Application of artificial intelligence to the electrocardiogram

by

Charalampos Tsitouridis

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master in Business Analytics and Data Science

Chairperson of Supervisory Committee: Evangelos Kalampokis

Date 2024

Acknowledgments

I would like to express my sincere gratitude to the following who have played a vital role in the completion of my master's thesis, my professor for his valuable guidance and patience, and my family for the immense support and encouragement.

Abstract

Technological advancements have resulted in an unprecedented volume of data, prompting the need for effective utilization and extraction of valuable information. The internet's evolution, widespread broadband access, and the advent of applications, especially in social networks, have amplified data availability across various sectors like healthcare, communications, and education. Processing this vast data efficiently has led to the emergence of new methods, particularly knowledge mining and machine learning.

In this context, the thesis delves into the intersection of machine learning, healthcare, and diagnostic tools, particularly focusing on electrocardiograms (ECG or EKG). The three main chapters explore machine learning methodologies, their applications in healthcare, and a practical application of deep learning in diagnosing arrhythmias using ECGs. Notably, the discussion covers deep learning techniques, neural networks, and their application to ECG analysis.

The key takeaway from the case study is the comparison of XGBoost and Neural Networks algorithms, revealing that the XGBoost algorithm proves more reliable, achieving accuracy rates exceeding 90%. The findings emphasize the algorithm's effectiveness in healthcare applications, showcasing its potential for automated pattern recognition and decision-making processes. The thesis provides valuable insights into the symbiotic relationship between machine learning and healthcare, with implications for diagnostic tools and algorithmic reliability.

Table of Contents

1 Introduction 1

2 Machine and Deep learning 4

2.1 Artificial Intelligence 4

2.2 Machine Learning 8

2.3 Machine Learning General Steps 9

2.3.1 Planning 9

2.3.2 Data Preparation 10

2.3.3 Model Engineering 10

2.3.4 Model Evaluation 10

2.3.5 Model Deployment 11

2.3.6 Monitoring and Maintenance 11

2.4 Machine Learning Processes Categories 11

2.4.1 Supervised Machine Learning 11

2.4.2 Unsupervised machine learning 14

2.4.3 Reinforcement machine learning 15

2.5 Training, Testing and Validation phases 15

2.6 Machine Learning Algorithms 17

2.6.1 Naïve Bayes 17

2.6.2 Decision Trees 18

2.6.3 Support Vector Machines 19

2.6.4 Nearest Neighbor 21

2.6.5 K-means 21

2.7 Deep Learning (DL) 21

3 AI in Healthcare 25

3.1 ECG and AI-ECG 26

3.1.1 Types of the ECG 28

3.1.2 Annotation, Modern Tools, and Formats 30
3.1.3 Available datasets 37
3.1.4 AI applications using ECG 39
3.2 Arrhythmia 44
3.2.1 Types of arrhythmias 44
3.2.2 AI applications on arrhythmia detection 48
4 Application and comparison of algorithms on case study 53
4.1 ECG Description 53
4.2 ECG Data 58
4.3 ECG Annotations 59
4.4 Methodology 60
4.4.1 Getting data 60
4.4.2 Data Analysis 61
4.4.3 Preprocess 63
4.4.4 Process 63
4.5 Convolutional Neural Networks 66
4.5.1 Convolutional Layers 68
4.5.2 Pooling Layer 70
4.5.3 Fully Connected Layer 70
4.5.4 General Operation 70
4.6 XGBoost Algorithm 71
4.7 Results 74
4.8 Discussion and Conclusions 83
5 References 84

Table of Figures

Figure 1: Knowledge mining techniques categories 5

Figure 2: Machine Learning Process 9

Figure 3: Classification 12

Figure 4: Clustering 14

Figure 5:Train, validation and test datasets relationships and use 17

Figure 6: Example of a Decision Tree 19

Figure 7: Artificial Neuron Typical Structure: 22

Figure 8: Artificial Neural Network Typical Structure 24

Figure 9 – Overview of artificial intelligence and machine learning in cardiac electrophysiology (source: [8]) 26

Figure 10 - Normal 12-lead ECG (source: ecglibrary.com) 27

Figure 11 – Framework for AI-ECG in clinical practice (source: [5]) 28

Figure 12 - Illustration of the spatial angles of the 12 single-lead ECG system (source: [21]) 28

Figure 13 - Holter monitor (source: hopkinsmedicine.org) 30

Figure 14 – Internal structure of the Heart (source: https://training.seer.cancer.gov) 30

Figure 15 – Sinus Rhythm Waveform (source: https://www.wikipedia.org/) 31

Figure 16 – Parts of the electrical system of the Heart (source: https://www.hopkinsmedicine.org/) 31

Figure 17 – Illustrated areas of the Heart and the leads which are used to determine the electrical activity of the area 32

Figure 18 - Grid boxes on an ECG strip 33

Figure 19 – BeatLogic platform flowchart (source: [28]) 35

Figure 20 – Annotation interface of the annotation system developed by this research paper (source: [31]) 35

Figure 21 – Waveform's architecture (source: [32]) 36

Figure 22 - Sample of hyperkalemia (source: https://ecglibrary.com/) 39

Figure 23 - Sample of Hypokalemia (source: https:/ecglibrary.com) 40

Figure 24 - Block diagram of algorithm for ECG interpretation by similar ECG (source: [X5]) 41

Figure 25 – Training, Validation, and Testing of an AI-Based Electrocardiography Screen for Hypertrophic Cardiomyopathy developed on study 'Detection of Hypertrophic Cardiomyopathy using a CNN-enabled ECG [49] 42

Figure 26 - ECG from an 83 year old man with aortic stenosis (source: https://ecglibrary.com) 43

