FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Going 2D: Exploring Learnable Bidimensional Representations for ECG Biometrics

Guilherme Augusto Tiritan Romano Barbosa

Mestrado em Bioengenharia

Supervisor: Jaime dos Santos Cardoso, PhD

Co-Supervisor: João Ribeiro Pinto

July 24, 2022

Going 2D: Exploring Learnable Bidimensional Representations for ECG Biometrics

Guilherme Augusto Tiritan Romano Barbosa

Mestrado em Bioengenharia

July 24, 2022

Abstract

For several years, security measures relied on physical devices such as keys and tokens, or passwords and PINs. However, keys and tokens can be lost. Passwords and PINs can be forgotten. These methods are not perfect, and biometrics offer a solution. A biometric signal can't be overlooked or lost, because it is a unique part of the individual. The global biometrics market reached a value of US$ 23.5 Billion in 2020. This market is expected to grow at a CAGR (Compound Anual Growth Rate) of around 17.5% from 2022 until 2027.

ECG (eletrocardiogram) biometrics is an exciting and promising section of biometrics in general. From a commercial perspective, ECG biometric systems expect market growth. Forecasts from 2021 to 2026 show an increase in market size, adhesion in the healthcare, banking, insurance, and aviation sectors. ECG provides the client with a convenient and safe security measure. Especially with off-the-person signals, where convenience to the user increases tremendously. However, accuracy in these systems can still be improved, with outside-the-box thinking and innovative ideas.

To further perfect these models, we propose a new approach. A 1D to 2D transition incorporated in the deep learning model. This way, an end-to-end model creates a 2-dimensional representation that optimizes the identification process. Our main goal is to use this idea and improve state-of-the-art, especially in off-the-person databases. This increased performance can have an impact on people´s lives. Improving the convenience and safety of access to devices and personal information.

This document provides an overview of ECG fundamentals and biometrics state of the art. This gives context to the solution proposed. In addition to summarizing important information for the development of this project. Our baseline one-dimensional model outperformed both the two-dimensional approach and the end-to-end model.

With the PTB database, the accuracy of the 1D model is 97.81 %. The best result within the two-dimensional approaches is 96.8 %. Finally, the end-to-end model obtained a result of 56,08%. Considering the results, the largest issue to tackle seems to be regularization, as larger models are more affected by overfitting, which calls for more efficient solutions

Keywords: ECG, CNN, Biometrics

Contents

List of Figures

List of Tables

Abbreviations

ECG	Eletrocardiogram
CNN	Convolutional Neural Network
MLP	Multilayer Perceptron
LDA	Linear Discriminant Analysis
PCA	Principal Component Analysis
MTF	Markov Transitional Field
GAF	Gramian Angular Field
PTB	*Physikalisch-Technische Bundesanstalt*

Chapter 1

Introduction

A mutating world now offers us multiple platforms in different formats that modify the way we pay, access services, and live our lives. The practicality of these connected platforms requires a secure authentication process to maintain our data safe. Biometrics is a popular answer to this authentication challenge. It is a largely growing field [19].

ECG (electrocardiogram) identification is gaining increasing importance inside biometrics in general. This method has distinct characteristics that set it apart from the remaining biometrics world. It can give users a more convenient and comfortable authentication and identification method, with reliable and robust security features.

Traditional biometric approaches such as face, fingerprint, iris, and voice recognition all have different strengths and flaws that can be explored. These methods complement each other in different dimensions. ECG biometrics can have a space here and might be a valuable player in the biometric world.

1.1 History of the ECG from Medical Grade Acquisition to Off-the-Person Signals

The history of the ECG begins around 1887 with Augustus Waller. Waller uses a Lippmann capillary electrometer (a device that detects small rushes of electric current) to record the electrical activity in both humans and animals through surface electrodes. In 1888, he used saline solutions connected to electrodes in the limb extremities to monitor the heart's electrical activity [9, 73]. This was the first methodical non-invasive study of the heart using electric signals.

Willem Einthoven continued the studies of Augustus Waller and improved his work with a Lipman capillary electrometer. Einthoven improved the distortions mathematically until he obtained a clear electrocardiogram signal. Later on, he included a string galvanometer (a single fine filament of wire suspended in a strong magnetic field that measures small currents) in the electrocardiogram measurements [68]. Einthoven even writes an article describing the recording of an ECG with a string galvanometer, that is transmitted by a telephone wire to a nearby hospital [25].

Einthoven presented the triaxial bipolar system, which uses 3 limb leads (Leads I, II, and III). Each lead is a dipole, that has a negative and positive pole. Leads I has a positive pole at the left arm and a negative pole at the right arm. Lead II has a negative pole at the right arm and a positive pole at the left leg. Lead III has a negative pole at the left arm and a positive pole at the left leg. This system is still in place all around the world [12].

Einthoven envisioned the equilateral triangle using leads I, II, and III as the sides and calculating the electrical axis as a vector in the center of this triangle projected onto the frontal plane. This knowledge and Einthoven´s law are base concepts in interpreting electrocardiography to this day [36, 68].

Waller and Einthoven's contributions to science helped develop the foundations for modern ECG clinical interpretation. These scientists were acclaimed for their work. In 1924, Einthoven was awarded the Nobel prize in physiology or medicine, "for his discovery of the mechanism of the electrocardiogram" [1]. At the time Einthoven received his prize, Waller had already been deceased for two years.

The ECG development continued with the precordial leads. Waller first experimented with precordial leads, because the Lipman capillary electrometer is not as sensitive to the electrocardiogram signals. To have a clearer result, Waller placed his electrodes over the precordium closer to the heart. However, after Einthoven's developments, only the three standard leads were implemented clinically.

Thomas Lewis and other scholars experimented with precordial leads. In 1932 Wolferth and Wood introduced the first precordial lead in clinical diagnostic cardiology. This development opened many new possibilities in medical diagnosis.

Afterwards CL, CR, CF, CB, and V leads were implemented. There was a lot of discussion regarding the right placement for these leads. In 1938, the American Heart Association and the Heart Society of Britain and Ireland published the standard precordial leads. Leads V1 through V6 eventually became the standard worldwide [11].

At last, in 1942 Goldberg developed the 3 augmented leads (aVR, aVL, and aVF), thus completing the 12 lead standard electrocardiogram present today [44].

Recently off-the-person signals have come into play. Off-the-person sensors are embedded into everyday objects and don't require voluntary action from the user. This is envisioned for healthcare to obtain a daily ECG signal acquisition from the patient. The gold standard in off-the-person signals is the non-gelled hand palm electrodes. Silva et al. [20] demonstrated that this signal is fairly correlated with the signal provided by Lead I.

In addition to its relevance in healthcare, the continuous acquisition of electrocardiogram signals opens new doors for biometrics. Nowadays, multiple platforms require secure access and authentication. ECG acquisition from keyboards, remote controls, and other day-to-day objects can provide a reliable source of authentication. The study of ECG biometrics taps into this field and benefits greatly from the use of off-the-person signals, mainly for its convenience.

[1] The Nobel Prize in Physiology or Medicine 1924

1.2 Biometric Applications

Between USB drives with fingerprint readers, face and voice recognition to access computers, and biometric locks with iris recognition or fingerprint identification - biometrics are becoming more and more a part of our daily lives. Their presence even gives birth to companies that operate solely to unify and manage this biometric information. For example, Aware [1] is a company that works with businesses and government agencies to offer biometric software solutions that address security problems.

The traditional biometric signals are the face, fingerprint, iris, and voice recognition. Iris and fingerprint are the most reliable. While face and voice recognition are more convenient for the user [1].

The ECG offers a new perspective, mixing convenient off-the-person acquisition with acceptable performance and reliability [56].

ECG biometrics have inherent features that offer an advantage over other biometric traits. The ECG signal allows for continuous identification. It can be acquired periodically without any major inconvenience to the user. It offers a safer alternative and improves security. In addition to this, the ECG signal has inherent liveness detection. Normally, the ECG can only be measured in a living person, and the signal is hard to circumvent. These features are advantages that favor the ECG as a biometric trait.

Applications in ECG biometrics depend on the sensors responsible for the acquisitions. A study by Plácido da Silva et al. [55] explores various forms of signal acquisition. For example, a computer panel, a video game controller, a mobile device, and a standalone module specific for signal acquisition. There are various options and possibilities for convenient acquisition in day-to-day devices, the ultimate goal of ECG biometrics.

Some companies have already explored this concept. For example, the Nymi band had originally a bracelet designed for identification based on ECG biometrics. CardioID [2] developed CardioWheel, a steering wheel that collects your electrocardiogram. The algorithm detects fatigue and drowsiness keeping the driver alert. It can also identify the driver using ECG biometrics.

ECG-based algorithms are used in multiple situations. Companies such as B-secur develop ECG-related software directed towards health, wellness, and even identification. This set of algorithms is called HeartKey [3] . HeartKey algorithms include heart rate measurement, heart rate variability, arrhythmia analysis, measuring stress state, calculating calories burned over time, and user identification (User ID). The User ID algorithm is used for biometric identification. This enables a more secure physiological data storage and accounts for another application within the ECG biometrics realm.

ECG biometric applications are becoming more widespread. The convenience associated with off-the-person signals makes ECG authentication-based processes quite appealing.

---

[1] Aware
[2] CardioID
[3] HeartKey

1.3 The Learnable 2D Approach

Given that in the broad spectrum of computer vision, 2D inputs are more advantageous than 1D inputs.

The deep learning world is directed towards images. The majority of pre-trained models rely on image inputs, data augmentation techniques are designed to be implemented on two-dimensional representations. The entire deep learning process is largely image based. In heart biometrics, Ribeiro et al [57] with 1D CNNs, and Luz et al. [22] with 2D CNNs show favourable results.

In our assignment, various 2-dimensional approaches are explored. Afterwards there is an end-to-end model development. The input signal is one dimensional. The model transforms the 1D representation into an image. The image is than classified according to the subject's identity.

We experimented with four different 2-dimensional approaches: an ordinary spectrogram, a Markov Transition Field (MTF), a Gramian Angular Field (GAF) and the S-transform. Within these models, the one with better results in biometric identification is transfered to the end-to-end model.

A Multilayer Perceptron is pre-trained to mimic the two-dimensional representation. Afterwards, we add the classification algorithm. The result is quite a complex model, where regularization is paramount. Various different regularization methods are applied.

Our goal is to improve the state-of-the-art especially in off-the-person signals. This accounts for real impact especially in ECG biometric applications.

1.4 Document Structure

The first part of this document, begins with a general knowledge of the electrocardiogram features and acquisition. From here, we move on to the available databases that might be used to develop our project and show the advantages of using an ECG compared to the general biometrics landscape.

Secondly, we summarize of state of the art, even presenting some devices that use the ECG signal. Finishing with an explanation of our solution and describing the impact it can have.

Chapter 2

ECG Fundamentals

The motion of the heart is the result of electrical stimulation. The heart contracts and relaxes due to sequential depolarizations followed by repolarizations. This electrical activity can be measured and visualized as an ECG signal. This signal varies for the same person in different conditions and differs between individuals.

2.1 ECG Morphology

The ECG morphology is a set of different curves combined. These curves are presented in Fig. 2.1.

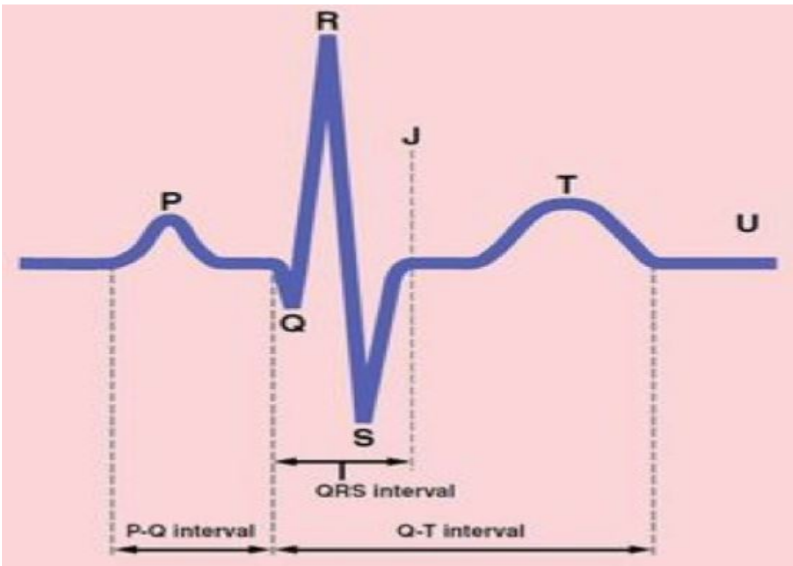


Figure 2.1: ECG Signal Segments ECG (EKG) Interpretation - OME .

First, the P wave reflects atrial depolarization. The QRS segment refers to ventricular depolarization. The left ventricle has a larger electrical vector than the right one. So, the QRS mostly accounts for the left ventricle´s electrical cardiac output. The ST segment (distance between S and T) is usually flat. The J point marks the beginning of the ST segment. The T wave is smooth and slightly asymmetrical, with a sharper downward slope. The U wave is not always present. Its amplitude is usually one-fourth of the T wave [1].

2.2 Variability

In inter-subject variability, there are static and dynamic cardiac features. Static cardiac features are related to the geometrical aspects of the heart and its orientation. Physique, age and gender have a big influence on static features. Dynamic cardiac features are related to the depolarization and repolarization timing, along with the electrical conduction within the thorax. Regular physical activity modifies some components of the signal [56].

Intra-subject variability is introduced by gradually changing factors like body habitus and age. Physiological responses to stress and other stimuli can alter the ECG within the same individual. Signal artifacts related to movement and electrode placement interfere with signal quality. The ECG also changes in response to pharmacological drugs and physical activity. Active people have a more variable ECG [66].

The ECG signal comes from the difference in potential between the electrodes. A positive electrode and a negative electrode (reference) comprise an ECG lead. Different leads, in different positions, provide projections of the heart´s electrical activity on multiple axes.

