Ministry of Higher Education
and Scientific Research
University of Diyala
College of Science
Department of computer science

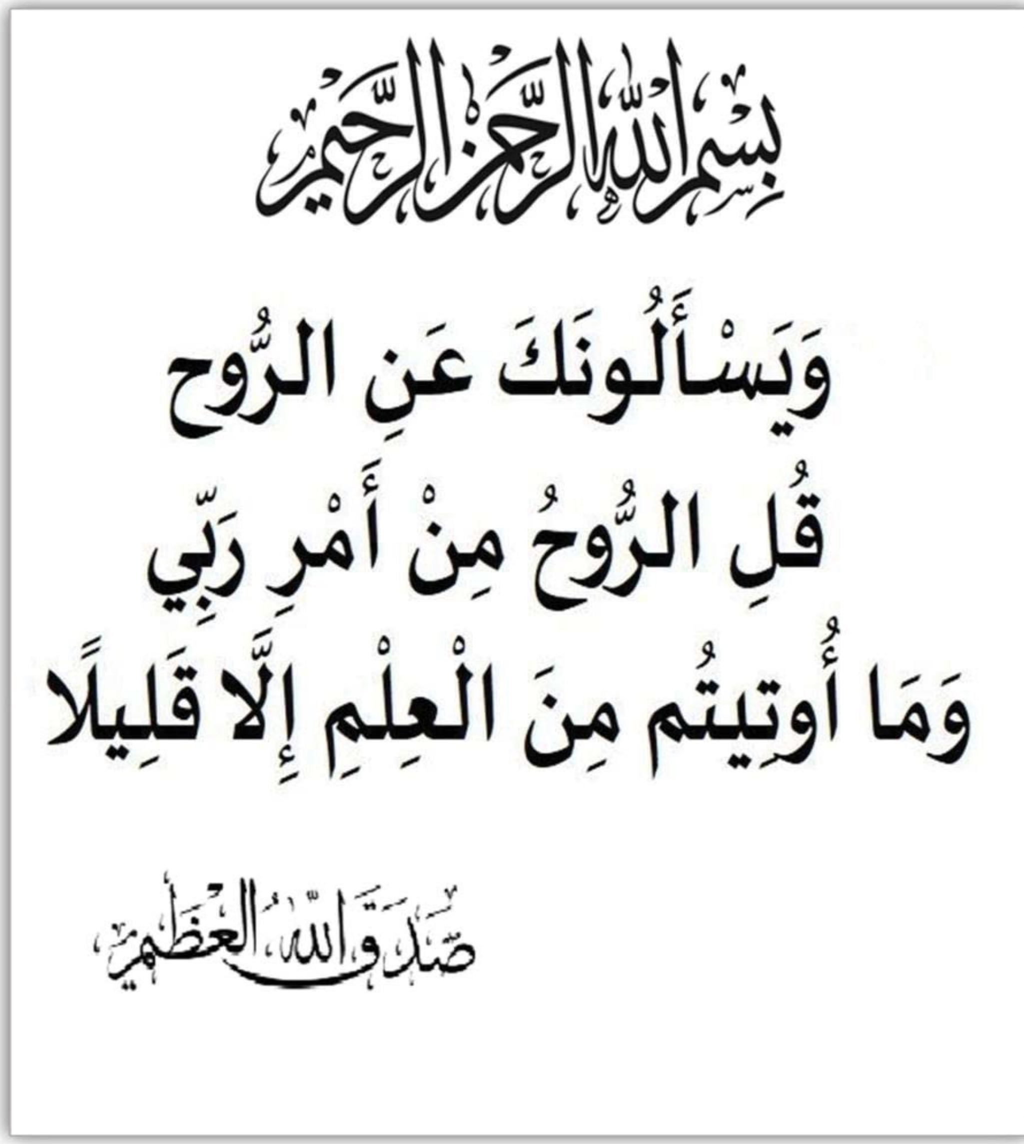
Electrocardiogram For Human Identification Based On Machine Learning Techniques

A Thesis
Submitted to the Department of Computer Science\ College of Science\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By
Nada Mahmood Hameed

Supervised By
Asst. Prof. Dr. Jamal Mustafa Abbas Al-Tuwaijari

2022 A.C. **1444 A.H.**



سورة الاسراء: (آية: ٨٥)

Linguistic Certification

This is to certify that this thesis entitled “[illegible]” was prepared under my linguistic supervision. It was amended to meet the style of English language.

