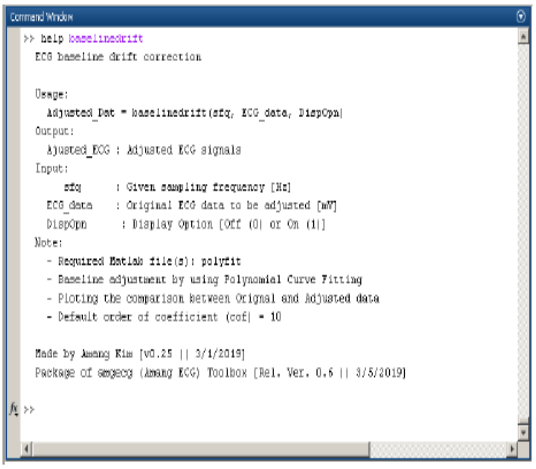

FIGURE 8. Using 'help' function in the Matlab command window

B. TOOLS FOR DATA QUALITY

The functions for evaluating the quality of the data are also included in the amgecg Toolbox. The toolbox contains the data quality indicators that were discussed in Section V, together with the following associated functions:

- *rangecontrol*
- *maer*
- *mseamg*
- *maerdataqualityengine*
- *msedataqualityengine*

A few fundamental and integrated functions are combined to provide the data quality functions. The "help" function in Matlab provides information about each function. Furthermore, the demo file may be accessed using the Toolbox, as seen in Figure 9.

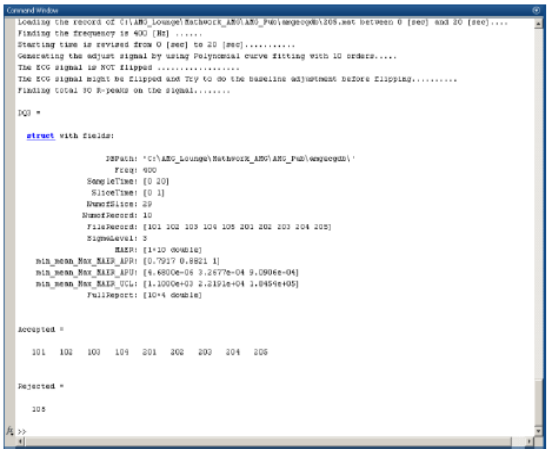

FIGURE 9. Demonstration of MAER based Data Quality Analysis

It is mentioned that readers are allowed to utilize the Matlab source codes (also known as the amgecg Toolbox) that are accessible on GitHub1. Additionally, YouTube2 provided demonstrations of how to use the amgecg Toolbox's features.

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 smart phones 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 research presents a generic 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). The proposed framework offers several data pre-processing techniques and newly defined measure metrics that can still be helpful to researchers developing ML-based schemes, even if they are not using ML technologies for ECG-based biometric authentication.

REFERENCES

[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.

[9] E. Saddik and et al., "Electrocardiogram (ECG) Biometric Authentication", U. S. Patent 9,699,182 B2, Jul. 4, 2017.