2.3 Signal Acquisition

There are various possible electrode configurations. The most common configurations are modified chest lead (MCL) and 12 lead. In MCL, there are 3 electrodes. A positive electrode placed in V_1, a negative one placed in LA (left arm), and the reference in RA (right arm).

In the 12 lead representation - there are three bipolar limb leads (I, II, and III), six unipolar chest leads (V_1 to V_6), and 3 augmented unipolar limb leads (aVR, aVL, and aVF). The placement of electrodes from the multiple representations is in Fig. 2.2.

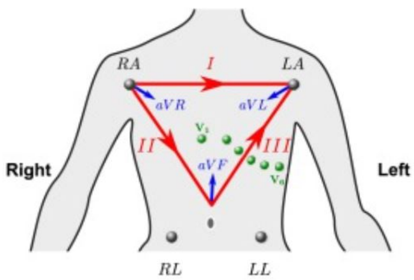


Figure 2.2: Lead and Electrode placement for MCL and 12 Lead Representations [61].

[1]ECG interpretation: Characteristics of the normal ECG (P-wave, QRS complex, ST segment, T-wave)

In the Frank Leads method, cardiac activity is represented by a vector of specific magnitude and direction, in a three-dimensional orthogonal axis. In this configuration, there are seven electrodes used to determine three components: mediolateral (right-to-left or X-axis), craniocaudal (head-to-feet or Y-axis), and anteroposterior (front-to-back or Z-axis). Pairs of these components create the orthogonal plains, where spatial curves of the heart´s electrical activity are projected [61]. Figure 2.3 represents the axis in Frank configuration.

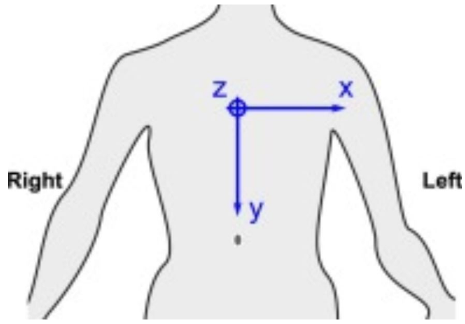
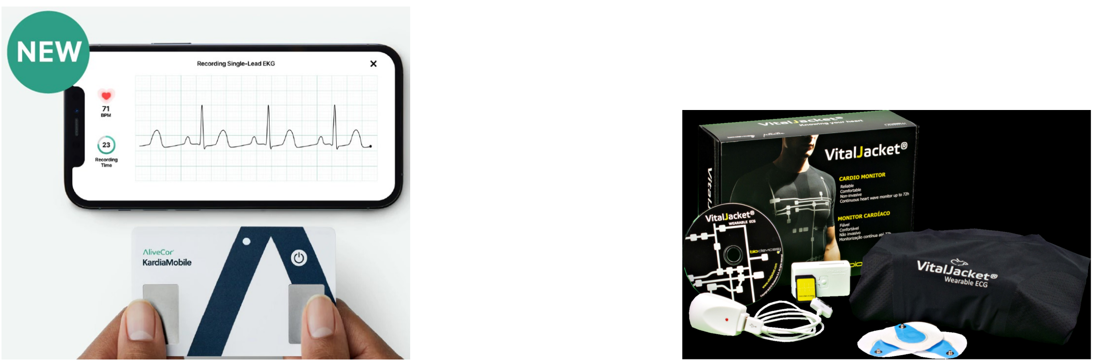


Figure 2.3: Frank Representation of ECG Signal [61].

In addition to the traditional methods, there are other ways to measure an ECG. A new concept emerging is off-the-person ECG. Non-gelled electrodes or conductive fabrics measure the signal. A less invasive method applied in everyday objects, like a computer keyboard or a phone. Off-the-person methods can even be contactless. Off-the-person ECG aims to acquire the ECG without any voluntary action from the individual [20].

On-the-person methods include both stationary and ambulatory systems. The stationary systems are the standard various leads used in hospitals, mentioned above. The ambulatory systems include wearable equipment, such as t-shirts, watches, and phones. Still, in these methods, the signal acquisition requires a voluntary action from the user [20]. An example of this devices is the Vital Jacket from Biodevices - or the AliveCor software accessory. Figure 2.4a shows the KardiaMobile Card a product designed by AliveCor that can read the ECG signal within 30 seconds. Figure 2.4b represents the Vital Jacket from Biodevices.



(a) Mobile Card. (b) Vital Jacket Wearable ECG.

Figure 2.4: Two Examples of Ambulatory Acquisition Systems

2.4 Summary and Conclusions

The electrocardiogram signal is fundamentally a difference in electrical potential between two electrodes. There are various signal acquisition methods related to the ECG. They vary in intrusiveness and comfort for the subject. In general, more intrusive methods are related to better signal quality.

In biometrics, off-the-person signals are quite interesting. These signals are more convenient for the user, and with the right algorithm may perform well in biometric tasks. Still, off-the-person are usually more noisy and present quite a challenge in biometrics.

Chapter 3

Use of Biometrics and its Advantages

3.1 Biometric in General.

A physiological signal or behavioral characteristic needs to comply with at least four requirements, to be a biometric feature:

- Universality - It should present in everyone.
- Distinctiveness - Has to vary between people to allow identification.
- Permanence - Must be invariant for a determined period, related to the matching criterion.
- Collectability - A measurable characteristic.

Nonetheless, there are still other issues to be addressed in a realistic biometric system:

- Performance - The accuracy and quickness of the process under different environments. The resources necessary to achieve this performance.
- Acceptability - The length of acceptance of this feature in daily lives.
- Circumvention - Fragility towards fraudulent methods.

Biometrics are remarkably convenient for users. There´s no need to remember multiple complicated passwords while maintaining a certain degree of security. External factors such as keys that may be stolen or lost are also unnecessary with biometrics [35].

The most common biometric modalities include face, fingerprint, iris, and voice recognition. In Table. 3.1 each of these methods is ranked according to the criteria mentioned above.

Table 3.1: Table with biometric identifier and their ranking (High, Medium, Low) by features adapted from [1].

Biometric Identifier	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	High	Low	Medium	High	Medium	High	High
Fingerprint	Medium	High	High	Medium	High	Medium	Medium
Iris	High	High	High	Medium	High	Low	Low
Voice	Medium	Low	Medium	Low	Low	Medium	High
ECG	High	High	High	Medium	Low	Medium	Low

Face and iris are universal methods, but some people do not have a voice or fingers. Iris and fingerprints are different, even in identical twins. Face and voice change with aging and other factors. Iris and fingerprints remain relatively constant. Iris, fingerprint, and voice systems require user participation, while face images are easily collected. Fingerprints and iris are usually more trustworthy. However, voice and face recognition have higher user acceptability.

Iris circumvention usually done with contact lenses is not very effective [35]. Fingerprints are circumvented with manufactured fingerprint patterns. A picture or mask maybe is enough to circumvent face recognition systems [27]. In voice recognition, recorded voice speakers are sometimes sufficient to sidestep the system [1].

3.2 ECG Biometrics Advantages and Flaws

Each biometric trait has advantages and disadvantages. ECG biometrics can bring something new to this field. Characteristics such as liveness detection and continuous authentication can be game-changers.

The ECG is a universal signal, that everyone has. In terms of distinctiveness - inter-subject variability of the ECG is covered in the ECG fundamentals.

In permanence, the ECG signal can vary due to the body's physiological conditions developed over time. On top of this, electrocardiograms are sensible to noise provoked by breathing, body motion, and even lead placement. This endangers the biometric security system and makes the process more susceptible to Denial-of-Service attacks.

Collectability is related to off-the-person and on-the-person acquisition methods. Off-the-person is much less invasive but has a lower quality.

In terms of performance, it is difficult to design capable heart biometrics with inexpensive and unobtrusive data.

As for acceptability, off-the-person signals will be easier to adjust into daily lives. However, this type of acquisition can compromise signal quality.

In terms of circumvention, ECG signals are hard to bypass. Because of signal intrinsic liveness detection and continuous authentication. Intrinsic liveness detection guarantees that only a living

[1] CSO Online

person can be recognized by this modality. This helps fight malicious attacks, that in other systems might require additional costs.

Continuous authentication, where the subject provides new biometric samples periodically. Offers an advantage compared to other biometric modalities with static biometric samples. That can be more easily deceived by ill-intentioned attackers [56].

Nonetheless, it has been proven that attacks on ECG biometric systems are possible. In a study conducted towards attacks on the Nymi Band, the ECG data acquired from the subject is stored in a device and injected by a different person. In case the owner´s ECG is obtained by another Nymi band there´s an 81% success rate. If the signal is acquired by other methods there is a 61% success rate, with the application of a mapping function developed by the researchers. Although, these attacks have moderate success. The method applied requires a knowledge of signal processing techniques that not everyone has [24].

Besides the Nymi Band, other technological enterprises are using ECG biometrics in the markets or simply monitoring ECG for healthcare. For example, Cardio ID [1] developed Cardio Wheel. A steering wheel that detects drowsiness, cardio health problems and does biometric identification using ECG. Even Samsung and Apple use ECG signals to monitor healthcare. Applications such as Samsung Health Monitor[2], show the user´s heart rate, blood pressure, and even an ECG representation. Apple Watch[3] also has an ECG app, that allows the user to visualize its ECG and monitor its heart rate.

Figure 3.1. summarizes the landscape regarding the systems mentioned in a class diagram.

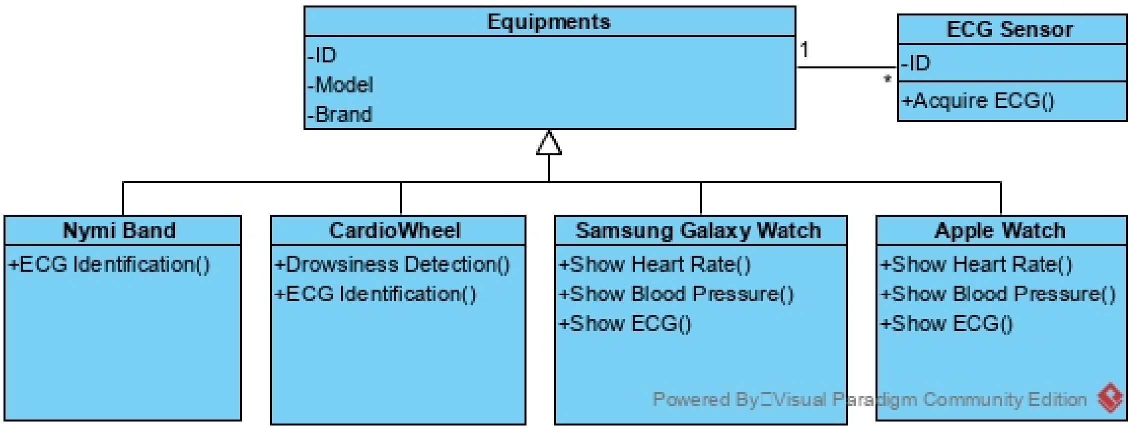


Figure 3.1: Class Diagram for Equipments that use ECG.

3.3 Summary

In conclusion, ECG biometrics is a very interesting and promising field. It certainly has advantages over its counterparts. Nonetheless, there are flaws that can be explored by a possible attacker.

[1] CardioID
[2] Samsung Health Monitor Samsung Portugal
[3] ECG app - Apple Watch

In addition to this, traditional ECG acquisition methods are fairly invasive and uncomfortable.

Off-the-person acquisition is quite promising. It allows for a more comfortable and convenient use of the signal. However, ECG performance with this acquisition type can be improved greatly . To further enhance performance there is a need to use different methods and think outside the box. The aim of this work is to try and develop an innovative algorithm. That improves the state of the art, and brings something new to this field.

Chapter 4

State of the Art

4.1 ECG Databases

4.1.1 Overview

ECG databases are a collection of ECG signals from various patients. The acquisition method and other characteristics change with different databases.

There are 8 main databases covered in this assignment: PTB (*Physikalisch-Technische Bundesanstalt*), PTB-XL (Larger database of *Physikalisch-Technische Bundesanstalt*), MIT-BIH (Massachusetts Institute of Technology - Beth Israel Hospital), ECG-ID (Electrocardiogram Identification), STAFF III, Long Term AF (atrial fibrillation), CYBHi (Check Your Biosignals Here Initiative) and UofTDB (University of Toronto Database).

The PTB has 549 records from 290 subjects, with one to five records by subject. One record contains 15 signals (12 conventional leads and 3 Frank Leads). A clinical summary is available for 268 subjects. Most of them have heart diseases [10]. The signal is measured in mV and the LSB (least significant bit) is 0.5 µV. There´s a total of 16 bits to account for the whole memory, so the signal varies between negative 16.384 mV and positive 16.384 mV - adding up to a range of 32.768 mV. The sampling frequency is 1000 Hz.

Although the PTB-XL database is from the same project as the PTB database, they don´t share any records. This database has 21837 records from 18885 patients. Each record has the traditional 12-lead ECG with a 10-second length. A large segment of the patients is healthy (9528), while the remainder has heart diseases [77].

The MIT-BIH database contains 48 half-hour records of two-channel ambulatory ECG recordings from 47 subjects. Most of the subjects in this database have a heart arrhythmia [48].

The ECG-ID has 310 records from 90 subjects. Each record has two channels: 1. a raw 20-second noisy signal from lead I and 2. filtered signal from lead 1 [42]. Lead I signals are extremely similar to some off-the-person representations [21].

The STAFF III database was developed with the goal of better understanding the ECG during acute ischemia, especially regarding high-frequency QRS components. The database consists of the standard 12 lead ECG. The ECG was acquired at a sampling rate of 1000 Hz and an amplitude resolution of 0.625 μ V. This database has records of 104 patients [46].

The Long Term AF (LT-AF) has 84 long-term ECG recordings from 44 subjects with atrial fibrillation. The sampling frequency is 128 Hz, and there is a 12-bit resolution with a 20 mV range. Record duration varies between 24 to 25 hours [53].