Figure 27 - Sinus tachycardia (source: https://ecglibrary.com) 44

Figure 28 - Sinus bradycardia (source: https://ecglibrary.com) 45

Figure 29 - Atrial Flutter (source: https://ecglibrary.com) 46

Figure 30 - Ventricular Fibrillation (source: https://ecglibrary.com) 46

Figure 31 - Long QT interval (source: https://ecglibrary.com) 47

Figure 32 - Participant Selection (source: [75]) 49

FIGURE 33 - SUMMARY OF DATA USED IN THE STUDY (SOURCE: [90]) 51
FIGURE 34: ECG'S PARTS 54
FIGURE 35: ECG PAPER 54
FIGURE 36: LEADS AND HEART VIEW ANGLE CORRESPONDENCE. 57
FIGURE 37: A 12 LEAD ECG EXAMPLE 57
FIGURE 38: SINUS-RHYTHM AND ARRYTHMIA COMPARISON 58
FIGURE 39: COMPARISON BETWEEN FULLY CONNECTED AND CONVOLUTIONAL LAYERS 67
FIGURE 40: TYPICAL CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE 68
FIGURE 41:GRADIENT BOOSTED TREES ALGORITHM 71
FIGURE 42: ALL EXPERIMENTS COMPARISON DIAGRAM 80
FIGURE 43: DNN 5 AND 2 CLASSES CLASSIFICATION 80
FIGURE 44: DNN 5 CLASSES CLASSIFICATION 6 AND 2 SECS WINDOW COMPARISON 81
FIGURE 45: DNN 2 CLASSES CLASSIFICATION 6 AND 2 SECS WINDOW COMPARISON 81
FIGURE 46: 5 CLASSES DNN (BLUE) AND CNN (RED) COMPARISON 82
FIGURE 47: XGBOOST ALGORITHM RESULTS 82

1 Introduction

New technological advancements have led to an abundance of data and the computational infrastructure to gather, utilize, and extract useful information. The development of the internet and its applications played an important role in this development. Access to broadband network connections has become possible for a large part of the world's population. Applications were created that addressed a wider gap than in the recent past (social networks are a typical example). More people were now using web applications and, in many cases, storing their own content on its infrastructures. Added to this content was that created by the sensors of the Internet of Things (IoT) devices. The amounts of data available on the internet are large.

Data can be found in various sectors of our society, such as healthcare, communications, education, banking, consumer trade amongst others. Thus, the need to transform raw data into valuable information grows rapidly. As large volumes of data can be produced and reached more easily nowadays, the problem has been moved from getting it to exploiting them. Conventional ways for processing data are not able to handle efficiently large amounts of data. So new ways for processing data had to be emerged in order to manage the high data availability.

Knowledge mining from large data sets is now one of the most basic research issues. Administrators of online IT systems are looking for ways to efficiently exploit the large availability of data. The techniques used aim to extract knowledge from large data sets. With these techniques, data scientists and analysts seek to extract knowledge structures that the data can describe. The structures developed include appropriate semantics and are based on common properties of the data under consideration. A typical definition for the concept of knowledge mining is: "knowledge mining is the deterministic process of identifying valid, innovative, useful and understandable patterns in data". Data describes entities or relationships between entities. Prototypes have the ability to define subsets of data whose elements exhibit common properties. One of the main challenges of knowledge mining processes is to create patterns that are consistent with new data. This criterion is also the most important for their evaluation.

A major role in this issue has been machine learning. The capabilities of machine learning exceed the simple notion of extracting and transforming data to automate complex procedures or even enhance existing practices. Its algorithms have been made

a useful tool for decision making processes. They offer efficient processes for manipulating large volumes of data to produce models for automated pattern recognition. These patterns are then used for decision orientation. Machine learning algorithms have been shortening the data processing time and made able the big data to be processed. There are several applications of machine learning processes to a large variety of sectors such as economy, education, marketing, and healthcare. These techniques give machines artificial intelligence so that they can automatically adjust their behavior according to the data they manage.

The focus of this thesis is centered on the healthcare sector and the implementation of machine learning techniques in diagnostic tools, specifically the electrocardiogram (ECG or EKG). The main structure of this paper includes three chapters:

- Chapter 2: The main goal of this chapter is to provide a synopsis about the definition and the characteristics of machine learning. The different methodologies followed for machine learning and their products are described. The applications of these methodologies are mentioned. Special mention is made of deep learning and neural networks.
- Chapter 3: The second chapter's content focuses on machine learning applications for healthcare. There are several healthcare sectors where artificial intelligence application has benefit relevant functions. ECG study is one of them. A detailed description of ECG features and processes is presented. Then, a comprehensive analysis, about how machine learning processes are being applied on ECGs, is provided. This chapter's content is essential for constructing basic knowledge about the processes the machine learning algorithms must be adapted.
- Chapter 4: In the third chapter, a practical application of deep learning methodologies in healthcare is presented. It examines how electrocardiograms can be evaluated to diagnose arrhythmias in heart function. The presentation includes the description of ECGs, how they are performed and the information they provide. The structures and characteristics of the developed neural networks are also described. Python's capabilities for developing machine learning mechanisms are reported. Finally, the test results of the developed neural networks are shown, as well as the comparison among them.

- Chapter 5: The last chapter contains the conclusions derived by the machine learning technics, their applications in healthcare and the case study performed.

2 Machine and Deep learning

2.1 Artificial Intelligence

The evolution of computing has been largely based on data processing. In the 1960s data was organized into data collections, databases, and related management systems. A few years later, the relational data model (RDBMS) was proposed, which evolved into special purpose databases (spatial, temporal, scientific, engineering). The increase in data availability has led to data mining and data warehousing. Gradually, mechanisms were also proposed for associating data with semantics (Linked Data, Ontologies).

Large amounts of data are collected and stored into data warehouses with high frequency. These data have complex relationships among them that are difficult to detect. The revelation of these relationships leads to knowledge. Knowledge gives an advantage over the competition, an opportunity to provide better quality services. With knowledge of mining techniques data can be grouped based on specific criteria. Knowledge mining contributes to the better formulation of hypotheses and the better visualization - presentation of information.