Signature :

Name [illegible]

Date : [illegible] 2022

Supervisor's Certification

We certify that this thesis entitled "[illegible]" was prepared [illegible] my supervision at the University of Diyala Faculty of Science Department of Computer Science, as a partial fulfillment of the requirements needed to award the degree of ***Master of Science in Computer Science***.

(Supervisor)

Signature :

Name [illegible]

Date : [illegible] 2022

Approved by University of a Diyala Faculty of Science Department of Computer Science

Signature :

Name : [illegible]

Date : [illegible] 2022

Examination Committee Certification

We certify that we have read the thesis entitled "[illegible]*" and as an examination committee, examined the student "*[illegible]*" in the thesis content and that in our opinion, it is adequate as fulfill the requirement for the Degree of Master in Computer Science at the Computer Science Department, University of Diyala.*

[illegible]

[illegible]

[illegible]

[illegible]

Approved by the Dean of College Science, University of Diyala

[illegible]

Scientific Amendment

*I certify that the thesis entitled "***Electrocardiogram for Human Identification Based On Machine Learning Techniques***" was prepared by* ***Nada Mahmood Hameed*** *has been evaluated scientifically; therefore, it is suitable for debate by the examining committee.*

Signature:

Name: [illegible]

Date: / / 2022

Signature:

Name: [illegible]

Date: / / 2022

Dedication

Firstly, praise be to God, Lord of all creation, for every blessing, whether or not it is, praise be to God who helped me complete this research to its end. I would like to express my thanks and appreciation to my supervisor, Dr. Jamal Mustafa Al-Tuwaijari, for his supervision of this research, and for his generosity, patience, and continuous guidance throughout the work. I have had the good fortune to get advice and guidance from him. My thanks to the academic and administrative staff in the Department of the Computer Science / University of Diyala. I would like to express my gratitude to my dear father, my beloved mother and the spirit of my husband who accompanied me at every stage, and there are not enough words to thank my dear sister and my second mother for their support and faith in me all the time and for their encouragement During my studies and my dear children. Not to mention the dear sister and colleague who helped me and stood by me since the beginning of my research journey until this day, all thanks and appreciation to everyone who stood by my side and extended a helping hand, morally or scientifically, without the help of God Almighty first and the help of those we love and love, this achievement would not have been possible

Acknowledgment

Initially, I extend my sincere thanks to "*Allah*" who does not abandon me and helped me in the completion of this work.

Many thanks are to my supervisor “Asst. Prof. Dr. Jamal Mustafa Abbas Al-Tuwaijari ", for his contribution, support, and discussions, which help me a lot during the research period.

Though only my name appears on the cover of this thesis, many people have contributed to its production.

I will never forget to thank the current staff of Department of Computer Science\ College of Science\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science, for their various forms of support during my master’s study. On top of them is the head of the department.

Abstract

Electrocardiogram (ECG) data are characteristically biomarkers that can be collected despite the limitations of time and space. The most important of its benefits are, it has been widely used in security and medical fields. Electrocardiogram, also known as an (ECG) is a recording of electrical activity of the heart and depicts the periodic patterns of the human heartbeats, therefore it is regard as the most reliable method for determining a person's identification. Due to the nature of an ECG's built-in capability of determining whether a person is still alive or not, it is possible to obtain a biometric sample from a genuine person who is still alive.

In this thesis, a biosafety system for person identification by using ECG based on two internal models is presented. The first one of which supports deep learning technique through the use of one neural network algorithm (CNN algorithm) and the other model supports machine learning technique (ML) through the use of three algorithms (decision tree, support vector machine, and navy base algorithms). The aim is to design a robust and realistic security system that has been functioned by using a certified global dataset, (the Physikalisch-Technische Bundesanstalt (PTB) dataset), and moreover to prove the importance of using these two techniques to determine the identity of a person based on the electrocardiogram data.

The results achieved by the proposed system were accuracy of 99.98% for the CNN algorithm the of deep learning model and an accuracy of (96% for the decision tree, 100% for the support vector machine, and 92% for the navy base) of machine learning model.