All the acquisition methods from the databases mentioned before are on-the-person.

The CYBHi database aims at biometric identification training. Dry electrodes in hand palms and fingers collect the signal - an off-the-person acquisition. This database has two types of signals: short-term signals (65 subjects) acquired with 2-day intervals between sessions; long-term signals (63 subjects) acquired with 3-month intervals between sessions [64, 21].

The UofTDB is the largest off-the-person database publicly available. The Vernier EKG Sensor [1] acquires the signal. There are six different states: sit, stand, exercise, supine (laid back and relaxed), and tripod (leaned forward while sitting on a chair). Each state is associated with multiple sessions. There are 1020 subjects, 398 males and 622 females, with ages between 18 and 52 years old. From these subjects, we have 1627 records, with lengths that vary from 2 to 5 minutes [78].

The UofTDB has a sampling frequency of 200 Hz, and a 12 bits resolution. All the subjects present had their signal acquired while sitting. However, 63 were recorded in supine, 63 in tripod, 71 in physical exercise, and 81 in standing conditions. From the entire subject pool, 72 participated in two sessions, 65 in three sessions, 54 in four sessions, 47 in five sessions, and 43 in six sessions. The signals in this database are filtered with a digital Butterworth bandpass filter having cutoff frequencies of 0.5Hz to 40Hz [47].

Both the PTB and UofTDB databases are used in this assignment.

All of these six databases are publicly available. Table 4.1 summarizes the databases´ information.

Table 4.1: Database Collection Summary.

Database Name	Nº of Subjects	Nº of Records	Sampling Rate	Number of Resolution Bits	Acquisition Type
PTB [10]	290	549	1000 Hz	16	On-the-Person
PTB-XL [77]	18885	21837	500 / 1000 Hz	16	On-the-Person
MIT-BIH [48]	47	48	360 Hz	11	On-the-Person
ECG-ID [42]	310	90	500 Hz	12	On-the-Person
STAFF III [46]	104	-	1000 Hz	-	On-the-Person
LT-AF [53]	44	84	128 Hz	12	On-the-Person
CYBHi Short Term [64, 21]	65	-	1 kHz	12	Off-the-Person
CYBHi Long Term [64, 21]	63	-	1 kHz	12	Off-the-Person
UofTDB [78]	1020	1627	200 Hz	12	Off-the-Person

[1] EKG Sensor

4.1.2 Conclusions

The main problem with the current databases is related to obtaining a good quality signal from off-the-person representations. Both the CYBHi and UofTDB databases are from 2013 and 2014. New more recent databases could bring new challenges to the table.

A wider variety of off-the-person databases would be beneficial to explore new postures for acquisition. The large majority of ECG databases were designed for clinical purposes, so new databases specified especially for ECG biometrics are a good addition.

Databases based on sensors from smart watches, phones and other everyday devices would also be of interest to test our models in real world conditions.

4.2 Algorithm Development

The process of ECG biometrics has five stages: preprocessing, feature extraction, feature selection, feature transformation, and classification [75]. With deep learning approaches, all of these steps are implemented simultaneously within the model.

4.2.1 Preprocessing

In the preprocessing stage, the aim is to reduce signal noise and motion artifacts. Both noise and artifacts can lead to incorrect authentication [75].

ECG preprocessing may include the following stages: mean removal, amplitude normalization, QRS detection, segmentation, and zero padding. Techniques such as discrete wave transform (DWT) or empirical mode decomposition (EMD) sometimes are implemented to clean the signal [75].

Preprocessing can also include signal segmentation in various heartbeats. Pan-Tompkins algorithm detects the QRS complex providing a marker for signal segmentation [71].

Traditional signal processing techniques such as: high pass filters, low pass filters, band pass filters, notch filters, median filters, and Savitzky-Golay filters are common.

High pass filters cut off the lower frequencies - reducing motion artifacts, respiratory variation, and baseline wander. However, this cutoff can cause alteration in the ST segment of the signal.

Low pass filters eliminate the higher frequencies - reducing high-frequency noise, background noise, and power line interference. Unfortunately, this alters the QRS complex. The recommended cutoff frequency in these filters is 150 Hz for adults and youngsters, and 250 Hz for infants [16].

Bandpass filters define a range of frequencies for the output signal. This filter combines both the low pass and high pass filters. Because of this, it shares the advantages and disadvantages of both.

Notch filters stop a very narrow range of frequencies. Usually set at 50 Hz or 60 Hz to erase the power line interference. But this may distort the output signal.

Median filters are a nonlinear form of signal processing used for baseline adjustment or removal. If the median filter is too big, there is a suppression of the T waves.

Savitzky-Golay filters have applications for smoothing and differentiation in ECG processing. However, they are seldom used in ECG biometrics. Despite this, Gupta et al. [29] obtained outstanding results in preprocessing for R peak detection [75].

4.2.2 Feature Extraction and Transformation

The feature extraction methods are fiducial or non-fiducial. Fiducial features use fixed points in the ECG for identification, called biomarkers. The distance between biomarkers is fundamental in the classification process.

On the other hand, non-fiducial features do not rely on fixed points. This approach uses entire segments of the signal. Fiducial extraction may lose crucial information for identification. On the other hand, non-fiducial feature extraction requires a more sophisticated classification method to select important information within the signal.

Feature transformation uses methods that reduce dimensionality: PCA (principal component analysis), LDA (linear discriminant analysis), SVD (singular value decomposition), and ICA (independent component analysis).

Some of these methods have applications in classification, usually preceded by an autocorrelation of the signal. Hejazi et al. [30] shows an equal error rate (EER) of 34% using autocorrelation of the signal followed by the LDA algorithm for classification. Once applied to the same individual, the EER drops to 14%. A significant number of subjects even have an EER between 0-5%.

4.2.3 Classification

For classification, simpler models based on LDA, PCA, K-nearest neighbors, or a support vector machine (SVM) need the five stages mentioned above for the algorithm to function correctly.

Deep learning approaches such as deep neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) [6, 81] are quite promising in the ECG biometrics field. Ribeiro et al. [57] uses CNNs that simplify this process, feeding the raw ECG to the model with remarkable results. CNNs prove to be a reliable alternative [39].

In Mariusz et al. [52], various baseline results are established in ECG biometrics. The accuracy results for the ECG-ID database and MIT-BIH database are the following for the mentioned methods:

- logistic regression (ECG-ID - 82.86%, MIT-BIH - 74.92%)
- SVM classifier (ECG-ID - 88.17%, MIT-BIH - 77.07%)
- LDA classifier (ECG-ID 93.28%, MIT-BIH - 90.17%)
- KNN classifier (ECG-ID - 89.03%, MIT-BIH - 79.67%)
- MLP - 1 hidden layer (ECG-ID - 89.33%, MIT-BIH - 89.25%)
- MLP - 2 hidden layer (ECG-ID - 89.76%, MIT-BIH 87.44%)

- MLP - 3 hidden layer (ECG-ID - 84.06%, MIT-BIH - 88.08%)
- PCA + Logistic Regression (ECG-ID - 82.86%, MIT-BIH - 73.35%)
- PCA + SVM (ECG-ID - 88.65%, MIT-BIH - 74.72%)
- PCA + LDA (ECG-ID - 95.36%, MIT-BIH - 87.98%)

The baseline provided in [52] shows the potential of deep learning within ECG biometrics. Although, principal component analysis followed by linear discriminant analysis obtained the best results.

CNNs in computer vision usually use a two-dimensional input and obtain noticeable results [23]. To further improve ECG classification with CNNs - researchers transform the 1D ECG signal into a 2D image. This image enters the neural network. The transition of the signal from one dimension to two dimensions can be related to spectrograms or other options [22, 37].

Benouis et al. [7] shows an EER between 10% and 14% in ECG-ID, MIT-BIH and PTB databases and 18% in CYBHi database. Using stacked autoencoders and a 2D grayscale image input obtained from the ECG.

Luz et al. [22] use a one-dimensional CNN, a spectrogram with a two-dimensional CNN, and a fusion model that uses both these methods. The Equal Error Rate (EER) for the CYBHi database in intra-session evaluation (training and evaluation within the same session) is 1.17 % for the one-dimensional CNN, 19.57 % for the spectrogram input, and 1.33 % for the fusion model.

For inter-session classification (training and evaluating on different sessions) with the CYBHi database it is 15.60 % or 14.13 % for the one-dimensional CNN, 26.38 % or 20.48 % for the spectrogram input and 12.78 % or 13.93 % for the fusion model. The different results in inter-session evaluation are related to the sessions used in training and evaluation. In the UofTDB database with inter-session evaluation, the EER for the one-dimensional CNN is 16.92 %, for the spectrogram input it is 19.37 %, and for the fusion model, it is 14.27 %.

Bento et al. [8] uses two distinct convolutional neural networks with spectrograms as input. The first CNN has four convolutional layers, while the second CNN is a DenseNet with nineteen blocks. There are two different databases approached in this method, the Fantasia database, and the ECG-ID. The results from this study are divided into within-session classification and across-session classification.

The simple CNN has an accuracy of 99.42 %, a specificity of 99.98 %, and a sensitivity of 99.42 % in the Fantasia database. In the ECG-ID database, it has an accuracy of 94.23 %, a sensitivity of 94.26 %, and a specificity of 99.94 %. Across-session for the ECG-ID database has an accuracy of 73.54 %, a sensitivity of 72.72 %, and a specificity of 99.70 %.

The DenseNet arrives at an accuracy of 99.79 %, a specificity of 99.78 %, and a sensitivity of 99.99 % in the Fantasia database. In the ECG-ID database, the within-session results are 96.88 % accuracy, 96.89 % sensitivity, and 99.96 % specificity. The across-session results are 73.28 % accuracy, 72.46 % sensitivity, and 99.70 % specificity.

In Min-Gu et al. [37] an ECG image is acquired by estimating a partial baseline using first-order regression analysis and projecting the signal onto the 2D space. The ECG segments are fiducial (single out the heartbeat) or non-fiducial. The researchers develop three different CNNs models (ConvNet1, ConvNet2, ConvNet3) and use them to create an ensemble network. The accuracies for fiducial segments obtained are ConvNet1 - 97.3%, ConvNet2 - 97.9%, ConvNet3 - 97.2% and Ensemble Networks - 98.9%.

The accuracies for non-fiducial segments obtained are ConvNet1 - 97.2%, ConvNet2 - 96.6%, ConvNet3 - 97.7% and Ensemble Networks - 98.6%.

In Lynn et al. [43], there are two initial approaches fiducial and non-fiducial. The raw ECG signals with the QRS complex were filtered and divided based on R-peak detection and heartbeat segmentation. The signals without a QRS complex are randomly segmented, and an autocorrelation technique is applied. The fiducial approach applies to a one-dimensional CNN with four hidden layers.

The non-fiducial approach uses Recurrent Neural Networks (RNN) architectures. The first non-fiducial model is a bidirectional RNN with a Long Short-Term Memory (LSTM) cell unit in the hidden layer state. This is referred to as the BLSTM. The other non-fiducial model has a Gated Recurrent Unit (GRU) in the hidden layer. This model is referred to as BGRU. There's a dropout added to all methods based on RNNs to prevent overfitting. Both these models are implemented in a bidirectional manner.

In this experiment, there are four databases used: ECG-ID, MIT-BIH, STAFF III, and Long Term AF Database (LT-AF). The results are the following, for the 1D-CNN with 3 heartbeats as input - 92,5% accuracy, and with 9 heartbeats as input - 91,1 % accuracy. For the BLSTM with 3 heartbeats as input - 98,2 % accuracy, and with 9 heartbeats as input - 99,3 % accuracy. Finally for the BGRU with 3 heartbeats as input - 92,1 % accuracy and with 9 heartbeats as input - 98,3 % accuracy.

In Chee et al. [14] transformer's self-attention mechanism is used in ECG biometrics. The electrocardiogram is segmented blindly without fiducial points, filtered, and standardized. In the study, we have the pair enconder. The ID enconder consists of 4 tranformers' enconder layers. The ID classifier is a 256-unit fully connected layer, a batch normalization layer with ReLU activation followed by a layer with a Softmax function to output the identification distribution.

In this paper, a long ECG is segmented into enrolment, afterward, there is a time separation until the classification window. When this time separation is 0, there's an accuracy of 96.35% in the ECG-ID database and accuracy of 98.10% in the PTB database. When the classification window is furthest from the enrolment segment, the results are the following 92.70% in the ECG-ID and 64.16% in the PTB database.

4.3 Two-Dimensional Representation

There are four two-dimensional representations approached in our work: spectrogram, S-transform, Markov Transition Field, and Gramian Angular Field.

spectrogram is a graph of the energy content of a signal expressed as function of frequency and time. The graph has time in horizontal axis, frequency in the vertical axis and the amplitude is in grayscale.

Spectrograms are often used as images in deep learning networks for audio tasks. The time-frequency representations of speech offer dynamic information of how the frequency changes over time [4].

In ECG biometrics, spectrograms are also being increasingly used. Odinaka et al. [50] and Bento et al. [8] are examples of studies where the spectrogram representation was leveraged for ECG biometrics.

The S-transform or Stockwell transform is a generalization of the short-time Fourier transform. It provides a frequency dependent solution, maintaining a direct relationship to the Fourier spectrum. It is an extension of the Continuous Wavelet Transform (CWT), based on a moving scalable Gaussian window. The mathematical expression 4.1 describes this methods implementation [70].

$$S(\tau, f) = \int_{+\infty}^{-\infty} h(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-i2\pi f t}\, dx \tag{4.1}$$

Recent applications in deep learning for the Stockwell transform include: modulation classification for Radio Frequency waves [31], multi-class diagnosis of brain pathologies [65], improve hand movement recognition accuracy [62], and studies in geophysics [86, 54].

The Gramian Angular Field (GAF) represents the time series in a polar coordinate system. The GAF maintains the temporal dependency of the signal, since time increases as the position moves from the top-left to the bottom-right.