Knowledge mining techniques are distinguished according to the type of models they produce into predictive and descriptive. Predictive techniques are used to produce estimates of the outcome of current situations. Descriptive ones identify current states. Predictive techniques contain the following mechanisms:

- Classification: It is defined by a well-defined set of classes as well as a set of pre-classified samples. Its purpose is to create a model for the classification of new samples whose class is unknown.
- Regression: It is the process of assigning data to a variable that refers to a future state. These variables usually take continuous values.
- Time Series Analysis: Time series are sequences of numbers each associated with a timestamp. Each timestamp is separated from the previous and next by a fixed amount of time. The purpose of these mechanisms is to predict future values of the series.

Descriptive techniques include:

- Clustering: It is the process of dividing a set of heterogeneous objects into groups based on the similarity or difference that the samples have.
- Summarization: They are feature reduction operations to identify objects.
- Association rules: They are processes of discovering association rules between objects. Associations are based on the values that the objects' specification parameters take.
- Sequence discovery: In these processes, temporal or other sequences are studied in order to identify patterns that determine how they are being developed.

The following chart shows the knowledge mining categorization.

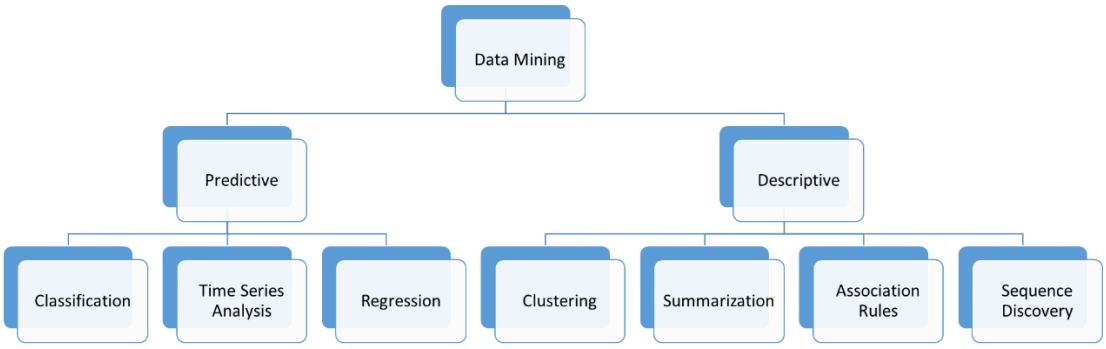


Figure 1: Knowledge mining techniques categories

Knowledge mining systems use but are not limited to data warehouses, statistical methods. The difficulties that need to be overcome when performing knowledge mining processes are:

- The large size of the data to be processed.
- The large number of dimensions of the data (the number of attributes that describe the entities).
- The inhomogeneity of the data might be derived by different sources formed in many ways.

Knowledge mining has many applications. The main ones are:

- Optimizing search engine performance: Data from past searches – either criteria or results – is used to improve the results of future searches.
- Personalization of content and services based on data related to the use of services and goods: Customer behavior is recorded, and its data is used to shape the organization's attitude towards them.
- Determining the strategies of commercial organizations: Customer preferences form the basis for decision-making processes.
- Scientific research: Knowledge mining techniques are used to carry out scientific research. The conclusions drawn from these processes contribute to the progress of science.

The concept of knowledge mining relates to efficient techniques for analyzing large collections of data to draw useful conclusions from them. The information resulting from these procedures is non-trivial, comprehensible, validated and provides useful conclusions. The relationships that emerge between the data and are not expected provide a greater amount of information.

The inability of organizations to process the data they have in a timely manner can have bad consequences for their viability. These problems are:

- Reduced productivity: Traditional data processing techniques lead to results at a slow pace. This has the effect of reducing the useful time of the workers.
- Inadequate exploitation of data: The inability of traditional data processing techniques to produce timely results deprives the heads of organizations of important tools for decision-making processes.
- Inability to exploit unstructured data: Conventional forms of data processing do not have the ability to handle unstructured data. In the modern era when the availability of unstructured data is great, the inability to draw conclusions from it deprives the organization of possibilities.

As knowledge mining is important for any kind of organization, artificial intelligence adapted to its algorithms. Artificial intelligence (AI) refers to machine-based data processing which otherwise would require human cognitive function. Nowadays AI has become ubiquitous and there are applications in multiple facets of life ranging from

language processing, computer vision, engineering, industrial, gaming, and other scientific fields. On daily basis one can detect these applications such as face and objects recognition in photographs, language detection and auto-completing words or even sentences in written communication, as well as verbal communication with certain applications as Alexa or Siri. Marketing also uses AI extensively, in order to suggest products and advertisements to users based on their profile and searches through the internet. Consequently, the advancement in the speed of computing and the development of new algorithms which partly mimic the human brain function have led to widespread implementation of AI-based practices.

Applications of artificial intelligence can bring many benefits to humans. The measures taken to deal with the consequences of Covid-19 made research look for ways to continue operations in various areas of human activity. AI offered significant solutions. In addition, it allows the development of new categories of products and services that companies can have in order to create new, innovative production lines. It also strengthens the existing production lines through the improvement of production mechanisms and the strengthening of its supporting operations. It is estimated that by 2025 business productivity will increase by 11% to 35%. AI applications can accelerate and improve the accuracy of public administration operations and upgrade its relationship with citizens. The use of AI in military applications and in general in the field of public security contributes to the development of a safe environment for modern societies.

