Evaluation of Electrocardiogram Biometric Verification Models Based on Short Enrollment Time on Medical and Wearable Recorders

Hazal Su Bı̇çakcı̇, Marco Santopietro, Matthew Boakes and Richard Guest

October 3, 2021

Evaluation of Electrocardiogram Biometric Verification Models Based on Short Enrollment Time on Medical and Wearable Recorders

Hazal Su Bıçakcı, Marco Santopietro, Matthew Boakes, Richard Guest
School of Engineering, University of Kent
{hb488, ms2101, mjb228, r.m.guest}@kent.ac.uk

Abstract—**Biometric authentication is nowadays widely used in a multitude of scenarios. Several studies have been conducted on electrocardiogram (ECG) for subject identification or verification among the various modalities. However, none have considered a typical implementation with a mobile device and the necessity for a fast-training model with limited recording time for the signal. This study tackles this issue by exploring various classification models on short recordings and evaluating the performance varying the sample length and the training set size. We run our tests on two public datasets collected from wearable and medical devices and propose a pipeline for ECG authentication with limited data required for competitive usage across applications.**

***Index Terms*—Biometric Authentication, ECG Biometrics, Performance Assessment, Wearable devices.**

I. Introduction

Over the last decade, with new technologies and the rise of mobile devices, people have started to accept biometric authentication and use it in a multitude of scenarios [1], [2], [3]. This advancement has allowed the development of an increasing number of authentication systems and modalities. Cardiac biometrics [4] and the use of ECG data as biometric features are relatively new. With the release of accurate sensors on mobile devices such as smartwatches and fitness trackers, it is now possible to consider them as a new authentication modality [5], [6] for the commercial user [7].

ECG measures the heart's electrical activity in terms of polarization and depolarization of the muscles according to the various phases of the heart's cycle. A standard medical diagnostic ECG system consists of 12 leads PC-based recorder. The first 6 leads are known as the limb leads, measuring the voltage differences between the right arm (RA), left leg (LL) and left arm (LA) as bipolar and unipolar (augmented). The latter 6 are the precordial leads, and they calculate the differences between electrodes on the chest and the average potential from the limb electrodes. Portable versions of 12-leads ECG recorders such as Holter devices that have 6 leads [8]. Holter devices are standard medical devices because of their 24 hours or 48 hours ECG recording and storage capacities [9]. Although using 12 channels would provide much more information for biometric purposes, it is unlikely to be used in real-life scenarios.

Standard medical devices like Holter devices [8] produce less noisy signals than wearable devices such as smartwatches, armbands, and chest bands [10] when collecting ECG signals. We can attribute the cause for this noise to multiple factors, including the electrode types (dry or wet types), numbers (1, 6 or 12 leads), and their locations (chest and wrist). While medical-based ECG recorders have 12 or 6 wet type electrodes, wearable devices have 1, 2 or 3 dry type electrodes. Usually, only the first lead, the RA-LA lead, is used as it is the easiest way to record and implement in a mobile device, although some portable devices [11] have been able to implement the first six leads. In addition, many chest bands like a Qardiocore device [12], armbands [13], and wrist bands like an Amazfit's Health Band 1S [14] have been used for ECG recordings via 1 or 2 leads.

Medical ECG recorders give more reliable data than wearable devices because of long and detailed recording and the higher complexity of the setup. The primary issue with ECG biometrics has been the time required to capture the sample or the long sessions for enrollment, compared to faster modalities (i.e. fingerprint or face recognition). We propose a verification algorithm for mobile devices that uses the first ECG lead with the lowest recording time and most minor training requirements but remains a competitive option in terms of authentication performance.

In a real-life verification scenario, a person who claims their identity on the system will be accepted or rejected according to the verification task. Therefore, verification time is crucial for all biometrics systems to ensure effective performance. We simulate a real-life scenario using a short time interval from the ECG records, which come from wearable and medical devices for a verification task and present our results.

II. ECG Overview

An ECG is a recording of the heart's electrical activity, captured using electrodes attached to specific body locations. These electrical activities are the sum of all the electrical waves occurring during the depolarization of the muscles. In a normal cardiac cycle, there are three phases (Fig. 1):

– Initially, the atrial depolarization is triggered by the sinoatrial node (P wave). Atrial repolarization follows this depolarization. Then a new trigger by the atrioventricular node after a short delay.
– The signal travels through the bundle of Hiss to the Purkinje fibres, activating the ventricular depolarization and contraction (QRS complex). During this phase, the

cardiac vector looks like a triangle, hence the resulting characteristic complex.