In the deep learning world, Gramaian Angular Fields have been used for sensor based human activity recognition (HAR) [82], functional near-infrared spectroscopy (fNIRS) processing for task classification [80] and EEG classification [26].

The Markov Transition Field is a visualization technique for time series. The technique begins with a discretization of the input signal. This discretization is used to build the Markov transition matrix and eventually create the Markov Transition Field. MTF images represent the first order Markov transition probability along one dimension and temporal dependency along the other [41, 79].

The Markov Transitional Field has various uses in deep learning. The MTF is used for classification of vibration-events and measure vibration-frequency [84], milling tool condition monitoring [72], and rolling bearing fault diagnosis [83]. Mostly methods based on information from a sensor with a uni-dimensional signal.

In addition to the representations used in our project, many studies have leverage a two-dimensional input for their assignment.

In Ciocoiu et al. [17] 2D/3D phase representations are classified by a convolutional neural network. The input signal is filtered by a bandpass filter with frequencies between 2 and 30 Hz. Outliers are rejected with Normalized Cross-Correlation Clustering (NCCC). Results from

20 subjects in the UofTDB dataset for sitting have an accuracy of 93.6% (Rank-1) and 97.2% (Rank-2). Results from all postures have an accuracy of 90.4% (Rank-1) and 95.8% (Rank-2).

In Ciocoiu et al. [18] multiple two dimensional representations are explored: the S-transform, the Gramian Angular Field, the recurrence plot and phase-space representations. The two-dimensional representations are classified by a CNN.

Benouis et al. [7], Luz et al. [22], Bento et al. [8] and In Min-Gu et al. [37] also explore different two dimensional representations in the field of ECG biometrics.

4.4 Metrics

Accuracy is the proportion of correct predictions both true positives and trues negatives. The accuracy gives an overall view of the models behaviour and model quality. In biometrics, accuracy is even known as identification rate. Equation 4.2 represents the accuracy metric [5].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.2}$$

In equations 4.2, TP represents true positives, TN - true negatives, FP - false positives and FN - false negatives.

The only metric used in our work is accuracy. However, there are other metrics used in biometrics such as: False Acceptance Rate (FAR), False Rejection Rate (FRR) and Equal Error Rate (EER) among others [38].

We chose the accuracy metric given that our main objective is to analyse the overall efficiency of the model, in order to select the best model architecture. A simple metric such as accuracy is enough to determine the overall quality of the model.

4.5 Loss Functions

Loss functions are crucial for any deep learning task. The problem itself is defined through the loss function specified. There are many loss functions available for a variety of different problems.

For multiple class problems Cross-Entropy loss is the right choice. This function compares the output probability vector of a model with the ground truth labels. The function computes a distance between these two vectors [45]. In equation 4.3 Cross-Entropy is defined for n classes :

$$L_{CE} = -\sum_{i=1}^{n} t_i log(p_i) \tag{4.3}$$

In equation 4.3, t_i represents the truth label for class i, while p_i is the probability associated with class i.

The labels in the categorical cross-entropy can be defined in different ways. The Categorical Cross-Entropy is used when the labels are one-hot encoded. A label vector is defined by multiple zeros and a number one in the position of the label intended. In Sparse Categorical Cross-Entropy the labels are encoded as integers. The value of the integer corresponds to class in question.

The probabilities used in the Cross-Entropy function, normally come from a Softmax activation function. This function converts the logits into probabilities. Equation 4.4 describes the Softmax function. The output of the Softmax function is always between 0 and 1 [32].

$$S(y)_i = \frac{exp(y_i)}{\sum_{j=1}^{n}(y_j)} \tag{4.4}$$

A simpler loss function is the Mean Square Error (MSE). The MSE is simply the average of the absolute difference between two vectors. One vector is the predicted value, while the other is the actual value. This function can be used for example to compare the pixels of two images and obtain some sense of their similarity [51]. There are various applications for the MSE function.

The mean square error function is defined in equation 4.5.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4.5}$$

4.6 Regularization Methods

There are plenty of regularization methods available. The goal of a regularization method is to restrain the models workflow and so obtain better results, mostly by avoiding overfitting.

L1 and L2 regularization are constraints added to the model to avoid overfitting. In equation 4.6 we have the L1 constraint, while in equation 4.7 we have the L2 constraint.

$$L1 = \lambda \sum_{j=0}^{M} |W_j| \tag{4.6}$$

$$L2 = \lambda \sum_{j=0}^{M} W_j^2 \tag{4.7}$$

The W argument in each of the restraints can be the activity of a certain model layer (meaning the output of that layer), the weights for a certain layer or part of the model or the bias of a layer in the model. In general, the L2 regularization leads to denser models, while the L1 regularization produces much sparser models [49].

The dropout layer is another regularization method available. In this layer a section of the outputs of a layer is randomly turned to 0 during training. This helps the model generalize better and avoid overfitting scenarios [67].

The total variation loss penalizes the derivative of the input. This loss can be applied to both one-dimensional signals and images. The loss acts as a filter that removes the Gaussian noise in the input. In our work we implement the two dimensional anisotropic version (Equation 4.8) . The ansiotropic version is easier to minimize, because it is differentiable [59].

$$V_{ansio}(y) = \sum_{i,j} |y_{i+1,j} - y_{i,j}| + |y_{i,j+1} - y_{i,j}| \tag{4.8}$$

Data augmentation methods consist in applying slight changes to the network input, in order to make a model more robust and generalize better. Example of this are very common in images, methods such as flipping, rotating, cropping, adding Gaussian noise among others. This slight variations create artificial data with a label similar to the original image [63].

Batch Normalization consists on normalizing the outputs of a layer during training. This help prevent the saturation of the network. Distribution of each layer's input change during training, because the previous layers change. This delays the training by requiring lower learning rates and careful parameter initialization. Makes it notoriously hard to train models with saturating nonlinearities [33].

This phenomenon is refered to as internal covariate shift, the solution is normalizing layer inputs. Batch Normalization allows us to use much higher learning rates and be less careful about initialization. It also acts as a regularizer, in some cases eliminating the need for Dropout [33].

4.7 Framework

The algorithm development in our work requires a computational tool to construct the necessary deep learning models, filter the signal and build the two-dimensional representations. Python provides multiple libraries useful both for deep learning and signal processing.

To develop the deep learning models necessary, our library of choice is Keras. Keras is built on top of TensorFlow2. This frameworks manages the GPU (Graphic Processing Unit) required to operate Keras and Deep Learning models. Figure 4.1 show the symbols related to these libraries.

(a) Keras

(b) Python

(c) Tensorflow

Figure 4.1: Framework Logos.

The SciPy library is responsible for most signal processing. Numpy, matplotlib and other crucial libraries are also implemented in our work.

4.8 Summary

Table 4.2 summarizes the results from the studies discussed in the State of the Art.

Table 4.2: State of the Art Results.

Study	Year	Feature Extraction	Classification	Database	Results
Hejazi et al. [30]	2016	AC/LDA	LDA	BioSec.Lab	EER - 34%
Luz et al. [22]	2017	Non-Fiducial	1D CNN, 2D CNN and Fusion Model	CYBHi, UofTDB	CYHBi (EER - 1D: 1.17 % to 15.60 %, 2D: 19.57 % to 26.38 %, Fusion: 1.33 % to 13.93 %) UofTDB (EER - 1D: 16.92 %, 2D: 19.37 %, Fusion: 14.27 %)
Min-Gu et al. [37]	2019	ECG Segments (fiducial and non-fiducial)	ConvNet1, ConvNet2, ConvNet3 and Ensemble Network	MIT-BIH Normal Sinus Rhythm Database (NSRDB)	Accuracy - Fiducial: ConvNet1 - 97.3%, ConvNet2 - 97.9%,ConvNet3 - 97.2%, Ensemble Networks - 98.9% Non-Fiducial: ConvNet1 - 97.2%, ConvNet2 - 96.6%,ConvNet3 - 97.7%, Ensemble Networks - 98.6%
Bento et al. [8]	2020	Non-fiducial	2D CNNs (simple CNN and DenseNet) with spectrogram input	Fantasia and ECG-ID database	(Accuracy- 99.42 %, a Specificity- 99.98 % and a Sensitivity- 99.42 %) ECG-ID (Accuracy - 94.23 %, Sensitivity - 94.26 %, Specificity - 99.94 %, within session), ECG-ID (Accuracy- 73.28 %, Sensitivity- 72.46 %, specificity- 99.70 %, across session)
Benouis et al. [7]	2021	Non-Fiducial	Autoencoder/Neural Network	ECG-ID, MITI-BIH, PTB, CYBHI	EER - 10% to 14% (ECG-ID, MITI-BIH, PTB) EER - 18% (CYBHi)
Lynn et al. [43]	2021	Fiducial and Non-Fiducial	1D-CNN, BLSTM and BGRU	ECG-ID, MIT-BIH, STAFF III and Long Term AF Database (LT-AF)	1D-CNN (acc) 3 heartbeats - 92,5%, 9 heartbeats - 91,1 %, BLSTM (acc) 3 heartbeats - 98,2 %, 9 heartbeats - 99,3 %, BGRU (acc) 3 heartbeats - 92,1 % , 9 heartbeats - 98,3 %
Chee et al. [14]	2022	Non-Fiducial	Transformer's self-attention Mechanism	ECG-ID, PTB	ECG-ID - (96.35% and 92.70%), PTB - (98.10% and 64.16%)
Ciocoiu et al. [17]	2019	Bandpass filter [2 - 30] Hz NCCC	2D/3D phase-space representation 2D CNN	UofTDB (20 subjects)	Sit: 93.6% (Rank-1), 97.2 % (Rank-2), All Postures: 90.4% (Rank-1), 95.8% (Rank-2)

In this assignment we use both the PTB and UofTDB. The PTB is an on-the-person database with clearer signals. The UofTDB is an off-the-person database with nosier signals, more difficult to identify. Table 4.3 summarizes the result for the PTB database.

Table 4.3: Results related to the PTB Database.

Study	Year	Feature Extraction	Classification	Result
Tantawi et al. [74]	2013	AC/DCT Auto correlation of the signal followed by a discrete cosine transform	Radial Based Function Network	Accuracy 97.7% (100 subjects)
Irvine et al [34]	2008	PCA use of eigen vectors (EigenPulse)	Distance in eigenspace	Accuracy 100% (43 subjects)
Zhao et al [85]	2011	Matching Pursuit	Support Vector Machine (SVM)	Accuracy 95.3% (20 subjects)
Labati et al [39]	2019	Deep CNN	Deep CNN	Accuracy 100% (52 subjects)
Benouis et al. [7]	2021	Non-Fiducial	Autoencoder/Neural Network	EER - 14%

The table 4.4 show the result for each distinct state of acquisition in the UofTDB database. With the exception of one study that evaluates the overall database.

Table 4.4: Results related to the UofTDB Database.

Study	Year	Feature Extraction	Classification	Result				
				Sit	Stand	Supine	Tripod	Exercise
Chan et al. [13]	2008	-	Wavelet distance measure (WDIST)	88%	94%	94%	96%	84%
Odinaka et al. [50]	2010	Use spectrogram for ECG	Compute Log Likelihood Ratio from the spectrogram features	66%	58%	84%	78%	90%
Irvine et al. [34]	2008	PCA use of eigen vectors (EigenPulse)	Distance in eigenspace	78%	88%	84%	82%	92%
Agrafioti et al. [2]	2008	Autocorrelation of the Signal	LDA	77%	75%	90%	94%	52%
Luz et al. [22]	2017	Non-Fiducial	1D CNN, 2D CNN and Fusion Model	EER-(1D:16.92%, 2D:19.37%, Fusion:14.27%)				

To further improve state of the art methods, we propose a new approach: a deep learning model with 1D to 2D transformation. An end-to-end model creates a 2-dimensional representation that optimizes the identification process. This transformation uses deep neural networks or more complex models.

Complex models such as GANs or autoencoders in ECG biometrics clean the ECG signal and optimize the overall process [7, 3]. Their application can extend to our 1D to 2D transformation.

Complex models offer a different perspective and alternative from the traditional methods. These methods can provide better results, although they require more processing power. The number of parameters in these models can also lead to overfitting. Complex models have advantages and disadvantages, but they´re a viable alternative worth exploring.

Chapter 5

Solution Overview

To improve the current state of the art, we propose an innovative method, a model with learnable 2D representations. There is a multitude of ways to do this. The process has two parts: 1D to 2D transition and classification.

In this project, first we studied a multitude of 2D electrocardiogram representations for classification. Afterwards, the representation that provides the best results incorporates the algorithm. A Multilayer Perceptron is trained to transform the one dimensional signal into a two dimensional representation. This Multilayer Perceptron is attached to the classification algorithm. The process is represented in Fig. 5.1.

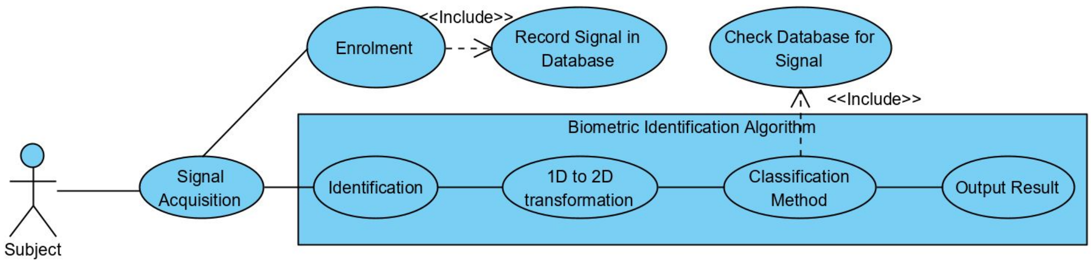


Figure 5.1: Use Case Diagram of the Biometric Identification Algorithm.