List of Contents

[illegible]

List of Tables

List of Algorithms

List of Abbreviations

Abbreviations	**Meaning**
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CNN	Convolutional Neural Network
CSP	Common Spatial Pattern
DL	Deep learning
ECG	Electrocardiogram
FFNN	Feed-Forward Neural Network verifies individuals
FN	False Negative
FP	False Positive
HPF	High Pass Filter
LPF	low Pass Filter
LSTM	Long Short-Term Memory Networks
ML	Machine Learning
PTB	Physikalisch-Technische Bundesanstalt
RBM	Restricted Boltzmann Machines
RELU	Rectified Linear Unit
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

Chapter One

General Introduction

1.1 Introduction

Due to the general rapid advancement of Artificial Intelligence technology, everyone is concerned about their data's safety. Many cases need the use of individuals' unique data as a "key" to their personal information. Fingerprints, face scans, and iris scans are some of the most widely used human identification technologies [1].

An electrocardiogram (ECG) is a physiologic signal that records the cardiac activity over time (ECG). It is used to detect heart disease and biometric security systems [2]. People with varied electrocardiograms (ECGs) may experience distinct P waves, QRS complexes, and T waves, but they are all related. However, the degree of difficulty may vary based on the individual 's ages, ethnicity, height, and weight, as well as their lifestyle. Athletes' QRS complexes are larger than the normal person's because of their stronger left ventricle [3]. Individuals' ECG morphology differs due to their lack of differentiating traits, allowing us to use ECG to identify individuals [4].

Electrocardiograms have unique qualities that aid in precisely identifying humans, is a fertile research ground, and are continually evolving[5]. This chapter deals with an overview of the topic of ECG and what are the problems of the system and a review of the most important previous studies of researchers.

1.2 Background of Electrocardiogram (ECG)

Electric cardiac currents were first recorded more than a century ago by scientists. Willem Einthoven, a Dutch scientist, was awarded the Nobel Prize in physiology or medicine in 1924 for being the first to invent the electrocardiogram as we know it today. This accomplishment came at the beginning of the twentieth century[3].

There are several complimentary medical tests that can be conducted utilizing electrocardiograms (ECGs) to look for and analyze cardiac abnormalities from both a medical as well as security perspective. The ECG measures variations in electrical potential between two sites throughout the heart's electrical activity [6].

Explains the physiological basis of ECG and what are its advantages in Figure (1.1) (ECG) of a healthy individual:

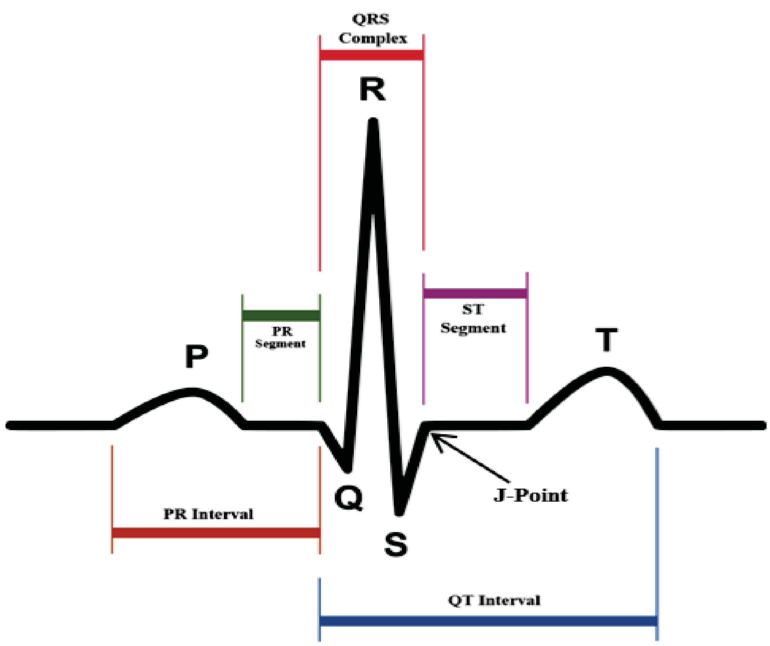


Figure (1.1): ECG of a healthy individual [8,9]