– During the last phase, ventricular relaxation and repolarization occur (T wave).

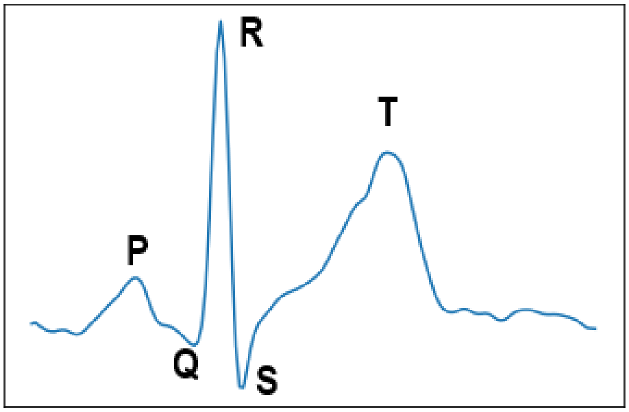

Fig. 1. ECG track for a single heartbeat. P wave, QRS complex and T wave are labeled in proximity of the peaks.

The bandwidth of the cardiac cycle is substantial. It may overlap with other signals (not necessarily biological), but it is consistent with itself (considering healthy subjects) and is unique to each subject. The entire cycle follows the same rules and timing, is triggered by the sinoatrial node. This trigger leads to two important conclusions: 1) another cycle cannot occur unless there is repolarization of the muscle fibres, 2) only the heart rate can be affected by external factors and cannot exceed a particular value.

Therefore, the heart rate is irrelevant, and the only helpful information we can extract is from the single cycle during the various phases.

III. Related Works

Table I for medical-based devices, and Table II for wearable-based devices, summarise the existing related work in the area of ECG authentication alongside corresponding accuracy rates and recording duration.

When considering a real-life application, it is necessary to assume the recording of only a single ECG channel from a mobile device in a short time window. The most probable situation is a verification task when login into a profile or for identity confirmation.

The state-of-the-art of the study is the comparison of medical and wearable-based ECG recorders biometric performances with low training samples using *Deep convolutional network (DL)* and classical *Machine learning (ML)* algorithms. Moreover, comparing the performances of DL and ML algorithms in the case of several training sample sizes. Training sample sizes are generally proportional to verification rates. High training samples return high verification rates. The study aims to present a robust verification system with several training sample sizes for medical and wearable-based ECG data. All the results from the previous works are promising. The diversity of data is essential for the reliability and broad applicability of authentication systems. However, each study lacks a realistic scenario, and in the majority of the cases, a tiny pool of subjects participated in the study, leading to over-fitting and unreliable data. Our proposed method provides consistent verification results for both the medical and wearable ECG data, with few training samples and short enrolment time, both of which would be essential when considering a commercial mobile system.

IV. Materials and Methods

We ran several experiments exploring classical ML models with extracted features and a DL model as a feature extractor followed by classification on a distance metric. All the data used for training and testing come from two datasets, WeSAD (Wearable Stress and Affect Detection) [25], and E-HOL-03-0202-003 [20], to compare verification performance with data obtained with different devices at different sampling rates under different circumstances. There is no previous study that includes the comparison of WeSAD and E-HOL-03-0202-003 dataset performances with different models. In this section, we describe the datasets, models and procedures of our experiments.

A. Databases

The WeSAD dataset [25] consists of 17 subjects ECG recordings collected from a RespiBAN device simultaneously for 36 minutes per subject. Data was collected for the subject while sitting, speaking, and watching video clips in the sitting position. RespiBAN has up to 16-bit sampling resolution and 700 Hz sampling frequency from the chest band. Genders did not distribute equally (3 females, 12 males) [25]. In this study, we used RespiBAN data from 15 subjects.

E-HOL-03-0202-003 [20] consists of 24 hours of continuous digital Holter recordings from 202 healthy subjects collected with three electrodes positioned to obtain the pseudo-orthogonal lead configuration. The dataset indicated that participants had no cardiovascular diseases or disorders, high blood pressure, and chronic illness. The population consists of equally distributed genders (100 male, 100 females and two undefined). The data capture was after 20 minutes resting (supine) period, and 200 Hz sampling frequency and 10mV of amplitude resolution recorded the ECG signals.