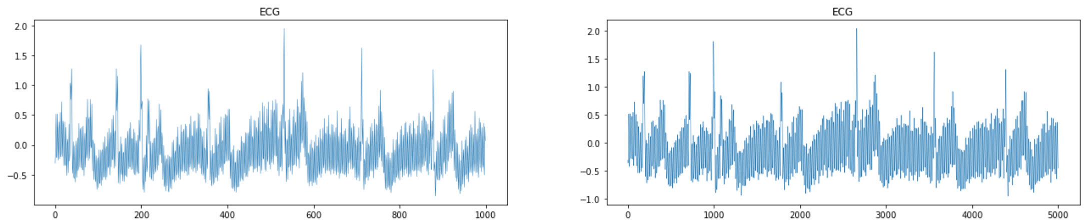
5.1 Signal Preprocessing

The first database used was the PTB. This database has on the person signals from 290 different subjects with an overall of 549 records. Each record is split into 5 second signal segments. The sampling frequency is 1000 Hz - so a 5 second signal has 5000 points.

Originally there are 11871 segments. These segments are divided into training, validation and test. First there is an 80-20 split to separate the testing samples. There are 7596 test segments.

Afterwards, there is a second 80-20 split to divide the training and validation subsets. The validation has 1900 samples, and the testing has 2375 samples. With data augmentation the training set doubles to 15192 segments.

The second database used was the UofTDB. This database has 1020 subjects and 1627 records. However, the records from subject 410 are all empty. So effectively there are only 1019 subjects. The sampling frequency of the UofTDB is 200 Hz. The signal is also segmented into 5 second segments. In this database the five second segments have 1000 points. The entire signal is up-sampled by a ratio of 5 to match the five thousand points segments obtained from the PTB database. Figure 5.2 shows the original signal and the up-sampled signal.



(a) ECG Segment from UofTDB.

(b) Up-sampled ECG Segment from UofTDB.

Figure 5.2: ECG segments from UofTDB.

For the UofTDB, after signal division, there are 40255 segments. These segments are divided with an 80-20 split to separate the test from the remainder of the ECG segments. There are 8051 test segments. The rest of the segments available are again divided in an 80-20 split. This gives us 25763 training segments and 6441 validation segments. With data augmentation, the training samples double to 51526.

The data augmentation method applied was random permutation. The signal is divided into three different parts and shuffled randomly. As a baseline, we used a uni-dimensional CNN based on Pinto et al. [58]. The study verified that random permutation as a data augmentation method leads to solid results and shows improvement compared to the original signal.

To verify these results, we defined our uni-dimensional neural network with the architecture in table 5.1.

Table 5.1: One Dimensional CNN Architecture.

Conv1D	Input: (5000, 1) Output: (5000, 24)
MaxPool	Input: (5000, 24) Output: (1000, 24)
Conv1D	Input: (1000, 24) Output: (1000, 24)
MaxPool	Input: (1000, 24) Output: (200, 24)
Conv1D	Input: (200, 24) Output: (200, 36)
MaxPool	Input: (200, 36) Output: (40, 36)
Conv1D	Input: (40, 36) Output: (40, 36)
MaxPool	Input: (40, 36) Output: (20, 36)
Conv1D	Input: (20, 36) Output: (20, 36)
Flat	Input: (20, 36) Output: 720
F.C.	Input: 720 Output: 100
F.C.	Input: 100 Output: 290 - Softmax

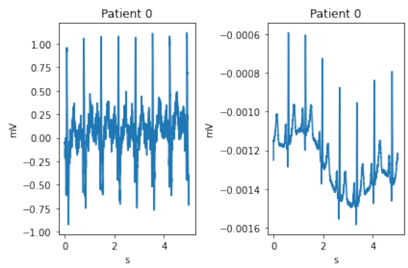
Each convolutional layer has a kernel with dimensions 5 by 1 and a stride of 1. The first layers have 24 filters while the last ones have 36. The max-pooling layers have a kernel of 5 by 1, except for the last layer where the kernel is 2 by 1. In the max pooling, the layers´ stride is the same as the pool size. The pool size is always 5, except for the last layers where it is 2.

After every convolutional layer, there´s a batch normalization layer to normalize the inputs. The model ends with two fully connected layers, the last one with 290 neurons. Every convolutional layer has a ReLU activation function. The first fully connected layer also has a ReLU activation function, while the last has a Softmax activation function.

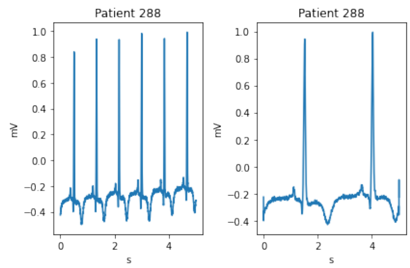
This CNN is a small simple model with 122,138 trainable parameters, that provides accurate results. This network was trained with various data augmentation methods:

1. Baseline Wandering- a sinusoidal wave of 1 Hz is added to the signal
2. Cropping- a small portion of the signal is oversampled to have the same number of points as the original signal
3. Flip- the data is flipped along the time axis
4. Gaussian Noise - random Gaussian noise added to the signal
5. Magnitude Scaling- the input is amplified by a factor of 1.3
6. Magnitude Warping- the input is multiplied by a sinusoidal wave with amplitude values close to one
7. Random Permutation- the signal is divided into 3 segments along the time axis, and the segments are joint randomly

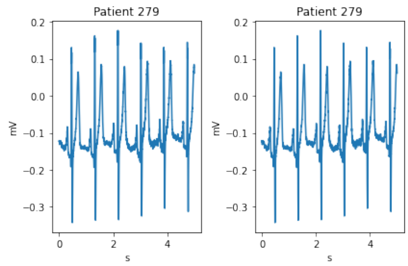
The multiple data augmentation methods are presented in Figure 5.3, besides the original signal.



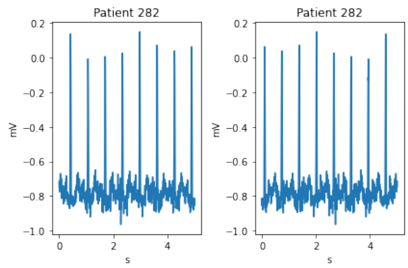
(a) ECG Segment with added Baseline Wander.



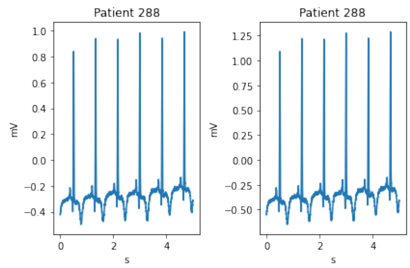
(b) ECG Segment Cropped.



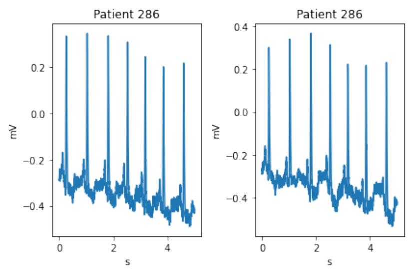
(c) ECG Segment with Gaussian Noise.



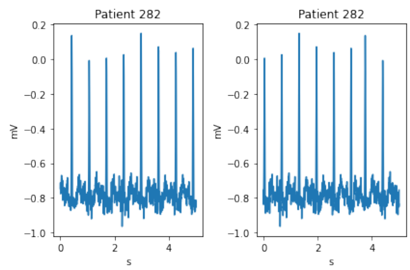
(d) ECG Segment Flipped.



(e) ECG Segment with Magnitude Scaling.



(f) ECG Segment with Magnitude Warping.



(g) ECG Segment with Random Permutation.

Figure 5.3: Data Augmentation of ECG Segments.

This model is trained using the Adam optimizer and the sparse categorical cross-entropy loss function. This function compares the predicted probability distribution with the targeted probability distribution. The comparison between multiple prediction values makes this lost function a good choice for models with various labels. Parameters such as the learning rate are fine-tuned to obtain the best result.

5.2 2D Representations

For the 2D representations, the signal is filtered. On the other hand, for the one-dimensional CNN the input signal is raw. In the one-dimensional approach, the noise functions almost as an intrinsic method of data augmentation. The slight variations between signals make the model more robust are reliable. However, in the two-dimensional approach, noisy signals have worse results, so a filtration step is implemented.

In the 2D representations, the signal is cleaned. First, a second-order IIR notch filter removes the 60 Hz frequency. This process reduces power-line interference. A corruption of the ECG signal, that occurs in the 50 to 60 Hz frequency band [40].

Secondly, a Butterworth high pass filter of order five applies a cutoff frequency of 3. This filter removes the baseline wander from the signal.

Table 5.2 represents the architecture of the CNN classifier for two dimensional approaches.

Table 5.2: Two dimensional CNN architecture

Conv2D	Input: (100, 100, 1) Output: (100, 100, 36)
Conv2D	Input: (100, 100, 36) Output: (100, 100, 36)
MaxPool2D	Input: (100, 100, 36) Output: (20, 20, 36)
Conv2D	Input: (20, 20, 36) Output: (200, 36)
MaxPool2D	Input: (20, 20, 36) Output: (4, 4, 36)
Flat	Input: (4, 4, 36) Output: 576
F.C.	Input: 576 Output: 100
F.C.	Input: 100 Output: 290 - Softmax

Figure 5.4 shows two different signals the original ECG segment present in the database and the ECG signal after it is filtered.

(a) ECG Segment from PTB database.

(b) Filtered ECG Segment from PTB database.

Figure 5.4: PTB Filtered and Unfiltered ECG Segment.

After every convolutional layer, there is a batch normalization layer to normalize the inputs. The model ends with two fully connected layers, the last one with 290 or 1019 neurons, depending on using the PTB or UofTDB dataset. After the two consecutive convolutional layers, there's a ReLU activation function. In the sequence of two convolutional layers, the first one has a kernel of size 5, while the following has a kernel of size 3. Each of the MaxPooling layers has a pool size of 2 and a stride of 5.

The first fully connected layer also has a ReLU activation function, while the last has a Softmax activation function. The input of this network varies regarding the 2D dimensional representation of the ECG. All the two-dimensional representations mentioned have an input size of 100 by 100, except for the spectrogram which has a size of 100 by 106. The spectrogram network has 126 302 trainable parameters for the PTB database and 199 643 for the UofTDB, while the rest of the two-dimensional models have 111 614 trainable parameters.

For the two-dimensional representations, random permutation is the data augmentation method of choice.

5.2.1 Markov Transition Field

The construction of a Markov Transition Field can be divided into multiple steps:

1. Signal Discretization
2. Build the Markov Transition Matrix
3. Calculate the Transition Probabilities
4. Compute the Markov Transition Field
5. Compute an Aggregated Transition Field

In signal discretization, the time-series is divided into multiple bins. The bins are determined by a range of values. The values within that range are classified as part of the bin in question.

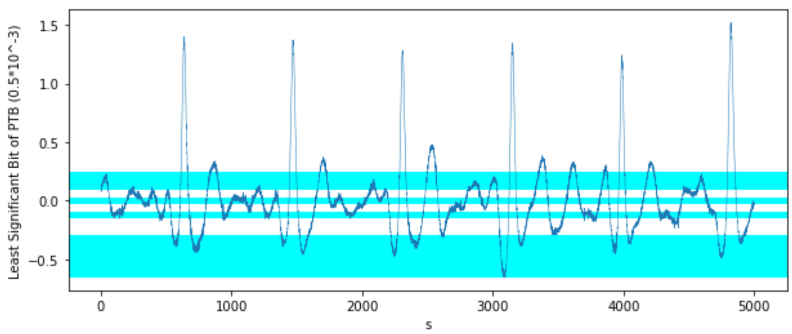


Figure 5.5: ECG Segment with Highlighted Bin Divisions.

To create our Markov Transition Matrix, the signal was first divided into 8 bin segments. In figure 5.5, the range of values related to each bin is clear. The white and blue background parts are related to the value range of each bin.

	Bins							
Bins	0	1	2	3	4	5	6	7
0	576	49	0	0	0	0	0	0
1	49	503	68	4	1	0	0	0
2	0	67	445	111	2	0	0	0
3	0	6	108	376	133	2	0	0
4	0	0	4	130	363	123	4	0
5	0	0	0	4	125	403	93	0
6	0	0	0	0	1	96	500	28
7	0	0	0	0	0	0	28	597

Figure 5.6: Markov Transition Matrix from ECG Segment.

The Markov Transition Matrix (Figure 5.6) is defined by the transitions between each bin. For example, in the first row and first column the number of points from bin 0 in the electrocardiogram - followed by points in the bin 0, as well, are quantified. The same happens for the remaining bins in columns and rows. In the fourth column and third row, there are 111 transitions from bin 2 to bin 3.

The transition probabilities are obtained from the Markov Transition Matrix. Transition probabilities correspond to the likelihood of a transitions between each bin. To calculate this, the number in each bin is divided by the total sum of the numbers in that column [41].

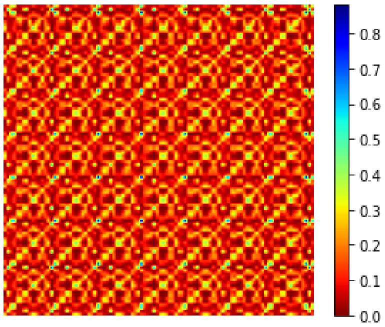
	Bins							
Bins	0	1	2	3	4	5	6	7
0	92.2	7.8	0	0	0	0	0	0
1	7.8	80.5	10.9	0.6	0.2	0	0	0
2	0	10.7	71.2	17.8	0.3	0	0	0
3	0	1	17.3	60.2	21.3	0.3	0	0
4	0	0	0.6	20.8	58.2	19.7	0.6	0
5	0	0	0	0.6	20	64.5	14.9	0
6	0	0	0	0	0.2	15.4	80	4.5
7	0	0	0	0	0	0	4.5	95.5

Figure 5.7: Markov Transition Probabilities Matrix from ECG Segment.