The use of AI also presents several challenges. Organizations should take care to monitor and adapt to developments in order to remain competitive. On the other hand, excessive and incorrect use of AI applications can lead to problematic situations and social issues. The errors of the automated processes of AI applications are difficult to attribute as the distinction of responsibility of the producers of the systems and the users is not clear. Their operation requires the storage, maintenance, and processing of large volumes of data, which in many cases are personally sensitive. In addition, the development of AI mechanisms may include the incorporation of bias to produce the result, to such an extent that the behavior of critical systems is manipulated. As with many modern technological achievements, it can lead to job cuts. Inadequate regulation and adaptation of AI systems to their environment and the demands of society can have bad results for their users[119].

2.2 Machine Learning

Machine learning (ML) is a sector of AI and provides a different approach on how scientists use AI traditionally. It is a tool which can learn patterns empirically from data and identify nonlinear relationships and interactions between multiple variables that may be challenging for traditional statistics. Additionally, instead of needing specific rules to input data into the program to generate results, machine learning uses data and extracts answers from the data to construct the rules. The rapid growth of Machine Learning became apparent during the latest decades due to the availability of large datasets (Big Data) and computing power. ML's place in the technological realm has been solidified as one of the most useful and popular subfields of AI.

The general purpose of machine learning is building models using large volumes of data. Its processes aim to generalize the relationships among them. These models need to learn how to distinguish certain objects in order to achieve a desired total result or come to a useful conclusion. A large variety of techniques and algorithms are being used. Machine learning applications learn from data and use fresh data to recognize known patterns. Those patterns have been derived from the study of objects characteristics that describe the data. They are features that can determine the species or the identity of the object to be recognized. Additionally, machine learning can determine through the patterns it creates what objects cannot be[126].

A variety of fields have implemented techniques from ML such as mathematical statistics, probability, decision theory and information theory. ML has the ability to tackle issues often involving extremely large, complex data of various types (like images, video which consist of thousands of pixels). As a result, ML has become a more engineering-oriented approach[1]. (and is becoming a growing part of medicine).

ML systems require training data from which they design and produce certain rules. Once the system is trained on the data provided, it proceeds to the testing phase in which a different set of data (testing dataset) is used to predict answers from new data based on the aforementioned rules. The ability of an ML algorithm to classify or predict, based on data which are detached from the training dataset is the central tenet in ML[2]. One additional step required for the training phase is often the preprocessing of data[3]. For instance, when the algorithm's aim is to classify pictures as dogs or cats, the images need be processed in order to have the same size or pixel intensities to a fixed range.

Preprocessing methods developed on the training dataset and applied to the test dataset improve the model's generalizability and functionality. In contrast, the traditional ML approach to feature extraction has been feature engineering, in which the researcher modifies the raw data based on experience to condense information into values to be used in ML.

Machine processes execution depends on the algorithm to be implemented. However, each process follows a sequence of general steps. These steps are planning, data preparation, model eengineering (train), model evaluation, model deployment, monitoring, and maintenance.

2.3 Machine Learning General Steps

The general process of machine learning projects consists of the stages described below.

Figure 2: Machine Learning Process

2.3.1 Planning

During this stage, the requirements of the project to be developed are defined. The goals of the machine learning process are defined based on the purpose for which it will be used. The type of procedure that is most suitable is then selected as well as the algorithm that is estimated to produce the best results. All implementation details are also defined and the tools to be used are selected. The sources of the data to be used in the next stages are identified and the way of gathering and processing them is planned. If necessary, a schedule of actions is drawn up. The evaluation criteria of the process to be developed, the tools and metrics to be used for this purpose are also defined.

2.3.2 Data Preparation

It is a process that begins with the collection of data from various sources and ends shortly before training the models. It starts with data collection. After data sources have been define, they are accessed and gathered by several mechanisms (depending on data forms and sources availability). The validity of collected data must be examined (this is mainly depended on source's reliability. In special case, data labelling might be needed (e.g. classification). After that data have to cleaned. Records with outliers should be removed (especially if they are a small part of the total), a decision should be made on how to use the missing values (usually the corresponding mean, some neutral value, or the entire record is removed). Records are also checked for incorrect values so that they can be corrected. Subsequently, the data undergoes processing in order:

- The amount of data to be sufficient for the development of the process
- The data set should present a quantitative balance in terms of the field values of the records. The percentage of field values that are input or output of the procedure should be as equal as possible.
- Data values to be normalized to common scales.

2.3.3 Model Engineering

It is the most important part of the process as this is when the model is created. The algorithms designed in the first stage are implemented and the data prepared in the second stage are used. The model is trained and at the end of this process it is ready to be evaluated. Usually during the training phase, the model parameters are adjusted in multiple steps, at each step its efficiency is checked and re-adjusted. This process is repeated until the value of some selected metrics reach a threshold.

2.3.4 Model Evaluation

The model developed should be evaluated for its ability to produce the desired results accurately and in an acceptable time. It mainly evaluates the ability of the model to work efficiently with data unrelated to the data used during training. Appropriate performance metrics are applied at this stage.

2.3.5 Model Deployment

At this stage it has already been decided that the model is capable of producing reliable results when fed with data under any conditions. It is developed and now normally used for the purpose for which it was designed.

2.3.6 Monitoring and Maintenance

It is a process that lasts until the end of the model's life cycle. This is constantly checked for its performance and whenever required adjustments are made to its parameters and it is retrained. At each revision of the model re-evaluations are carried out.

2.4 Machine Learning Processes Categories

Depending on the purpose served by each machine learning-based process, the appropriate procedure and algorithm is chosen. In addition, the appropriate parameter values for the selected algorithm are determined. The techniques used in machine learning processes fall into three main categories: supervised, unsupervised and reinforcement learning. Each of them is described to the following paragraphs[126].