We chose these two datasets for the large amount of data provided from healthy subjects and the differences in recording devices, the wearable-based chest bands versus medical-based Holter ECG recorders. Furthermore, the lack of existing studies about verification performance assessments using these selected databases and the many participants in the E-HOL database played a vital role in our dataset selection. We aim to prove that the models are consistent and not biased by device specifications, providing reliable biometric verification with wearable devices.

B. Preprocessing

We filtered the ECG signals from all recordings for low-frequency noise removal and powerline band removal with a

TABLE I
RELATED WORKS OF ECG AUTHENTICATION WITH MEDICAL ECG DEVICES

Studies	*Features*	*# of Subjects*	*Datasets*	*Record Duration*	*Accuracy Results*	*Train/Test Sizes*
Biel et al. [5]	Fiducial Points	20	Private	Unkown, 4,5 and 10 times	100 %	85 samples in training, 50 samples in testing
Israel et al. [6]	Fiducial Points	29	Private	14 min	98 %	20 seconds training
Page et al. [15]	QRS	90	Private	20 sec	99.96 %	70/15/15% training, validation and testing
Wang et al. [16]	Temporal Points	26 (13,13)	PTB & MIT-BIH	Vary	PTB: 84.61%, MIT-BIH:100%	50% training sample sizes
Chiu et al. [17]	QRS	45 (Private), 35(QT)	QT and Private	15 min	100 % (Healthy)	1 min training, 1 min testing
Chan et al. [18]	Non-fiducial	50	Private	270 sec	89 %	90 sec training, 180 sec testing data
Shen et al. [19]	Fiducial points	20	MIT-BIH	48 half-hour excerpts	95% TM, 80% DBNN, 100% Combine	400 training and 200 testing heartbeat samples
Proposed Model	Fiducial and DL	160	E-HOL [20]	24 hours	ML: 3.64 - 6.3% EER in NB classifier, DL: 4.98 - 5.76% EER	ML: 50, 150, 250, 500 sec. , DL: 5, 50, 150, 250, 500 sec in training

TABLE II
RELATED WORKS OF ECG AUTHENTICATION WITH WEARABLE ECG DEVICES

Studies	*Features*	*# of Subjects*	*Datasets*	*Record Duration*	*Accuracy Results*	*Train/Test Sizes*
Sriram et al. [21]	Fiducial and non-fiducial	17	Private	12-15 mins (training), 5-7 mins (Testing)	88 %	4 sec test, 400 secs training per person
Wieclaw et al. [22]	DL	18	Private	147x 10 seconds	88.97 %	30% testing and 70% training samples
Luz et al. [23]	CNN	65 and 100	CYBHi and UofTDB	Vary	1.33 – 14.27% EER	10% validation in training phase
Choi et al. [24]	QRS	175	Private	60 sec	1.87% EER	15 seconds testing time
Proposed Model	Fiducial and DL	15	WeSAD [25]	36 minutes x 15	ML: 3.02 - 4.57% EER in NB classifier, the best 1.6% in DT classifier, DL: 3.79 - 4,98% EER	ML: 50, 150, 250, 500 sec. , DL: 5, 50, 150, 250, 500 sec in training

high pass Butterworth cutoff at 0.5 Hz and a notch at 50/60 Hz (powerline frequency). After this, we created two different preprocessed data structures, one to train the classical machine learning algorithms and one for the deep learning architecture, since the two methodologies work on different conditions. Firstly, we filter the recording for each subject, and N number of samples per subject is trimmed and stored in a matrix. Each sample with the window of size T (number of temporal points for each sample) and delay between windows s (shift time), resulting in a matrix of dimensions $[(N\_subject*N)\times(T+1)]$. Each sample will be further processed to extract a feature vector. The window size considered for each sample is ten seconds, with a one-second shift.

In the second case, we created another data structure with each row representing a single heartbeat cycle. Noise removal filtering follows the same procedure. We store the first 10000 cycles centred on the R peak as single samples for each subject in the datasets. Each sample consists of 150 points, equal to 0.75 seconds of recording. We take the R peak as reference (from the peak, previous 0.25 seconds and following 0.5 seconds are selected). For both datasets, after trimming with the 0.7 seconds window, all samples are downsampled to match the 150-point size, resulting in a matrix of dimensions $[(N\_subject * N) \times 150]$. To detect R peaks, we used the *neurokit2* libraries with the implemented Pan-Tompkins algorithm [26].