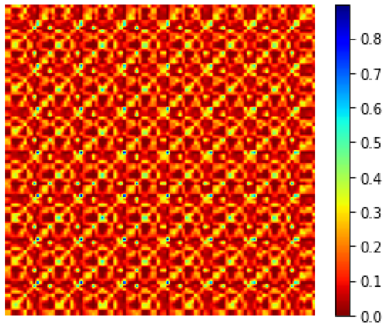
In the last step the Markov Transition Matrix is calculated. The result is a square matrix with a length and width similar to the size of the input signal.

In the matrix (Figure 5.7), the point from the first row and the second column corresponds to the probability of a transition between the bin of the first point and the bin of the second point. The same happens through out the rest of the matrix. Given that our signal has a length of 5000 points, the Markov Transition Matrix from our signal has a width of 5000 and a length of 5000.

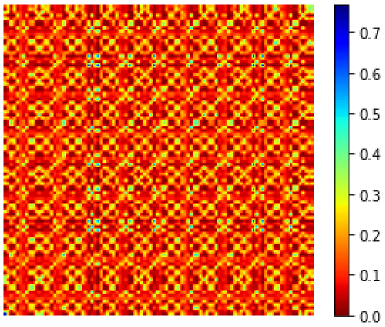
This matrix is to large to manage, so to simplify this process, we implement one last filter to aggregate the Markov Transition Field. This reduces the image to a manageable size. In our case we chose 100 by 100.



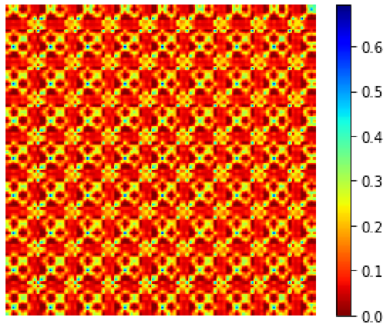
(a) Markov Transition Field from ECG Segment with Label 1.



(b) Markov Transition Field from ECG Segment with Label 1.



(c) Markov Transition Field from ECG Segment with Label 94.



(d) Markov Transition Field from ECG Segment with Label 94.

Figure 5.8: Example of Markov Transition Field Images from ECG Segments.

In figure 5.8, we have four examples of a Markov Transition Field obtained from an ECG segment. These examples are from different labels, a) and b) are from the patient with label 0, while c) and d) both have label 93. There is some difference visible from label to label. However, in general,the MTFs are somewhat similar.

5.2.2 Gramian Angular Field

Before exploring the Gramian Angular Matrix, two key concepts must be approached. The dot product of two vectors and the Gram Matrix.

The dot product is an application that measures the similarity between two vectors. In a two-dimensional space, the dot product can be characterized in two different ways.

$$< u,v >= u1 * v1 + u2 * v2 \tag{5.1}$$

$$< u,v >= \|u\| * \|v\| * cos(\sigma) \tag{5.2}$$

In 5.1, the coordinates of each vector are multiplied. While in 5.2 the norms of each vector are multiplied by the angle between them. If we consider that the norm of both vectors is equal to 1, the result of the dot product is only the angle of the two vectors.

The Gram Matrix is applied to a set of vectors. It is defined as the dot product of each vector by the remaining vectors in the set (Figure 5.9).

$$G = \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle & \dots & \langle v_1, v_n \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle & \dots & \langle v_2, v_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \langle v_n, v_2 \rangle & \dots & \langle v_n, v_n \rangle \end{pmatrix}$$

Figure 5.9: Gram Matrix Encoding Time Series as Images.

In case all the vectors in consideration have an unitary norm. The Gram matrix can be simplified by simply considering the angles between each vectors (Figure 5.10).

$$G = \begin{pmatrix} \cos(\phi_{1,1}) & cos(\phi_{1,2}) & \dots & cos(\phi_{1,n}) \\ \cos(\phi_{2,1}) & cos(\phi_{2,2}) & \dots & cos(\phi_{2,n}) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_{n,1}) & cos(\phi_{n,2}) & \dots & cos(\phi_{n,n}) \end{pmatrix}$$

Figure 5.10: Gram Matrix based on the Vector Angles Encoding Time Series as Images.

The Gramian Angular Field application can be characterized in three steps:

1. Scale the series to [-1, 1] with a Min-Max Scaller
2. Calculate the Polar Coordinates of the Scaled Time-Series
3. Create the Gramian Angular Matrix

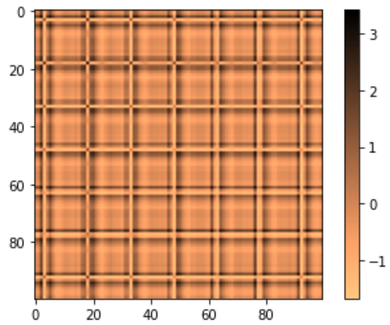
The time series needs to be scaled to [-1, 1], because a polar representation of a unitary vector always has a value between [-1, 1]. After the conversion to polar coordinates, we can compute the Gramian Angular Matrix with the angles obtained, by summing the angles of two vectors and using a cosine operation (Figure 5.11).

$$G = \begin{pmatrix} \cos(\phi_1 + \phi_1) & cos(\phi_1 + \phi_2) & \dots & cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & cos(\phi_2 + \phi_2) & \dots & cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & cos(\phi_n + \phi_2) & \dots & cos(\phi_n + \phi_n) \end{pmatrix}$$

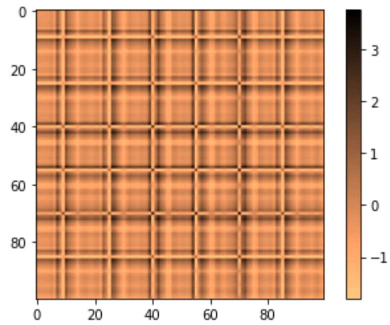
Figure 5.11: Gramian Angular Matrix using a Dot Product Alternative Encoding Time Series as Images.

The output of this operation is quite large. To obtain a smaller image, we use a bicubic interpolation method. As mentioned previously, our input has 5000 points. So the resulting output is a square matrix of 5000 by 5000. After the bicubic interpolation, the output has a size of 100 by 100.

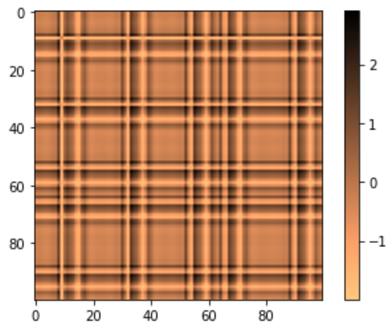
The bicubic interpolation method has a ratio of $\frac{1}{50}$. The a coefficient is $-\frac{1}{2}$. Bicubic interpolations is the preferred method in image processing, often preferred to bilinear interpolation, nearest neighbours interpolation, image resampling and others [15].



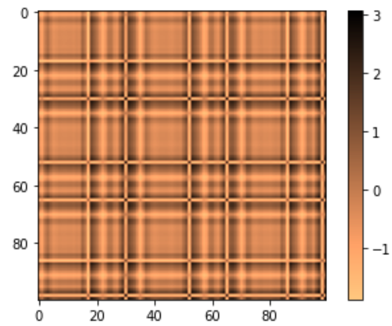
(a) Gramian Angular Field from ECG Segment with Label 6.



(b) Gramian Angular Field from ECG Segment with Label 6.



(c) Gramian Angular Field from ECG Segment with Label 51.



(d) Gramian Angular Field from ECG Segment with Label 51.

Figure 5.12: Examples of Gramian Angular Field Images from ECG Segments.

Figure 5.12 shows examples of the Gramian Angular Field. The four examples account for two different labels. The images with similar labels have certain points in common, such as the width of the darker lines in the GAF. In general, all GAFs are represented by squares of multiple sizes.

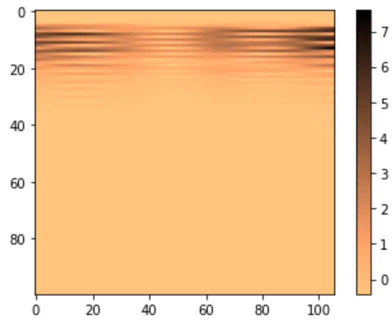
5.2.3 Spectrogram

The spectrogram is an application of the Fourier Transform in the signal by sequential windows. In our application, we use the Hanning window.

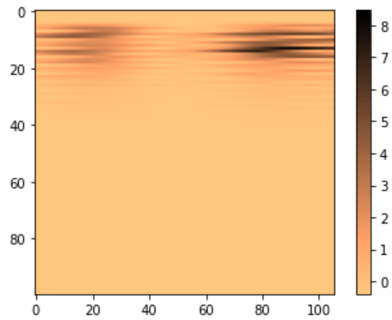
$$x(t) = \frac{1}{2} * [1 - cos(\frac{2\pi t}{T_H})], t \in [0, T_H] \tag{5.3}$$

Equation 5.3 describes the Hanning window. Given that $T_H = \frac{N_B}{f}$, where N_B is the length of the Hanning window and f is the signal frequency. The Hanning window reduces the excitation of less relevant frequencies, providing a smoother spectrum after applying a Fast Fourier Transform (FFT) [28]. The window used has a size of 2048 elements, and there is an overlap of 2020 elements between each window application. For our signal with 5000 points, this leads to 106 windows.

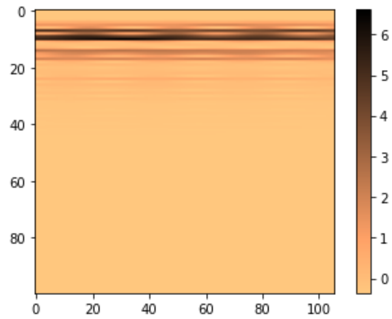
The FFT changes the signal from the time domain to the frequency domain [76]. The spectrogram is then clipped into an array of 100 by 106. This represents the frequencies from 0 Hz to 48,33 Hz - approximately 50 Hz. These frequencies are selected because they contain most of the information necessary to identify the person in question. The QRS segment and the remaining ECG waves are within this frequency bandwidth [60]. This more precise selection allows the model to focus on relevant information.



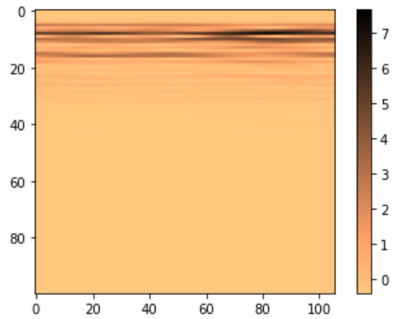
(a) Spectrogram from ECG Segment with Label 253.



(b) Spectrogram from ECG Segment with Label 253.



(c) Spectrogram from ECG Segment with Label 280.



(d) Spectrogram from ECG Segment with Label 280.

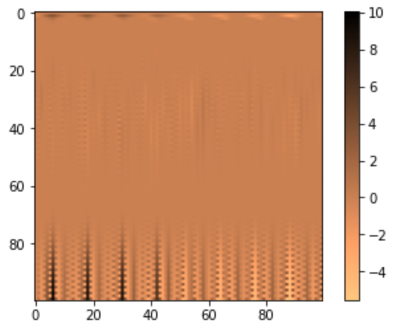
Figure 5.13: Examples of Spectrogram Images from ECG Segments.

Figure 5.13 shows four examples of the spectrogram. Periodically appearing blurry parts characterize the spectrogram in an image with a clear background. The pattern of these blurs is similar between spectrograms from the same subject. These blurry parts represent the higher intensity of the signal in certain frequencies at that time.

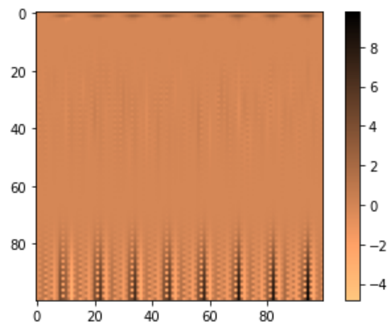
5.2.4 S-Transform

The S-Transform or Stockwell transform is a generalization of the continuous wavelet transform (CWT) with a scalable localizing Gaussian window. The S-transform provides a frequency-dependent solution from a signal in the time domain. It combines elements from the Fourier transform (FT) and the CWT [69].

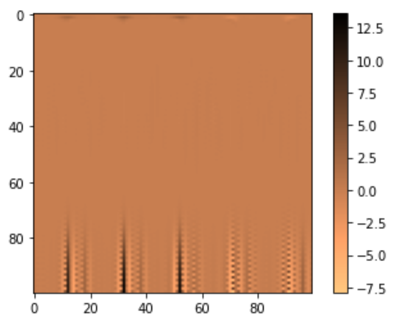
Our S-transform frequency bandwidth ranges from 0 to 50 Hz. The result is a matrix of size 5001 by 5000. To manage the resulting image more easily, we use bi cubic interpolation to resize the image into a 100 by 100 matrix. Figure 5.14 show examples of S-transforms.



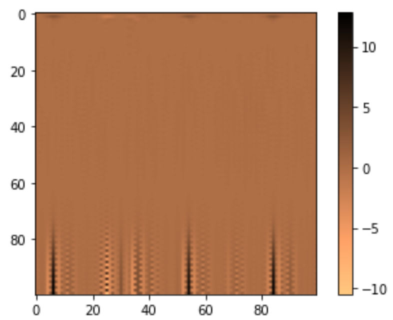
(a) S-Transform from ECG Segment with Label 93.



(b) S-Transform from ECG Segment with Label 93.



(c) S-Transform from ECG Segment with Label 256.



(d) S-Transform from ECG Segment with Label 256.

Figure 5.14: Examples of S-Transform Images from ECG Segments.

Figure 5.14 has four examples of S-transforms. Periodic and almost vertical lines characterize the S-transform. The lines represent a higher intensity for certain frequencies at that time. For different subjects, the number of lines changes. This is related to the heartbeat frequency of each subject, and the differences inherent to each subject's heartbeat.

5.3 Multilayer Perceptron

The multilayer perceptron (MLP) has two or three layers, (both architectures were explored). The first layers have a ReLU activation function, while the second one has a linear activation function. This is because the spectrogram has negative values that are cut off with the ReLU function. In table 5.3 bellow there's a representation of the MLP with two layers. At the table 5.4 the 3 layer architecture is revealed.