Based on the above Figure, the essential features of the cardiogram can be summarized in the following points [10, 11]:

i. □□wave an early inversion of the positive component of the QRS complex
ii. □□a part of the "QRS complex" that consists of the first upward deflection that comes after the "P wave in an electrocardiogram. Also known as the "positive initial deflection".
iii. □□After the "R wave," this is the first time the "QRS complex" has deflected downward.
iv. □□ This is the initial positive deflection on the ECG that represents the atrial depolarization.
v. □□□ The intrinsic characteristics and autonomic aspects of the sinus node control the duration elapsed between two successive R-waves of the QRS signal. Complex relationships exist between blood pressure and vascular resistance (the ability of arteries to dilate or contract) (the amount of blood that is pumped out of the heart in a single minute).
vi. □□□□ The complex is a representation of ventricular depolarization and the electrical impulse that results from it. As it moves through the complex, an electrical impulse travel through each ventricle of the heart. The QRS complex begins just before ventricular contraction, like the P wave.
vii. □□□Action potential transmission across the AV node in an isoelectric (AV delay) manner (whereas a Q wave does not always appear, this time period is sometimes referred to as the PR interval).
viii. □□Repolarization of the ventricles is represented by this wave.

ix. Isoelectricity is the term used to describe the nerve impulse of a ventricular myocyte's ventricular potential plateau. After the completion of ventricular depolarization, this point in time is also frequently "referred to as the end of the QRS complex" and the "beginning of the T wave." This terminology is used in cardiology.

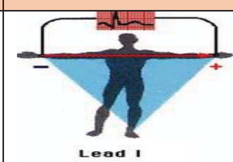
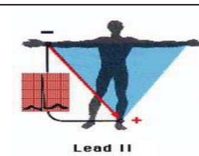
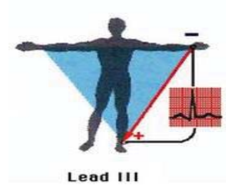
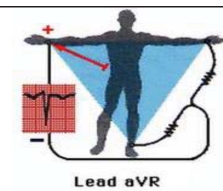
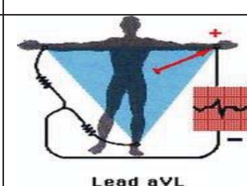
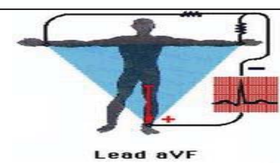
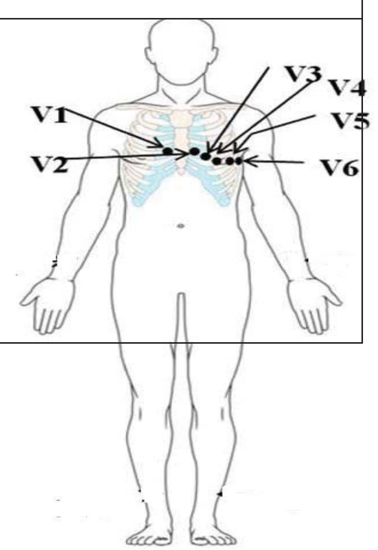
x. A cardiac rhythm disorder's QT interval is the period between the onset of the QRS complex and its conclusion; in this scenario, the amount of time that elapses between the commencement of the Q wave and its conclusion is referred to as the QT interval.

xi. subsequent to the ventricular repolarization and may not always be noticeable due to its modest size "U" waves are used to depict the repolarization of Purkinje fibers.

a s

A 12-lead electrocardiogram (ECG) is a medical test that is recorded using body-attached leads or nodes. An electrocardiogram, also known as an electrocardiogram, captures the electrical activity of the heart and transmits it to graph paper or an electronic vault in an electronic form that can be used from a medical and security standpoint by constructing a diverse dataset and presenting it to researchers for use in biometric systems [11]. The following Table (1.1) shows the location of each lead and how to capture the signal that is stored in electronic records.

□ab□ □□□□□□□□□ □ar□ □ □□□□□□a□s □□s t□□[13 , 14]□

□□□	□□a□s	□□s t□□□	□□□ □□□□art
□	□	From the right arm to the left	
□	□□	To the left leg, from the right arm	
□	□□□	Between the left arm and left leg.	
□	a□□	right arm	
□	a□□	left arm	
□	a□□	left leg	
□	□□	Positioned in the right fourth intercostal space with respect to the sternum	
□	□□	Intercostal gap to the left of the fourth major intercostals	

□	□□	between "V2 and V4" of the power supply.	
□□	□□	Midclavicular intercostal space fifth intercostal area	
□□	□□	At the same level as V4 along the line of the left anterior axillary	
□□	□□	In accordance with the midaxillary axis of V5 and (Directly under the midpoint of the armpit)	

□□□ □□ctr□car□□□ra□ □□□ □ □□□st□□

Since the electrical impulses of the heart reveals periodic patterns of beats, an electrocardiogram (ECG) is finest device for identifying a person. This ability to identify life in the ECG enables the collection of biometric data from a real, breathing human being [13]. The method is more dependable and less susceptible to fraud because to the use of an ECG biometric [14,15].