C. Feature Extraction

From each sample in the preprocessed matrix, we extract a feature vector with 15 features. The extracted fiducial features are: *QS, meanRR, PQ, Pamp, ST, minRR, maxRR, Qamp, medRR, Ramp, Samp, stdRR, RR50pRatio, Tamp, RR50p* With P, Q, R, S, T waves of the signal, RR50p the number of occurrences of R to R distances shorter than 50 points.

Distances between peaks (*QS*, *PQ* and *ST*) are calculated considering the position of the max value of the two peaks and averaged over the recording. Peak *amp* refers to the mean of the maximum values of the related peaks over the recording. To calculate each peak amplitude in each cycle, we took as reference the R peak position (detected with Pan-Tompkins algorithm). We evaluated the maximum (in case of P and T) or minimum (in case of Q and S) value in a specific confident interval of milliseconds before or after the R peak (Table III). We did not include the number of peaks or specific peak positions because such features depend on the starting value of the recording (which is not fixed) and on the subject activity (an increase in HR implies an increasing number of cycles and, therefore, peaks). The only features partially affected by this issue are the RR distances.

D. Models

We explored the performance of several classification models, fixing the hyperparameters (such as the number of neighbours and number of hidden layers for the DL classifiers) and changing the training size for both datasets. Used models are *Naive Bayes (NB)*, *Decision Tree (DT)*, a *LDA* and a *DL*.

These classifiers are prominent, frequently-used examples of ML models on ECG-based applications [27]. NB and LDA are operated for linear data classification, while DT works on non-linear data. LDA supposes Gaussian circumstantial density model, and this classifier uses normally distributed data. LDA is useful for removing outliers. The calculation time of K-NN is generally longer than NB for significant input data cases, while LDA is fast, easy to use and a simple classifier. NB is the best option to reduce computation costs. NB works based on prior class probabilities for test samples. Unless it encounters the problem of zero probability, NB works appropriately. DT is an effective way to solve classification and regression problems. DT also helps to reduce computation costs. It is more tolerant of missing values and non-normalized data.

TABLE III
DESCRIPTION OF THE EXTRACTED FEATURES. *t* IS THE TIME VECTOR OF THE SAMPLE, N THE NUMBER OF HEARTBEATS IN A SAMPLE. P, Q , R S, T REFER TO WAVE PEAKS.

Features	***Description***	
QS	$\frac{1}{N}\sum_{i=1}^{N} t_{S_i} - t_{Q_i}$	Distances between peaks (averaged over sample)
PQ	$\frac{1}{N}\sum_{i=1}^{N} t_{Q_i} - t_{P_i}$	
ST	$\frac{1}{N}\sum_{i=1}^{N} t_{T_i} - t_{S_i}$	
Pamp	$\mu\{P_i\}_{i=1}^{N}$	Peak amplitude (averaged over sample)
Qamp	$\mu\{Q_i\}_{i=1}^{N}$	
Ramp	$\mu\{R_i\}_{i=1}^{N}$	
Samp	$\mu\{S_i\}_{i=1}^{N}$	
Tamp	$\mu\{T_i\}_{i=1}^{N}$	
minRR	$\min\{t_{R_i} - t_{R_{i-1}}\}_{i=2}^{N}$	R-R distances statistics
maxRR	$\max\{t_{R_i} - t_{R_{i-1}}\}_{i=2}^{N}$	
medRR	$m\{t_{R_i} - t_{R_{i-1}}\}_{i=2}^{N}$	
meanRR	$\mu\{t_{R_i} - t_{R_{i-1}}\}_{i=2}^{N}$	
stdRR	$\sigma\{t_{R_i} - t_{R_{i-1}}\}_{i=2}^{N}$	
RR50p	$\dim(A), A = \left\{RR50p \;\middle|\; RR50p = RR_i < 50\right\}$	
RR50pRatio	$\frac{RR50p}{\dim(RR)}$	