The Multilayer Perceptron is first pre-trained before it is introduced into the end-to-end model. During this training, the loss function is the mean square error between the output of the network

Table 5.3: Multilayer Perceptron Architecture two Layers.

FC	Input (5000) Output(8000)
FC	Input(8000) Output(10600)

Table 5.4: Multilayer Perceptron Architecture three Layers.

FC	Input (5000) Output(7000)
FC	Input(7000) Output(9500)
FC	Input(9500) Output(10600)

and the flat spectrogram itself. To help regulate this process, we experimented with activity constraints and the total variation loss. The activity constraint keeps the output values of the network from being too sparse. While the total variation penalty reduces the noise at the network output.

The Multilayer Perceptron is also trained with no restrictions, however this results in tremendous overfitting. The regularization in this process is an instrumental point. In addition to regularization, freezing certain parts of the overall model can be an advantage. The MLP with 2 layers has 124,834,600 trainable parameters, while the one with 3 layers has 202,260,100 trainable parameters.

5.4 End-to-End Model

The end-to-end model is the combination of the pre-trained MLP and the classifier. The classifier, in this case, is the CNN trained to identify the spectrogram. Within the spectrogram, both models with and without the dropout regulation are used.

To help regulate the model even further, in the end-to-end model the mean square error of the MLP output and the real spectrogram image is added to the overall loss function.

In addition to this, we experimented by freezing the classifier, and freezing the first layers of the MLP and the last layers of the classifier to regulate the model even further.

From the various combinations applied in this section, overfitting is the most pressing issue. The end-to-end model with a 3-layer MLP has 202,386,402 trainable parameters when used with the PTB database and 202,459,743 for the UofTDB. While the end-to-end model with a 2-layer MLP has 124,960,902 trainable parameters, and is only used in the PTB database.

5.5 Summary

The goal of our work is to build a two-dimensional learnable approach to ECG biometrics. The baseline here is the uni-dimensional signal. With hope of improving results, we use various two-dimensional approaches. From these multiple approaches, the best one is selected for the end-to-end model.

The UofTDB database is only used with the two-dimensional representation which provides the best results and the end-to-end model that provides the best results for the PTB.

The models within our project become increasingly more complex. This presents a problem - overfitting. To fight this and obtain the best results possible, there is a need for regularization. Both regularization and freezing parts of the model are viable approaches to try and stop this overfitting phenomenon.

Chapter 6

Results

6.1 One-dimensional Model

The methodology described in Chapter 5 is applied to two different databases. The PTB database and the UofTDB database. The PTB database has clearer signals that account for better results, while the UofTDB is an off-the-person database with a more convenient acquisition for the user.

In the table 6.1, we have the results from the one dimensional model with various data augmentation applications. These results are related to the PTB database.

Table 6.1: One dimensional CNN Data Augmentation Results (PTB)

Representation	Accuracy		
	Train	Validation	Test
Random Permutation	99.99%	99.68%	98.99%
Gaussian Noise	100%	98.47%	98.15%
Magnitude Scaling	100%	98.74%	97.94%
Normal	100%	98.58%	97.81%
Flip	100%	98.05%	97.39%
Base Line Wondering	100%	97.53%	96.46%
Magnitude Warping	100%	97.32%	96.08%
Cropping	99.99%	95.68%	94.65%

Of the various methods approached the random permutation stands out, so this method is applied for the rest of the project.

The UofTDB database was only tested with the uni-dimensional model, without any data augmentation, and with random permutation. Table 6.2 shows the results.

For the UofTDB, random permutation also show improvement compared to no data augmentation at all.

The random permutation method creates artificial signals that keep the features responsible for the identity of the subject intact. Nonetheless, it modifies the signal enough to provide robustness

Table 6.2: One dimensional CNN Data Augmentation Results (UofTDB).

Representation	Accuracy		
	Train	Validation	Test
Random Permutation	99.26%	87.12%	87.12%-
Normal	88.57%	61.62%	62.28%

to the model. There was some concern about whether this method could cause discontinuities in the signal, but the results show that it is the better option.

6.2 Two-dimensional Model

In the two-dimensional model, the PTB database is used for every two-dimensional representation. However, the UofTDB database is only used with the most promising representation related to the PTB, the spectrogram. The results for PTB from the two-dimensional representations are in Table 6.3:

Table 6.3: Two-dimensional PTB Results.

Representation	Accuracy		
	Train	Validation	Test
MTF	100%	75.2%	76.0%
GAF	100%	68.5%	67.6%
spectrogram	100%	96.7%	95.2%
spectrogram Dropout (0.2)	100%	97.7%	96.8%
S-Transform	100%	82.5%	80.3%

Given that the spectrogram obtained the best results from the two-dimensional representations, to further develop this approach we implement a dropout. This results in less overfitting and better results.

The UofTDB results for the two dimensional approach use the spectrogram and implement the spectrogram with a dropout. These are the better options within the PTB results. The results for the UofTDB are presented in Table 6.4.

Table 6.4: Two-dimensional UofTDB Results.

Representation	Accuracy		
	Train	Validation	Test
spectrogram	84.45%	62.01%	61.37%
spectrogram Dropout (0.2)	83.47%	66.57%	66.49%

The spectrogram with dropout implementation provides the best results. The dropout successfully reduces overfitting and helps generalize the model.

The dropout application in the two-dimensional scenario is key. The increasing number of trainable parameters within the two-dimensional model requires regularization. The dropout is a clear solution here.

6.3 End-to-End Model

Table 6.5 shows result from an end-to-end model trained without any regularization using the 2-layer MLP.

Table 6.5: End-to-End Model without Restrictions Results.

Train	Validation	Test
96.87%	41.63%	39.66%

The overfitting in this model is clear. The first step in the solution is to apply restrictions that fight this phenomenon.

The L1 activity regularization in the first layer (sum of the absolute value of the layer output) is multiplied by a lambda of 1^{-4}. The total variation loss is applied to the output of the last layer with a lambda of 1^{-6}. These restrictions are a starting point to stabilize the model. Table 6.6 shows the results.

Table 6.6: End-to-End Model with L1 and Total Variation Restrictions.

Train	Validation	Test
91.82%	55.95%	42.95%

The spectrogram has cyclical patterns, blurs, and the dynamic range is between -0.6 and 35. The spectrogram does not have any absurdly high values, so a sparsity constraint (L1 activity regularization) helps control the model. The total variation loss reduces noise from the MLP output, so it resembles the blurs in a spectrogram.

The changes implemented improved the results, nonetheless there is still room for improvement. The next step after this adjustment is to freeze the classifier, while maintaining the previous restrictions. The results are presented in Table 6.7.

Table 6.7: End-to-End Model with L1 and Total Variation Restrictions Classifier Frozen.

Train	Validation	Test
85.43%	43.58%	33.94%

Freezing the classifier didn't lead to better results. This restriction was too aggressive and prevented the model from developing a solution. To achieve a better regulation, a different part of the model was frozen. The first layers of the MLP, and the last layers of the classifier are frozen. This way the connection between the MLP and the classifier is trainable. The MLP in this situation has 3-layers, to better mimic the spectrogram representation.. The rest of the model is static to avoid overfitting. Table 6.8 shows the results.

Table 6.8: End-to-End Model with L1 and Total Variation Restrictions First Layer of the MLP and Last Layers of the Classifier Frozen.

Train	Validation	Test
71.16%	33.63%	33.39%

The increased regulation has a harsh effect on the model and the results are worse in the train, validation and test datasets. To achieve a better result, the mean square error between the output of the MLP and the actual spectrogram is added to the final loss of the model, with a lambda of 1. The model is not frozen, and the MLP has 2-layers. Table 6.9 shows the results.

Table 6.9: End-to-End Model with L1 and Total Variation Restrictions and MSE.

Train	Validation	Test
94.37%	55.21%	43.41%

This new restriction penalizes differences between the MLP representation and the original spectrogram. This force the MLP to produce a representation very similar to the original spectrogram itself.

The results became more promising. The next step is to use the 3-layer MLP, joining a L1 activity restriction in the last layer with a lambda of 1^{-4} and apply the total variation loss with a lambda of 1^{-6}. In addition to having the classifier frozen and adding the mean square error of the last layer of the spectrogram with a lambda of 2. The results are on Table 6.10.

Table 6.10: End-to-End Model with 3-layer MLP L1 and Total Variation Restrictions and 2*MSE.

Train	Validation	Test
99.42%	46.58%	45.73%

These various restrictions heavily regularize the end-to-end model. This is advantageous and certainly helped with overfitting, giving better results. In order to develop even further the results achieved, the classifier model added is now restrained by a dropout layer similar to the dropout explored in the spectrogram classification. The classifier is not frozen, however, it has a dropout layer to manage its performance and avoid overfitting. Table 6.11 has the results for this model.

Table 6.11: End-to-End Model with 3-layer MLP L1 and Total Variation Restrictions 2*MSE and Dropout 0.2 PTB Results.

Train	Validation	Test
99.69%	60.05%	56.08%

The dropout is more effective in regularization than freezing the classifier. With the dropout, the model remains flexible, while fighting overfitting by losing partial information. This provides the best generalization.

All of the previous models were only trained with the PTB database. This last model was also trained with the UofTDB. This allows an understanding of the effect various models with different dimension inputs have on off-the-person signals. Table 6.12 has the results from the UofTDB database.

Table 6.12: End-to-End Model with 3-layer MLP L1 and Total Variation Restrictions 2*MSE and Dropout 0.2 UofTDB Results.

Train	Validation	Test
94.60%	62.01%	32.49%

The baseline uni-dimensional models provide the best results for both databases. In the two-dimensional models, the PTB database achieves good results for the spectrogram representation. However, in the end-to-end model, there is still room for improvement. It is also clear that the identification in an off-the-person database such as the UofTDB is more challenging.

In future work, more constraints can be explored to regularize the model more efficiently. For example regularization terms such as the multiscale structural similarity index matrix (MS-SSIM) can help create a representation from the MLP equivalent to the spectrogram.

Chapter 7

Conclusions and Future Work

The initial part of the work, especially the work related to the two-dimensional model, was successful for the PTB database. With the PTB database, the results in both the uni-dimensional model and the spectrogram-based model are compatible with the results found in state-of-the-art performance. The end-to-end model, on the other hand was more complicated.

As for the UofTDB database, biometric identification is much more challenging. Similar models have worse results with the UofTDB dataset. This off-the-person representation is the most valuable for biometrics, given that its acquisition method is far more convenient for the user.

The problem with the end-to-end model in both the PTB and UofTDB databases is overfitting. The models are increasingly complex, and overfitting becomes a problem. Solutions for this problem are increasing the dataset size with different databases, experimenting with different architectures, or using different regularization methods.

One out of the box idea might be using an MLP to create a short-term Fourier transform STFT. The various STFTs are joined together to make a spectrogram. Using the various STFTs produced by the MLP, we create the spectrogram that is classified afterward. This maintains the original idea of an end-to-end model and gives a simpler task to the MLP.

Better and new regularization is also necessary. The use of other methods applied in image similarity, for example the multiscale structural similarity index matrix (MS-SSIM) can go a long way in improving the end-to-end model.

Exploring different architectures for the 1D to 2D transformation is also an option. Perhaps a model such as a CNN or a GAN could more accurately mimic the spectrogram in question and obtain better results overall.

References

[1] M. Abo-Zahhad, Sabah M. Ahmed, and S. N. Abbas. Biometric authentication based on PCG and ECG signals: present status and future directions. *SIViP*, 8(4):739–751, 2014.

[2] Foteini Agrafioti and Dimitrios Hatzinakos. Ecg based recognition using second order statistics. In *6th Annual Communication Networks and Services Research Conference (cnsr 2008)*, pages 82–87, 2008.

[3] Karol Antczak. A generative adversarial approach to ECG synthesis and denoising. *CoRR*, abs/2009.02700, 2020.

[4] T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster. Multi-channel spectrograms for speech processing applications using deep learning methods. *Pattern Anal Applic*, 24(2), 2021.

[5] Mohamed Bekkar and Dr Hassiba Kheliouane Djemaa. Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, page 13, 2013.

[6] David Belo, Nuno Bento, Hugo Plácido da Silva, Ana Fred, and Hugo Gamboa. Ecg biometrics using rnn and cnn. 01 2020.

[7] Mohamed Benouis, Meriem Reguide, Alfredo Rosado-Munoz, and Lotfi Mostefai. 2d ecg image based biometric identification using stacked autoencoders. In *2021 IEEE International Workshop on Metrology for Industry 4.0 IoT (MetroInd4.0IoT)*, pages 286–289, 2021.

[8] Nuno Bento, David Belo, and Hugo Gamboa. Ecg biometrics using spectrograms and deep neural networks. 2020.

[9] E Besterman and R Creese. Waller–pioneer of electrocardiography. 42(1):61–64.

[10] Ralf-Dieter Bousseljot, D Kreiseler, and A Schnabel. The PTB diagnostic ECG database. Type: dataset.

[11] GE Burch. History of precordial leads in electrocardiography. *European journal of cardiology*, 8(2):207—236, September 1978.

[12] Christian Cajavilca and Joseph Varon. Willem einthoven: The development of the human electrocardiogram. *Resuscitation*, 76:325–8, 04 2008.

[13] Adrian D. C. Chan, Mohyeldin M. Hamdy, Armin Badre, and Vesal Badee. Wavelet distance measure for person identification using electrocardiograms. *IEEE Transactions on Instrumentation and Measurement*, 2008.

[14] Kai Jye Chee and Dzati Athiar Ramli. Electrocardiogram biometrics using transformer's self-attention mechanism for sequence pair feature extractor and flexible enrollment scope identification. *Sensors*, 22(9), 2022.

[15] Gyu-Ho Choi, Eun-Sang Bak, and Sung-Bum Pan. User identification system using 2d resized spectrogram features of ecg. *IEEE Access*, 7:34862–34873, 2019.