2.4.1 Supervised Machine Learning

The main characteristic of the methodologies belonging to this category is that the set is used to train the model, it includes entities described by a set of attributes. The set of these attributes includes the one for which the model is developed. Therefore, the model is trained based on the values of the parameters that characterize each entity, knowing with which value of the target characteristic they are associated. After the model is trained, the new data is applied to the trained model and the estimated value of the target feature is assigned. These techniques are usually used when the data can predict possible events. Algorithms of this kind learn a set of inputs along with corresponding correct outputs and learn by comparing their actual outputs with correct outputs to find errors. The detection of errors determines the corrective actions in the model[120].

The main procedure included in this category is Classification. It is referred to the prediction attempt of output of a specified input. The classification process takes an unknown group of entities and upon completion of its operation, places each entity in

the most relevant category. In order for these processes to work better, enough data is required in a crowd which is evenly distributed among the different categories[123].

The next figure shows how supervised machine learning works.

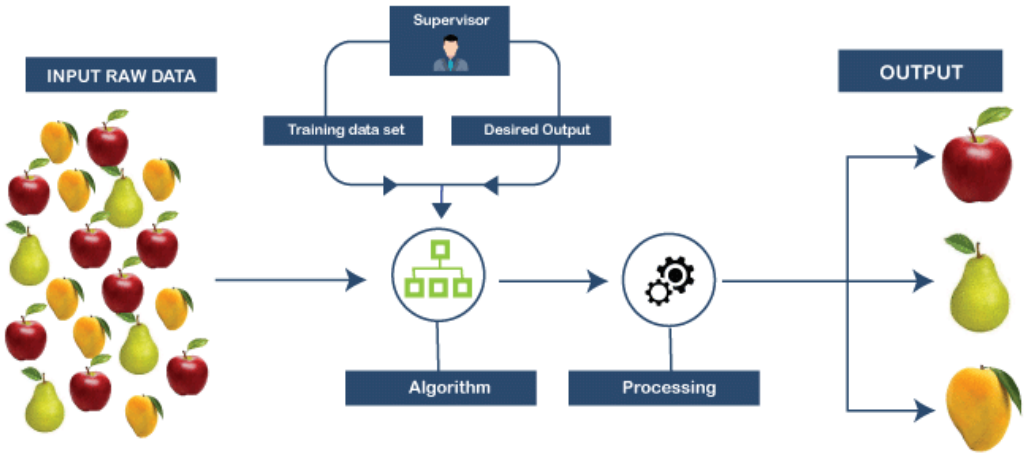


Figure 3: Classification

Built models' quality are measured by various metrics. The most important and mainly used of which are:

- Classification accuracy: It is described by the ratio of the number of correct predictions to the total number of input samples.

$$accuracy = \frac{correct\ predictions}{total\ number\ of\ samples}$$

- Logarithmic loss): It is described by the equitation:

- $LL = \frac{-1}{N}\sum_{i=1}^{N} \sum_{j=1}^{N} y_{ij} \log(p_{ij})$

where y_{ij} points if i-th sample belongs to the j-th class, p_{ij} is the probability the i-th to belong to j-th class. Logarithmic loss values are in the space of $[0, \infty)$. Zero value refers to absolute accuracy.

- Confusion Matrix: It is a two-dimensional table whose cells record the forecasts as follows:

- True Positives (TP): The cases in which the prediction was A and the actual value was A.
- True Negatives (TN): Cases in which the prediction was not A and the actual value was not A.
- False Positives (FP): The cases in which the prediction was A and the actual value was not A.
- False Negatives (FN): The cases in which the prediction was not A and the actual value was A.

- Confusion Matrix is the base for useful metrics computations:
  - Accuracy: It refers to how close predictions are to the corresponding true values.
    - $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
  - Precision: It refers to how close predictions are to each other. It is equal to the ratio of the number of true positives values to the summary of positive predictions.
    - $Precision = \frac{TP}{TP+FP}$
  - Recall: It measures the model's ability to recognize the samples classes. It is equal to the ratio of the number of true positive predictions and the summary of true positive and false negative predictions.
    - $Recall = \frac{TP}{TP+FN}$
  - F1 Score: It is a metric for evaluating the model's ability for predicting both positive and negative samples.
    - $F1 - Score = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}$

- Mean Absolute Error: It is a metric used when the target attribute takes continuous values. The mean absolute error is the average of the difference between the original values and the predicted ones. It gives

the measure of how far the predictions are from the actual values. Also used is the mean squared error (MSE) which calculates the average of the square of the difference between the original values and the predicted values[124].

2.4.2 Unsupervised machine learning

Learning processes of this kind, manipulate samples the target characteristic values are not known during the training of the model. The relevance of the entities to each other or lack of relevance is calculated, is estimated depending on the objects' attributes values. Through suitable processes patterns emerge without having to know any previous data or information. A typical case where unsupervised machine learning is used is clustering. In such procedures, the elements of a set are placed into the same clusters according to the degree they match each other or into different clusters according to how different they are from each other. This type of learning works well with clustering, which refers to data placing into groups of similar data[126].

The next figure shows an example of samples clustering in two-dimensional space.

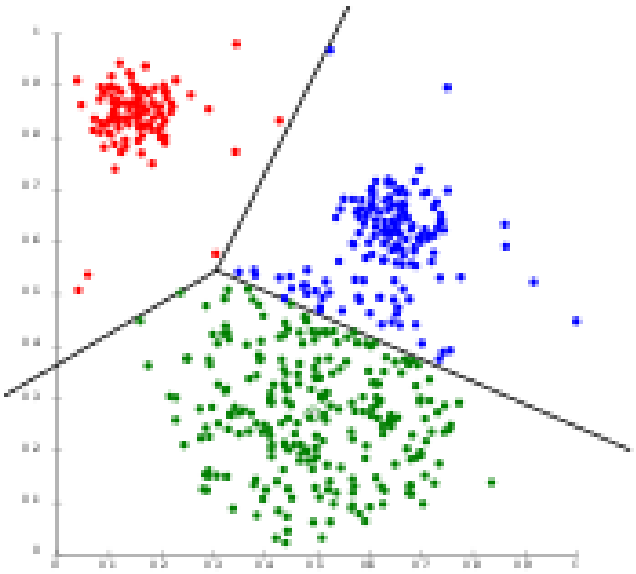


Figure 4: Clustering

The most common metric used for evaluating clustering processes quality is silhouette. Silhouette coefficient value is in the space of [-1,1].