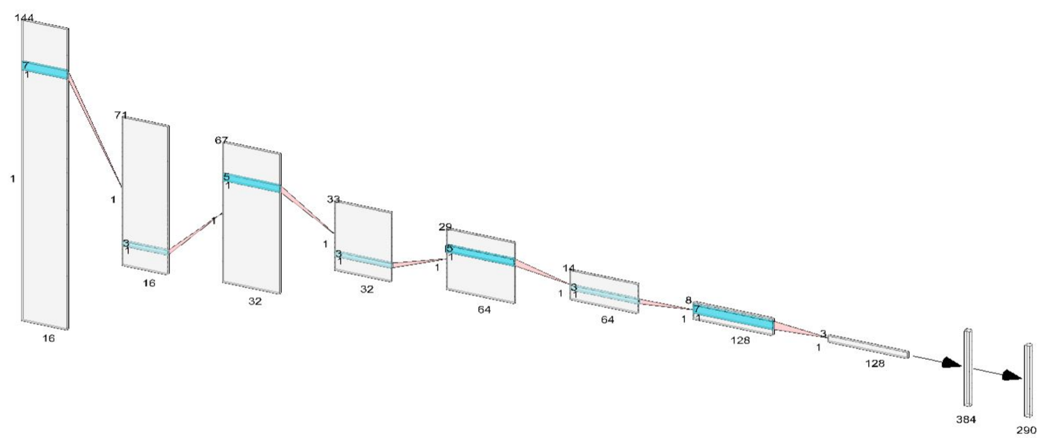

Fig. 2. Visualization of the deep learning model. Blue sections inside layers represents the filters. Numbers describe layers and filters dimensions (height, width and depth).

For this reason, it might be an effective way for wearable-based data. We use NB for solving linear, parabolic and elliptic decision boundaries. For this reason, we used several classifiers to compare their results in different training sample size cases. The DL model is a unidimensional convolutional network based on a previous implementation by *Labati et al.* [28], but with a smaller structure and different set of weights and parameters to make it faster. It has four successions of convolutional and pooling layers, a dropout, a flatten layer and a fully connected layer (Fig. 2). During the training phase, we used a softmax layer and a cross-entropy loss function. During the validation with unseen subjects, we used the output of the fully connected layer as a feature extractor.

We created the user pattern averaging the extracted features from five samples after an L2 normalization, and later, we performed authentication calculating the euclidean distance between the pattern and testing sample.

V. EXPERIMENTS AND RESULTS

We performed several experiments on the different classifiers with fixed hyperparameters and varied both databases' train and test sizes.

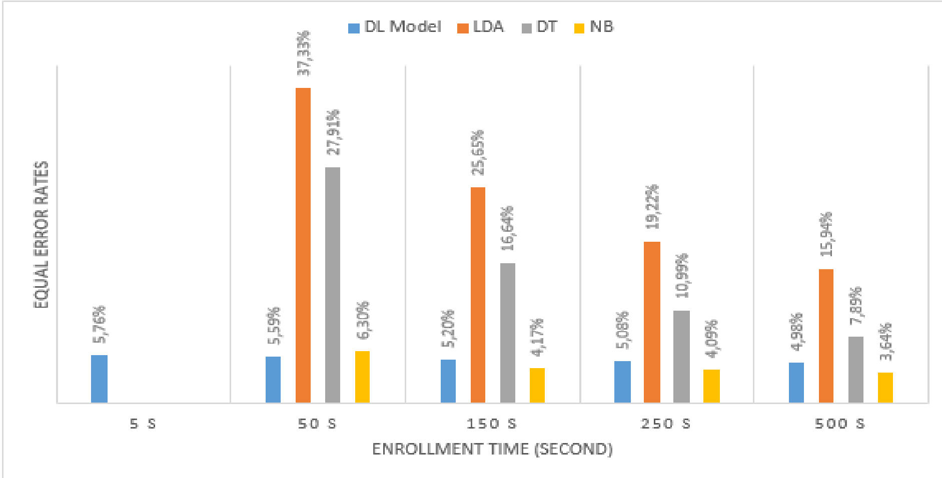

Fig. 3. Comparison of results in terms of EER with different classifiers varying the enrollment time for E-HOL dataset. Authentication is performed with a 10 seconds window (3 seconds or less for the convolutional model), therefore 5s ML results cannot be calculated.