[16] Ivaylo Christov, Tatyana Neycheva, Ramun Schmid, Todor Stoyanov, and Roger Abächerli. Pseudo-real-time low-pass filter in ECG, self-adjustable to the frequency spectra of the waves. 55(9).

[17] Iulian B. Ciocoiu and Nicolae Cleju. Off-the-person ecg biometrics using convolutional neural networks. In *2019 International Symposium on Signals, Circuits and Systems (ISSCS)*, pages 1–4, 2019.

[18] Iulian B. Ciocoiu and Nicolae Cleju. Off-person ecg biometrics using spatial representations and convolutional neural networks. *IEEE Access*, 8:218966–218981, 2020.

[19] Adrian Condon and Grace Willatt. Ecg biometrics: the heart of data-driven disruption? *Biometric Technology Today*, 2018(1):7–9, 2018.

[20] Hugo Plácido da Silva, Carlos Carreiras, André Lourenço, Ana Fred, Rui César das Neves, and Rui Ferreira. Off-the-person electrocardiography: performance assessment and clinical correlation. 4(4):309–318.

[21] Hugo Plácido da Silva, André Lourenço, Ana Fred, Nuno Raposo, and Marta Aires-de Sousa. Check your biosignals here: A new dataset for off-the-person ECG biometrics. 113(2):503–514.

[22] Eduardo José da Silva Luz, Gladston J. P. Moreira, Luiz S. Oliveira, William Robson Schwartz, and David Menotti. Learning deep off-the-person heart biometrics representations. *IEEE Transactions on Information Forensics and Security*, 2018.

[23] Shubhajit Datta. A review on convolutional neural networks. In Rabindranath Bera, Prashant Chandra Pradhan, Chuan-Ming Liu, Sourav Dhar, and Samarendra Nath Sur, editors, *Advances in Communication, Devices and Networking*, pages 445–452, Singapore, 2020. Springer Singapore.

[24] Simon Eberz, Nicola Paoletti, Marc Roeschlin, Andrea Patani, Marta Kwiatkowska, and Ivan Martinovic. Broken hearted: How to attack ECG biometrics. In *Proceedings 2017 Network and Distributed System Security Symposium*. Internet Society.

[25] W. Einthoven. The telecardiogram. *American Heart Journal*, 53(4):602–615, 1957.

[26] Chi Nok Enoch Kan, Richard J Povinelli, and Dong Hye Ye. Enhancing multi-channel eeg classification with gramian temporal generative adversarial networks. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1260–1264, 2021.

[27] Nesli Erdogmus and Sébastien Marcel. Spoofing face recognition with 3d masks. *IEEE Transactions on Information Forensics and Security*, 9(7):1084–1097, 2014.

[28] Victor Giurgiutiu. Chapter 12 - wave propagation shm with pwas transducers. In Victor Giurgiutiu, editor, *Structural Health Monitoring with Piezoelectric Wafer Active Sensors (Second Edition)*. Academic Press, Oxford, second edition edition, 2014.

[29] Varun Gupta, Monika Mittal, and Vikas Mittal. Performance evaluation of various pre-processing techniques for r-peak detection in ecg signal. *Iete Journal of Research*, pages 1–16, 2020.

[30] Maryamsadat Hejazi, S. A. R. Al-Haddad, Yashwant Prasad Singh, Shaiful Jahari Hashim, and Ahmad Fazli Abdul Aziz. ECG biometric authentication based on non-fiducial approach using kernel methods. 52:72–86.

[31] Shrishail M. Hiremath, Sambit Behura, Subham Kedia, Siddharth Deshmukh, and Sarat Kumar Patra. Deep learning-based modulation classification using time and stockwell domain channeling. ISBN: 9781538692868.

[32] Yaoshiang Ho and Samuel Wookey. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access*, 8, 2020.

[33] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[34] John M. Irvine, Steven A. Israel, W. Todd Scruggs, and William J. Worek. eigenPulse: Robust human identification from cardiovascular function. 41.

[35] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. 14(1):4–20.

[36] Benjamin E Jin, Heike Wulff, Jonathan H Widdicombe, Jie Zheng, Donald M Bers, and Jose L Puglisi. A simple device to illustrate the einthoven triangle. 36(4). Publisher: American Physiological Society.

[37] Min-Gu Kim, Hoon Ko, and Sung Bum Pan. A study on user recognition using 2d ECG image based on ensemble networks for intelligent vehicles. 2019:6458719. Publisher: Hindawi.

[38] Eric P. Kukula, Mathias J. Sutton, and Stephen J. Elliott. The human–biometric-sensor interaction evaluation method: Biometric performance and usability measurements. *IEEE Transactions on Instrumentation and Measurement*, 59(4):784–791, 2010.

[39] Ruggero Donida Labati, Enrique Muñoz, Vincenzo Piuri, Roberto Sassi, and Fabio Scotti. Deep-ecg: Convolutional neural networks for ecg biometric recognition. *Pattern Recognition Letters*, 126:78–85, 2019.

[40] Chavdar Levkov, Georgy Mihov, Ratcho Ivanov, Ivan Daskalov, Ivaylo Christov, and Ivan Dotsinsky. Removal of power-line interference from the ECG: a review of the subtraction procedure.

[41] Lu Liu and Zhiguang Wang. Encoding temporal markov dynamics in graph for time series visualization. *CoRR*, abs/1610.07273, 2016.

[42] Tatiana Lugovaya. The ECG-ID database. Type: dataset.

[43] Htet Myet Lynn, Pankoo Kim, and Sung Bum Pan. Data independent acquisition based bi-directional deep networks for biometric ecg authentication. *Applied Sciences*, 11(3), 2021.

[44] John Madias. On recording the unipolar ecg limb leads via the wilson's vs the goldberger's terminals: avr, avl, and avf revisited. *Indian pacing and electrophysiology journal*, 8:292–7, 02 2008.

[45] Shie Mannor, Dori Peleg, and Reuven Rubinstein. The cross entropy method for classification. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, New York, NY, USA, 2005. Association for Computing Machinery.

[46] Juan Pablo Martínez, Olle Pahlm, Michael Ringborn, Stafford Warren, Pablo Laguna, and Leif Sörnmo. The staff iii database: Ecgs recorded during acutely induced myocardial ischemia. In *2017 Computing in Cardiology (CinC)*, 2017.

[47] Mario Merone, Paolo Soda, Mario Sansone, and Carlo Sansone. Ecg databases for biometric systems: A systematic review. *Expert Systems with Applications*, 2017.

[48] G.B. Moody and R.G. Mark. The impact of the MIT-BIH arrhythmia database. 20(3).

[49] Robert C. Moore and John DeNero. L1 and l2 regularization for multiclass hinge loss models. In *Symposium on Machine Learning in Speech and Natural Language Processing*, 2011.

[50] Ikenna Odinaka, Po-Hsiang Lai, Alan D. Kaplan, Joseph A. O'Sullivan, Erik J. Sirevaag, Sean D. Kristjansson, Amanda K. Sheffield, and John W. Rohrbaugh. Ecg biometrics: A robust short-time frequency analysis. In *2010 IEEE International Workshop on Information Forensics and Security*, pages 1–6, 2010.

[51] Jakub Pęksiński and Grzegorz Mikołajczak. *The Synchronization of the Images Based on Normalized Mean Square Error Algorithm*, pages 15–25. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[52] Mariusz Pelc, Yuriy Khoma, and Volodymyr Khoma. Ecg signal as robust and reliable biometric marker: Datasets and algorithms comparison. *Sensors*, 19(10), 2019.

[53] Simona Petrutiu, Alan V. Sahakian, and Steven Swiryn. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *EP Europace*, 9(7), 05 2007.

[54] C. Robert Pinnegar and Lalu Mansinha. The s-transform with windows of arbitrary and varying shape. *Geophysics*, 68:381–385, 2003.

[55] Hugo Plácido da Silva, Andre Lourenco, Filipe Canento, Ana Fred, and Nuno Raposo. Ecg biometrics: Principles and applications. 01 2013.

[56] Aditya Singh Rathore, Zhengxiong Li, Weijin Zhu, Zhanpeng Jin, and Wenyao Xu. A survey on heart biometrics. 53(6):1–38.

[57] João Ribeiro Pinto and Jaime Cardoso. *An End-to-End Convolutional Neural Network for ECG-Based Biometric Authentication*.

[58] João Ribeiro Pinto, Jaime Cardoso, and Andre Lourenco. *Deep Neural Networks For Biometric Identification Based On Non-Intrusive ECG Acquisitions*. 11 2019.

[59] Paul Rodríguez. Total variation regularization algorithms for images corrupted with different noise models: A review. 2013. Publisher: Hindawi Publishing Corporation.

[60] Norhashimah Mohd Saad, Abdul Rahim Abdullah, and Yin Fen Low. Detection of heart blocks in ecg signals by spectrum and time-frequency analysis. In *2006 4th Student Conference on Research and Development*, 2006.

[61] João Loures Salinet and Olavo Luppi Silva. Chapter 2 - ecg signal acquisition systems. In João Paulo do Vale Madeiro, Paulo César Cortez, José Maria da Silva Monteiro Filho, and Angelo Roncalli Alencar Brayner, editors, *Developments and Applications for ECG Signal Processing*, pages 29–51. Academic Press, 2019.

[62] Haotian She, Jinying Zhu, Ye Tian, Yanchao Wang, Hiroshi Yokoi, and Qiang Huang. SEMG feature extraction based on stockwell transform improves hand movement recognition accuracy. 19(20):4457.

[63] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

[64] Hugo Plácido da Silva, André Lourenço, Ana Fred, Nuno Raposo, and Marta Aires-de Sousa. Check your biosignals here initiative (CYBHi) dataset for off-the-person electrocardiography (ECG) biometrics. Type: dataset.

[65] Mohsen Soleimani, Aram Vahidi, and Behrouz Vaseghi. Two-dimensional stockwell transform and deep convolutional neural network for multi-class diagnosis of pathological brain. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, PP:1–1, 11 2020.

[66] Janani C. Sriram, Minho Shin, Tanzeem Choudhury, and David Kotz. Activity-aware ECG-based patient authentication for remote health monitoring. In *Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI '09*, page 297. ACM Press.

[67] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. page 30.

[68] Barold SS. Willem einthoven and the birth of clinical electrocardiography a hundred years ago. *Card Electrophysiol Rev*, 2003-01.

[69] R.G. Stockwell, L. Mansinha, and R.P. Lowe. Localization of the complex spectrum: the s transform. *IEEE Transactions on Signal Processing*, 1996.

[70] R.G. Stockwell, Lalu Mansinha, and Robert Lowe. Localization of the complex spectrum: the s transform. ieee trans signal process 44: 998-1001. *Signal Processing, IEEE Transactions on*, 44:998 – 1001, 05 1996.

[71] Balambigai Subramanian. ECG signal classification and parameter estimation using multiwavelet transform. 28(7):7.

[72] Wei Sun, Jie Zhou, Bintao Sun, Yuqing Zhou, and Yongying Jiang. Markov transition field enhanced deep domain adaptation network for milling tool condition monitoring. 13(6):873.

[73] A. H. Sykes. A. d. waller and the electrocardiogram, 1887. *British Medical Journal (Clinical Research Edition)*, 294(6584):1396–1398, 1987.

[74] Manal M. Tantawi, Kenneth Revett, Abdel-Badeeh Salem, and Mohamed F. Tolba. A wavelet feature extraction method for electrocardiogram (ECG)-based biometric recognition.

[75] Anthony Ngozichukwuka Uwaechia and Dzati Athiar Ramli. A comprehensive survey on ecg signals as new biometric modality for human authentication: Recent advances and future challenges. *IEEE Access*, 9:97760–97802, 2021.

[76] Wim van Drongelen. Chapter 6 - continuous, discrete, and fast fourier transform. In Wim van Drongelen, editor, *Signal Processing for Neuroscientists (Second Edition)*. Academic Press, second edition edition, 2018.

[77] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. 7(1):154.

[78] Saeid Wahabi, Shahrzad Pouryayevali, Siddarth Hari, and Dimitrios Hatzinakos. On evaluating ECG biometric systems: Session-dependence and body posture. 9(11). Conference Name: IEEE Transactions on Information Forensics and Security.

[79] Zhiguang Wang and Tim Oates. Imaging time-series to improve classification and imputation. *CoRR*, abs/1506.00327, 2015.

[80] Sajila D. Wickramaratne and Md Shaad Mahmud. A deep learning based ternary task classification system using gramian angular summation field in fnirs neuroimaging data. In *2020 IEEE International Conference on E-health Networking, Application Services (HEALTHCOM)*, pages 1–4, 2021.

[81] Lukasz Wieclaw, Yuriy Khoma, Pawel Fałat, Dmytro Sabodashko, and Veronika 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)*, volume 1, pages 129–133, 2017.

[82] Hongji Xu, Juan Li, Hui Yuan, Qiang Liu, Shidi Fan, Tiankuo Li, and Xiaojie Sun. Human activity recognition based on gramian angular field and deep convolutional neural network. *IEEE Access*, 8, 2020.

[83] Jialin Yan, Jiangming Kan, and Haifeng Luo. Rolling bearing fault diagnosis based on markov transition field and residual network. *Sensors*, 22(10), 2022.

[84] Xiaoting Zhao, Hongbin Sun, Bo Lin, Hongmin Zhao, Yingli Niu, Xiang Zhong, Yidan Wang, Yiming Zhao, Fanchao Meng, Jinmin Ding, Xinghua Zhang, Liang Dong, and Sheng Liang. Markov Transition Fields and Deep Learning-Based Event-Classification and Vibration-Frequency Measurement for ϕ-OTDR. *IEEE Sensors Journal*, 22(4):3348–3357, 2022.

[85] Zhidong Zhao and Lei Yang. Ecg identification based on matching pursuit. volume 2, 2011.

[86] Huailai Zhou, Yaming Tian, and Yan Ye. Dynamic deconvolution of seismic data based on generalized s-transform. 108:1–11.