- 1: Value close to 1 means that clusters have been distinguished perfectly and the probability the samples to be placed to other clusters is small.
- 0: Value close to zero means the samples have been placed to correct clusters but they would be placed to other clusters as well.
- -1: Value close to -1 means that clusters have not been formed correctly.

The silhouette value is computed by the following equitation:

$$Silhouette = \frac{b - a}{\max(a, b)}$$

Where:

- α: is the sample's mean distance among all pairs of samples in a cluster built.
- b: is the sample's mean distance among other samples placed on different clusters[116].

2.4.3 Reinforcement machine learning

The reinforcement learning approach is based on algorithms interaction the environment they are being executed. The environment status is being detected periodically or continuously and it is being recognized by appropriate sensors. The algorithm adapts this status, and it behaves depending on it. Behavior is a combination of selected actions from a set of dispositions. These options as well as their results are recorded. The set of entries constitutes the experience of the system on which the algorithm is based in order to make decisions. The main idea of such algorithms comes from the field of psychology and is related to reinforcing behavior, learning by doing and getting relative reward. A machine can approach a result considered ideal through successive trial and error. Over time, it learns to choose certain actions that result in the desired output. This type of learning is often used in applications such as games and the movement of unmanned vehicles and robots[126].

2.5 Training, Testing and Validation phases

Training and testing phases are the most important ones for model quality. The training process is based on a loss function used to determine the distance between the computed

and the actual value of target features. It is desired to be as minimized as possible. This process of minimizing the loss function is iterative and lasts where its output is less than a predefined threshold. Parameters that are considered for building the model are called hyperparameters. During the training phase hyperparameters are being regulated in order to minimize the loss function output value.

A portion of the available dataset is being kept for using it for testing purposes (usually the 25% - 30% of the entire dataset). This is called test set while the rest is called train set. After having constructed the model, the test dataset is used on the model for evaluating metrics values. n most cases where machine learning techniques are used, the aim is to produce models using more than one algorithm. Each model is evaluated with the test dataset in order to select for use the one that appears best based on the evaluation metrics used[127].

It is a good practice to choose a portion of train dataset to use it as a validation set. This dataset would be used for configuring hyperparameters during the train phase in order to decrease the loss function value in each iteration. Checking the model's behavior during the training phase helps the trainer to configure hyperparameters in a manner that makes the model better than the one provided by the previous iteration. The most common technique used for validating training is the k-fold cross-validation. The dataset is being split into a number of folds (subsets). One of them is randomly chosen to be the validation dataset while the union of the rest is used as the train dataset. The model validation process is being performed repeatedly. The model performance is computed as the average value of the validation processes performed. The main advantage of this method is that the performance estimation would not be affected by the subset chosen for train or validation dataset. Its drawback is the fact that its high complexity[117]. Bootstrap is a popular method too. It uses resample with replacement, creating new datasets having the same number of elements with the original ones. That means that some records might be more than once in the set. The generalization error computed between performance on the bootstrapped set compared to original set is the measure for evaluating the derived mode. This method tends to produce overfitted models in some cases[127].

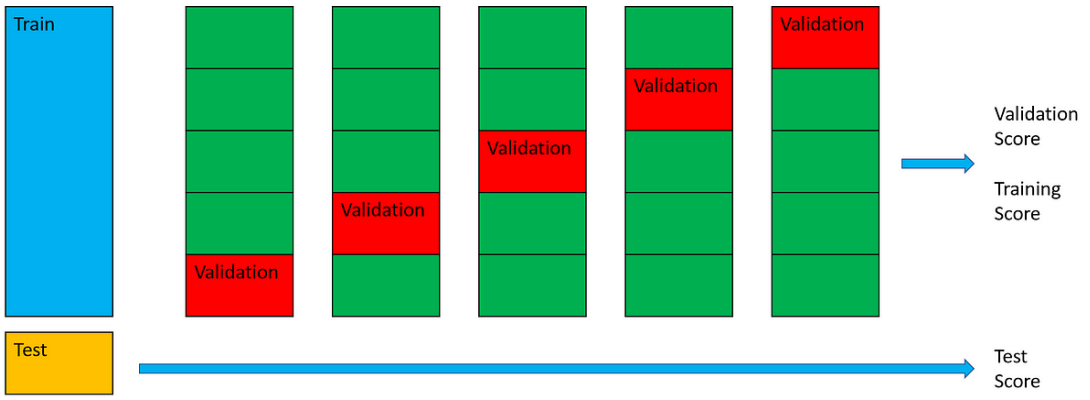


Figure 5:Train, validation and test datasets relationships and use

To sum up, the training phase consists of a repetitive sequence of train – validation cycles, resulting to the model testing. Train and test dataset are separated while the validation dataset is a portion of the train one. The above image shows schematically the train, validation and test set relationship into the model training and test process.

2.6 Machine Learning Algorithms

2.6.1 Naïve Bayes

This method is based on Bayesian statistical theory. The goal is to categorize a sample X into one of the classes C1, C2, … Cn, using a probability model defined according to Bayesian theory. Each class is characterized by a prior probability of observing the class Ci. It is also assumed that the given sample X belongs to a class Ci, with the conditional probability density function p(X|Ci) in the interval [0,1]. Then using the above definitions, the probability is determined as follows:

$$p(C_i|X) = \frac{p(C_i|X)p(c_i)}{p(X)}$$

The simplest classifier is Naïve Bayes which assumes that the effect of a feature on a given class is independent of the values of other features.