To illustrate the uniqueness of each person's ECG, the researchers employed signal processing, pattern detection, and technical learning methodologies. Morphology can be differentiated by its differences in the ECG. Changes in ionic potential, plasma electrolyte concentrations, and rhythmic variations all contribute to the distinctive appearance of the ECG [15]. Physiological variations in the thorax, other factors, such as "chest geometry," "heart location," "size," and "physical condition," may also play a part in a person's pulse rhythm. In the form, amplitude, and intervals of the major components of each person's pulse, changes show the cumulative influence of this uniqueness. In addition, cardiac arrhythmias and other diseases can be diagnosed using the ECG's information [16].

Studies of electrocardiograms have seen an increase in popularity due to technological advances in sensing (ECGs). Electrocardiograms (ECGs) are being studied as a potential non-intrusive biometric, similar to fingerprints [17].

The quality of the ECG signal and the ability to capture particular properties of individual heartbeats will determine the efficacy of ECG data gathering methods. Preprocessing and characterization of the ECG signal are therefore trivial. It has been shown that skin- or finger-contact ECG recordings are commonly tainted by noise and other abnormalities [18].

Because of the need for ever-improving technology that can both gather high-quality data as well as differentiate the unique pulse characteristics of a subject, the future of ECG biometrics looks bright. For biometric experiments with unique identification, the ECG pattern data and management representation must be strong enough [19].

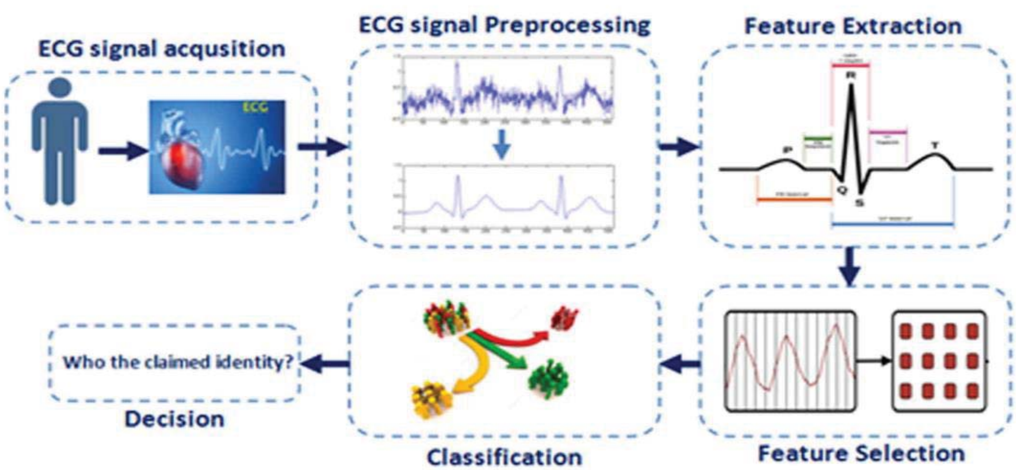
[illegible] Process

In order to verify whether or not the biometric data is accurate, the system uses pattern recognition to compare feature points to a database template [12].

Biometric systems can be used for verification or identification, depending on their intended usage. One-to-one matches between a biometric characteristic and a database registration are used for verification and identification. If these two characteristics do not change, the visitor is not real, then admission should be denied [20].

Observation reveals that a unique identifier must be used for the database identification code in the verification mode. Unlike the password verification approach, the biometric data must be same. All visitors' information must be entered into the template database [21].