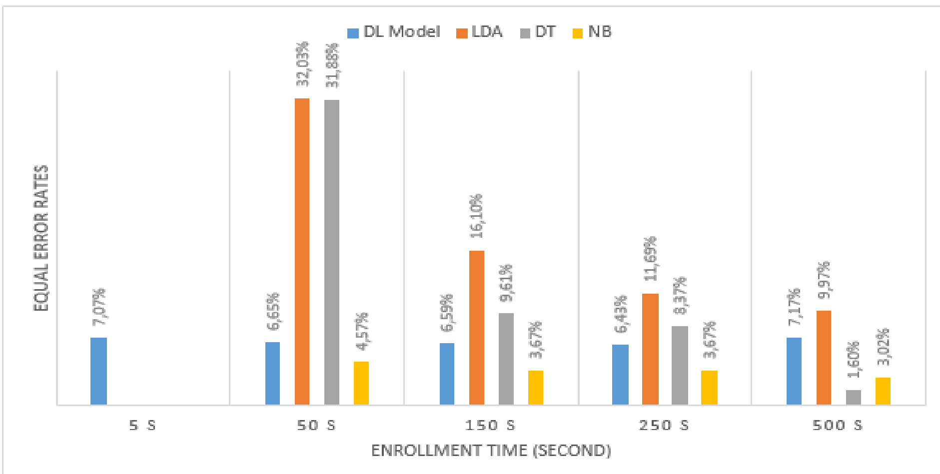

Fig. 4. Comparison of results in terms of EER with different classifiers varying the enrollment time for WeSAD dataset. Authentication is performed with a 10 seconds window (3 seconds or less for the convolutional model), therefore 5s ML results cannot be calculated.

For classical ML classifiers, we used a 10 seconds time window for each sample during the preprocessing. We calculated the mean EER for each model. We used 50, 150, 250 and 500 seconds of samples from genuine and imposter subjects for training; in the case of the DL algorithm, we also included a smaller training window of 5 seconds that classical models could not achieve. We selected imposter samples at random from the remaining users in the dataset different from the enrolled subject. In the testing protocol, we used randomly selected impostors and genuine samples for binary classification. The authentication time corresponds to the single sample window of 10 seconds (2 seconds for the DL model).

In the DL algorithm, we used all the preprocessed samples from 160 subjects of the E-HOL dataset to train the weights and the remaining for validation. Then, we used the various samples from each unseen subject to create the pattern and used the euclidean distance as the similarity metric according to the different enrolling times. Thus, we had two test sets: unseen users from the E-HOL dataset (Fig. 3) and unseen users from the WeSAD dataset (Fig. 4).

We achieved the best results with NB and DT classifiers, both models achieving an EER lower than 5% and 10%, respectively, with an enrollment time of 150 seconds. The DL model also achieved 5.76% EER on the E-HOL dataset and 7.07% EER on the WeSAD dataset with only 5 seconds of enrollment. In both cases, the short time required for enrollment and verification makes it the best choice.

VI. Conclusions

The results lead to many conclusions; it is vital to observe how statistical methods perform better than hyperplane separators. NB has the best performance in terms of EER. However, the DL model outclasses it in terms of enrollment samples required and the time window on verification. Our results are slightly worse than previous studies; however, our evaluation is against many subjects using the most miniature recording time windows and training samples. We outperform all the previous works in terms of enrollment samples required and verification time. Only [21] provides a challenging methodology with only four seconds of recordings required for authentication. However, due to the five minutes of enrollment required, their proposal becomes hard to be accepted by the public. Our DL model requires a shorter enrolment, only 5 seconds of recorded data, providing fast authentication on the query. We also proved the consistency and generalisation of our work over different datasets and devices, as can be seen in Fig. 3 and Fig. 4. The novelty of the proposed method is a robust authentication for medical and wearable ECG recorders, even in the case of short enrollment times.

We can conclude that ECG biometrics will be a valid verification option and could be used in the future as an authentication method, not only continuously, through wearable devices.

REFERENCES

[1] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, *Guide to biometrics*. Springer Science & Business Media, 2003.

[2] J. Chirillo and S. Blaul, *Implementing biometric security*. Hungry Minds, Incorporated, 2003.

[3] G. Kaur, G. Singh, and V. Kumar, "A review on biometric recognition," *International Journal of Bio-Science and Bio-Technology*, vol. 6, no. 4, pp. 69–76, 2014.

[4] A. S. Rathore, Z. Li, W. Zhu, Z. Jin, and W. Xu, "A survey on heart biometrics," *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, pp. 1–38, 2020.