2.6.2 Decision Trees

A decision tree is a map that shows all the possibilities and outcomes that can occur when a particular topic is discussed. It is a sequence of related options and enables individuals and groups to weigh possible outcomes against costs, priority, and benefits. Decision trees are used to guide informal discussion or create an algorithm that mathematically predicts the most important choice. Furthermore, a decision tree starts with a root node, which branches out too many possible outcomes. Each possible product also comes with additional nodes that result from the results and can be branched. When all possible outcomes have been branched, a tree-shaped diagram will be created. There are several types of nodes you can see in your decision tree: opportunity nodes, decision nodes, and termination nodes. A circle represents an opportunity node and shows the probabilities of the outcomes you can get. A square shape represents a decision node, indicating the decision to be made. And finally, the terminal node represents the result of the decision tree.

The criterion considered for the quality of decision trees is their height. The smaller the height of the tree, the faster it can lead to a decision. The height of the tree depends on the choice of characteristics of the samples to be examined first. The most popular algorithms of this kind first select those features that yield the greatest informational gain.

It is a useful tool for data interpretation and suitable for handling numerical and non-numerical data. Decision tree methodologies can be implemented easily and combined with other techniques. In the phase of creating the tree, its complexity should be limited so as not to lead to overfitting. They cannot be easily used for entities described by continuous numeric values. Decision trees produce lower prediction accuracy compared to other prediction methods[123].

Next figure shows an example of a decision tree.

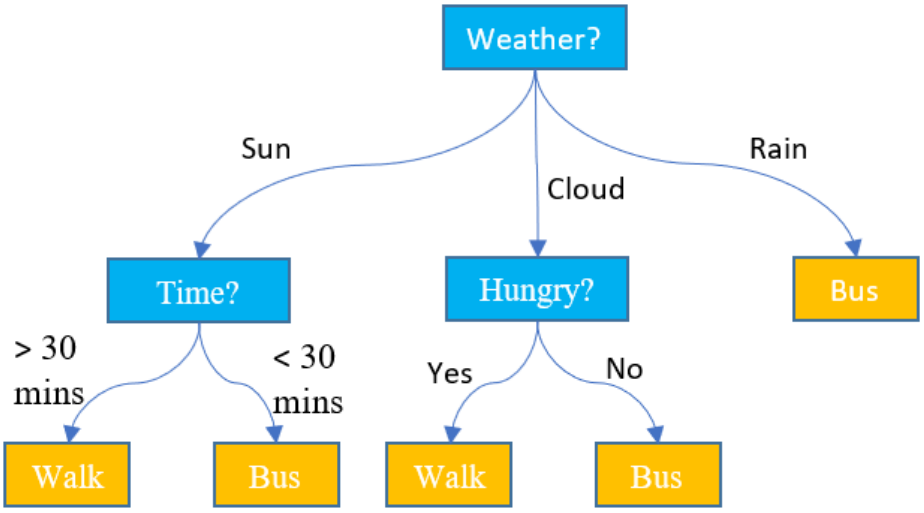


Figure 6: Example of a Decision Tree

2.6.3 Support Vector Machines

Suppose that n observations are recorded and each of them consists of a pair of the form: Xi ∈ Rn , and a label of the target category, yj. Suppose also that there is some unknown probability distribution P(x,y) based on which these observations are produced. The goal is to compute the set of parameters α of the function f(x,α) such that f realizes the mapping xi → yi. A particular choice of α uniquely determines the corresponding training engine, f(x,α).

Unlike traditional methods, which minimize the error, an SVM aims to minimize the upper bound of the generalization error. It achieves this goal by learning α of f(x, α) so that the training engine satisfies the maximum margin property, i.e. the decision boundary that represents the maximum - minimum distance from closest point calculated during training.

The expected control error for a trained machine is:

$$R(\alpha)\ (\tfrac{1}{2}\int |y = f(z,a)dP(x,y).$$

R(a) is called expected risk, or simple risk. Although it gives a good way of calculating and representing the error of the true mean, it requires the estimation of P(x,y). The empirical risk Remp(α) is then defined as the error rate measured over the training set:

$$Remp\ (\alpha) = \ (\tfrac{1}{2}\ \Sigma|yi\ -\ f\ (xi,\alpha)|)$$

In addition:

$$R(a) <= R(a) + C(h)$$

where h is the Vapnik Chervonenkis constant (VC) and describes the measure of the machine's ability to learn by any training set without error. The term C(h) is called confidence VC. Given a family of functions f(x, α), it is desirable to select the machine that gives the lowest upper bound on the risk. The first term, Remp(a), represents the accuracy achieved on a particular training set, while the second term, C(h), represents the ability of the machine to learn from any training set without error. The terms Remp(a) and C(h) define the bias and variance of the generalization error. The best generalization error is achieved when a proper balance is achieved between these two terms. This property is the basis for a method of selecting a learning machine for a particular task in order to minimize the construction error. An SVM engine maps the data into a high-dimensional space (feature space) and defines a separating hyper-plane in that space. Translating the training set into a high-dimensional space increases the overall learning cost. SVMs avoid overfitting by choosing a specific hyper-layer to partition the data into the feature space.

A function, the kernel function, is used to construct the feature vectors, instead of explicitly recording them. Therefore, an SVM can detect a cutoff in a feature space and categorize the points in that space without always mapping that space uniquely.

Additionally, overfitting is avoided by using the maximum margin level, which leads to a training algorithm with characteristics of an optimization problem solving method. In order to train the SVM the unique minimum of a function must be found. Consequently, SVMs do not have the local minima problem that affects many training schemes, and unlike the back propagation learning algorithm for neural networks, they always converge deterministically to the same solution for a given data set, regardless of the initial assumptions.