A person's identification is verified by matching their biometric data to a template recorded in the system database in the verification mode. Multiple one-to-one comparisons are performed between the real-time characteristic and all biometric patterns, culminating in a uniquely identifies pattern. In terms of public safety, the identification mode is often employed to detect biometric characteristics that match the template's data [22]. Figure (1.2) General procedure for identifying an individual based on ECG data.



r ctr car ra st r a

t cat [23]

at r

Many publications in the field of ECG recognition have been published in recent years, and this thesis highlights a few of them:

- **[illegible]** [24]. Based on the electrocardiogram readings, a neural network of different shapes was created to distinguish between humans. This experiment used the "Physikalisch-Technische Bundesanstalt (PTB)" database to obtain 87 data sets from 14 individuals. MLP classification was performed on 14 individuals using “Q-R-S Feature Score”. For training purposes 66% of this data is used; For testing purposes only 34% of this data set is used; An individual can be identified with an accuracy (96) percent using only three credit scores, the proposed system is limited to assessing the system score on 14 subjects out of 290 from the PTB data set, and is considered too small to assess whether or not the algorithm is effective.
- **[illegible]** [25]. An ECG-based framework has been proposed for efficient and usable user authentication. With the PTB dataset, the ECG is preprocessed to remove any noise or distortions before it is translated into numerous sets of characteristics. These characteristics are inputs into classification models, neural networks can also be used to create classification models that can predict whether an EKG belongs to a particular individual or not, depending on the feature sets generated during the transformation phase and according to the results, deep learning neural networks are more accuracy than traditional neural networks. The methods used are Ft-Dct-Dwt-Knn-Mlp-Rbfn-Rbfn-Random Forest Classifier and the results of the proposed system are: Accuracy (DL): 50 = 79.54%, 40 = 85.79%, 30 = 89.91%, 20 = 92.32%, 10 = 95.39% Accuracy (RFN): 50 = 89.05%, 40 = 87.19%, 30 =

85.12%, 20 = 79.5%, 10 = 83.44%. The limited system of the proposed system uses only 50 subjects out of 290 from the PTB data set.

- **[illegible]** [26]. Two biometric methods were used to build an EKG-based human authentication protocol. The first is a more advanced type of Bio-Hashing, while the second is a more traditional form of matrix operations. Three publicly accessible ECG databases were used to evaluate the two reversible methods. Two data sets were used for PTB: 50 subjects from 290 subjects, Mitb: 48 subjects, and Cybhi: 65 subjects. The result of the proposed system is ERR: PTB: 0.32 SYBHI: 0.17 MITB-BIH: 0.34. The finite system in the proposed system does not use fusion techniques to create a secure multi-biometric system; Nor does it integrate the two methods used in the proposed algorithm to create a cancelable system.
- **[illegible]** [27]. To obtain an electrocardiogram, electrodes are implanted on the skin to capture the electrical activity of the heart throughout, and then individual and multiple ECG readings are checked for each subject separately for different amounts of time (1, 3, 5, 7, 10, or 15 seconds), The CSP projection matrix is then used to extract the characteristics of each signal segment, after which the obtained features are used to train the SVM classifier. 10, 20, ..., 200 individuals who are part of the PTB ECG database were used to evaluate the recommended identification approach. By using only one limb (I) and 200 patients as references, the system was able to achieve a 95.15 percent identification rate with only 0.1 percent errors. An identification rate of 98.92 percent was achieved using “single chest-based bullet (V3)” on “200 reference patients,” while the error rate was only 0.08 percent. The recommended

system does not stress the flexibility of the proposed selection strategy when exposed to the influence of these stimuli.

- [illegible]**a** [illegible] **t a** [illegible][28]. Electrocardiogram (ECG) signals have the potential to be retrieved for person identification by using the method to feature representation provided by a deep convolutional neural network. It is possible for it to extract unique features from an ECG segment without detecting any reference points, hence doing away with the labor-intensive method of acquiring signal fiducial characteristic points. Furthermore, the mean scores of feature maps should be used as global categorization characteristics. The architecture method that was recommended is, to the best of my knowledge and understanding, simple to train and optimize, and it does not call for any specific domain expertise. The voting method is straightforward. Personal recognition based on ECG can be utilized in more practical ways. To do so, three public ECG datasets were employed. The dataset used PTB:234 subject, Cebsdb 20 subject and Nsrdb 18 subject, and the result of the proposed system is Accuracy PTB (NNS):100 subject=98.54%. ,234 subject: 96.35%, Cebsdb (NNS): 20 subject=97.65%. ,Cebsdb (SVM): 20 subject=99.70% , Nsrdb(NNS): 18 subject =89.83. and Nsrdb(SVM): 18 subject=93.56. the proposed system can work on real-time ECG recognition bat not doing.
- [illegible]**t a** [illegible][29]. A proposed electrocardiogram (ECG)-based identification technique is presented. Without the necessity for biometric template generation, identifying an unknown beat bundle can be discovered using a deep learning (Duckworth-Lewis DL) technique. As a consequence of this, it will no longer be possible to dedicate a person's physiological and pathological status using the stolen templates of that