[5] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "Ecg analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[6] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "Ecg to identify individuals," *Pattern recognition*, vol. 38, no. 1, pp. 133–142, 2005.

[7] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, "Ecg authentication for mobile devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 591–600, 2015.

[8] J. E. Deanfield, "Holter monitoring in assessment of angina pectoris," *The American Journal of Cardiology*, vol. 59, no. 7, pp. C18–C22, 1987, a Symposium: The Concept of the Total Ischemic Burden. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0002914987901913

[9] S. Ali, F. Sobhy, M. Abd El Aal, and S. Mahamoud, "Prospective study of holter monitoring in pediatric patients at sohag university hospital," *World Journal of Cardiovascular Diseases*, vol. 10, no. 7, pp. 460–472, 2020.

[10] Q. Zhang, D. Zhou, and X. Zeng, "Pulseprint: Single-arm-ecg biometric human identification using deep learning," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE, 2017, pp. 452–456.

[11] T. Martin, E. Jovanov, and D. Raskovic, "Issues in wearable computing for medical monitoring applications: a case study of a wearable ecg monitoring device," in *Digest of Papers. Fourth International Symposium on Wearable Computers*. IEEE, 2000, pp. 43–49.

[12] Qardio, "Qardio smart wearable ecg/ekg monitor," https://www.getqardio.com/qardiocore-wearable-ecg-ekg-monitor-iphone/, 2018.

[13] J. Lázaro, N. Reljin, M.-B. Hossain, Y. Noh, P. Laguna, and K. H. Chon, "Wearable armband device for daily life electrocardiogram monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 12, pp. 3464–3473, 2020.

[14] E. Chen, J. Jiang, R. Su, M. Gao, S. Zhu, J. Zhou, and Y. Huo, "A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation," *Heart rhythm*, vol. 17, no. 5, pp. 847–853, 2020.

[15] A. Page, A. Kulkarni, and T. Mohsenin, "Utilizing deep neural nets for an embedded ecg-based biometric authentication system," in *2015 IEEE biomedical circuits and systems conference (BioCAS)*. IEEE, 2015, pp. 1–4.

[16] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP journal on Advances in Signal Processing*, vol. 2008, pp. 1–11, 2007.

[17] C.-C. Chiu, C.-M. Chuang, and C.-Y. Hsu, "A novel personal identity verification approach using a discrete wavelet transform of the ecg signal," in *2008 International Conference on Multimedia and Ubiquitous Engineering (mue 2008)*. IEEE, 2008, pp. 201–206.

[18] A. D. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE transactions on instrumentation and measurement*, vol. 57, no. 2, pp. 248–253, 2008.

[19] T.-W. Shen, W. Tompkins, and Y. Hu, "One-lead ecg for identity verification," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, vol. 1. IEEE, 2002, pp. 62–63.

[20] J.-P. Couderc, X. Xiaojuan, W. Zareba, and A. J. Moss, "Assessment of the stability of the individual-based correction of qt interval for heart rate," *Annals of Noninvasive Electrocardiology*, vol. 10, no. 1, pp. 25–34, 2005.

[21] J. C. Sriram, M. Shin, T. Choudhury, and D. Kotz, "Activity-aware ecg-based patient authentication for remote health monitoring," in *Proceedings of the 2009 international conference on Multimodal interfaces*, 2009, pp. 297–304.

[22] L. Wieclaw, Y. Khoma, P. Fałat, D. Sabodashko, and V. Herasymenko, "Biometrie identification from raw ecg signal using deep learning techniques," in *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 1. IEEE, 2017, pp. 129–133.

[23] E. J. Silva Luz, G. J. P. Moreira, L. S. Oliveira, W. R. Schwartz, and D. Menotti, "Learning deep off-the-person heart biometrics representations," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1258–1270, 2018.

[24] H.-S. Choi, B. Lee, and S. Yoon, "Biometric authentication using noisy electrocardiograms acquired by mobile sensors," *IEEE Access*, vol. 4, pp. 1266–1273, 2016.

[25] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

[26] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. A. Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, pp. 1–8, 2021.

[27] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–16, 2019.

[28] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ecg: Convolutional neural networks for ecg biometric recognition," *Pattern Recognition Letters*, vol. 126, pp. 78–85, 2019.