individual. This form of identification that is based on DL also overcomes the problem of being susceptible to people who have not registered. Experiments are utilized to highlight the usefulness of the suggested strategy, and these experiments involve both actual and synthetic ECGs. 200 registered persons were subjected to an onslaught of one thousand artificial single-lead ECGs, and they were able to obtain an identification rate of 97.84 percent, with a false-positive identification rate of 0.6 percent. Proposed system using PTB 200 subject from 290 subject and dell with a dataset as image. And the proposed system does not address this issue.

- **A a s a t a** [illegible] [30]. The proposed technique analyzes how deep learning algorithms can be used to exploit a small portion of an electrocardiogram (ECG) signal to identify biometrics. The amplified component of the cardioid signal can be used to form a “small convolutional neural network (CNN). Two databases, one with single-session records and one with multi-session records, were used to test the performance. Moreover, the performance of the proposed classifier was compared with four well-known CNN models: “GoogLeNet” and “ResNet,” “MobileNet,” and “EfficientNet.” The proposed model achieved accuracy by using a time-frequency domain representation of a small portion of the ECG signal surrounding the R vertex. The PTB dataset had 99.90 percent efficiency, ECG-ID mixed session datasets with 98.20 percent efficiency, multi-session ECG-ID data sets 94.180 percent throughput A 97.28 percent accuracy rate was achieved using pre-printed ResNet Using a PTB dataset of 200 subjects and 50 subjects, the current proposal in PTB has a 99.90% accuracy rate of 200 topics and 50 topics.

- **[illegible]**[31].This technology combines the deep properties of several models into a single feature, which is then handed over to a specialized classifier for validation. A support vector machine (SVM) is one type of classifier to look at. For the purpose of determining how well the proposed authentication system works, experiments including cross-validation are carried out using two public databases and not one. The performance of the merger model on the MWM-HIT database had a validation accuracy of 99.4 percent. The proposed method uses the PTB dataset with an accuracy rating of 99.6% and data of 200 out of 290 people.

As well as the most relevant paper is (B. Abd El-Rahiem & M. Hammad (2022)[31]. Only in terms of principles can the concept of integrating deep learning and machine learning make sense.

[illegible] Problem Statement

The majority of research focused on the use of ECG in the medical aspect and diagnosing diseases, not the security aspect, because all global databases on official websites are taken from hospitals for sick or healthy people for the purpose of monitoring or diagnosis, and there were no databases taken for security purposes such as human identification, verification This is one of the challenges and difficulties that we encountered in this research, as there has been little work to explore the importance and accuracy of using electrocardiograms in distinguishing or identifying people. Another challenge for the system is that dealing with any electrical signal that has its frequencies Its own data according to the device used to register it was difficult and must go through several pre-processing processes before moving to the classification stage, and this prevents us from using any type of different databases on the sites except after passing through all these stages, and this is a waste of time and effort. Therefore, a solution to these challenges was found, which is that this medical data was used in the security aspect for human identification by taking the electrical recording of the heart in each record and neglecting the medical information that accompanies it because it does not benefit us besides the security of the data.

Aim of the Thesis

The aim of this thesis is to design and implement a robust biometric system capable of distinguishing and identifying people with high accuracy for all cases and problems they face on electrocardiograms depending on computer facilities based on machine learning and deep learning techniques and employing them to build this system using an ECG data.

Layout of the Thesis

The research conducted in this treatise is organized into the following chapters. In addition to the first chapter, this consists of four other chapters and as follows is the structure of these chapters:

Chapter Two: This chapter presents the theoretical basis for the general algorithm and methods used in this thesis.

Chapter Three: This chapter describes the proposed system, its associated algorithm in detail, and the procedures involved.

Chapter Four: This chapter contains the results obtained after using the proposed system on the respective data set, in addition to an explanation of the results.

Chapter Five: This chapter summarizes the study's results and provides recommendations for future research.