In the simple case of two linearly distinguishable classes, an SVM selects that classifier that minimizes the upper bound of the generalization error. It achieves this goal by computing the classifier that satisfies the maximum margin property (the one whose threshold decision has the maximum minimum distance from the nearest training point). If the two classes are indistinguishable, SVM searches for a level that maximizes

the margin and minimizes an amount proportional to the number of classification errors[125].

2.6.4 Nearest Neighbor

K-Nearest Neighbor is a machine learning technique that operates by measuring samples similarity with other nearby ones. For each of the samples for which the class to which it belongs is searched, the class to which the k samples closest to it belong is examined. The category to which most of these samples belong is the category to which the examined sample belongs[118].

2.6.5 K-means

It is an algorithm mainly used in clustering problems. It is initialized by selecting k samples to be the centers of k different clusters. Each of the remaining samples is then assigned to each of the centers. When the clusters form, their new centers are located. This process is repeated until at some step the centers of the clusters do not change[118].

2.7 Deep Learning (DL)

Deep learning is based on multilayered models. Each layer represents a stage of abstraction to the learning process. In recent years, scientific research has led to important applications of these models, in various and diverse fields, with excellent success. Speech recognition, visual object recognition, object detection, drug discovery and genomics are some of these applications. Deep learning discovers intricate structures in large data sets. It performs backpropagation to indicate how a machine should change its internal structure and structure elements relationships through the layers of the entire model[122]. Deep Learning techniques are processes that tend to simulate the functioning of the human brain and the behaviour of the human central nervous system. Based on this, a neural network, consisting of artificial neurons, can perform machine learning and pattern recognition efficiently. Artificial neurons are interconnected and support functions of processing a set of input parameters and producing a customized output format. A neural network is an oriented graph. Its edge corresponds to dendrites and synapses. Each edge is assigned a weight and connects a

pair of neurons – nodes. The values that enter each node constitute its input. In it they are applied on a predefined activation function[115].

An artificial neural network is a data processing structure that works like the way the human brain processes information. The human brain is made up of millions of neurons. It sends and processes signals in the form of electrical and chemical signals. These neurons are connected by a special structure known as synapses. Synapses allow neurons to pass signals. Artificial neural networks include a large number of connected processing units that work together to process data. Beyond categorization procedures, it is suitable for regression of continuous-valued features[114].

Neural networks are made up of neurons based on the neuronal structure of the brain. They process the elements one at a time and "learn" by comparing their categorization for a record (which, to begin with, is largely arbitrary) with the known actual categorization of the record. The errors from the initial classification of the first record are fed back to the network and are used to modify the network-based algorithm again. This process continues iteratively.

The structure of neurons in the artificial neural network contains[121]:

- A set of input parameters that take the values of the input (Xi).
- A set of weights applied to each of the input parameter
- A function (g) that sums the weights and maps the results to an output (y)

The following figure shows the typical structure of neurons.

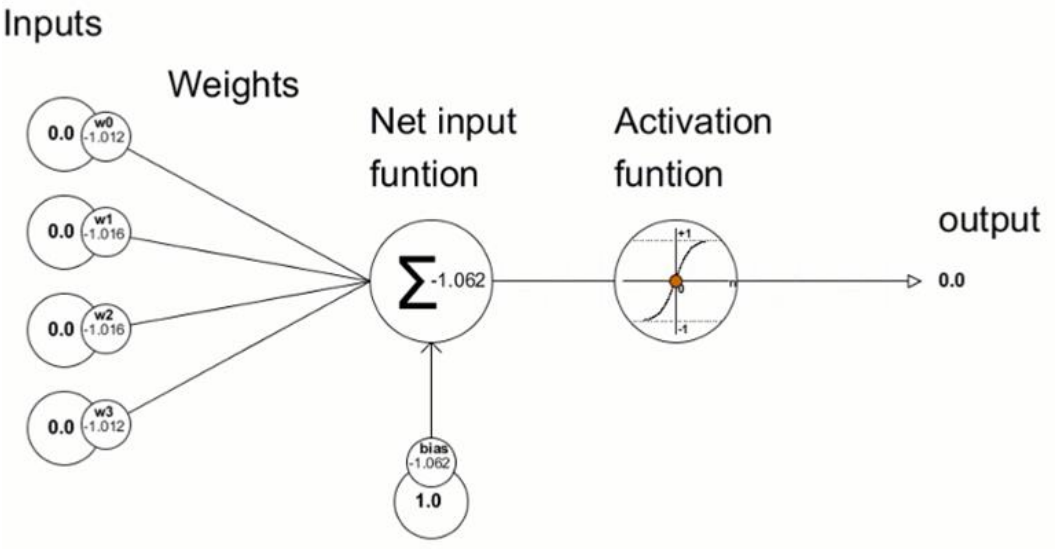


Figure 7: Artificial Neuron Typical Structure:

Artificial neural network consists of at least three layers[113]:

- Input layer: The purpose of the input layer is to take as input the parameter values for each observation. Typically, the number of input nodes is equal to the number of independent variables. The input layer is linked to one or more hidden layers. Input layer nodes are passive, meaning they do not change data. They take a single value on their input and copy the value to their outputs. From the input layer the values are sent to nodes of the hidden layers.
- Hidden Layer: Hidden layers apply given transformations to the input values within the network. In them, incoming edges are imported from other hidden nodes or from input nodes connected to their nodes. It is connected by outgoing edges to output nodes or other hidden nodes. In the hidden layer, the actual processing is done through a system of weighted connections. There may be one or more hidden layers in the entire network. The values entering a hidden node are multiplied by the weights. The weighted inputs are then added to produce a single number.
- Output layer: Hidden layers result in an output layer. It receives connections from hidden layers or from the input layer. Returns one or more output values corresponding to the prediction of the response variable. In classification problems, there is usually only one output node. The active nodes of the output layer combine and condition the data to produce the output values.

The next image shows a typical form of an artificial neural network.

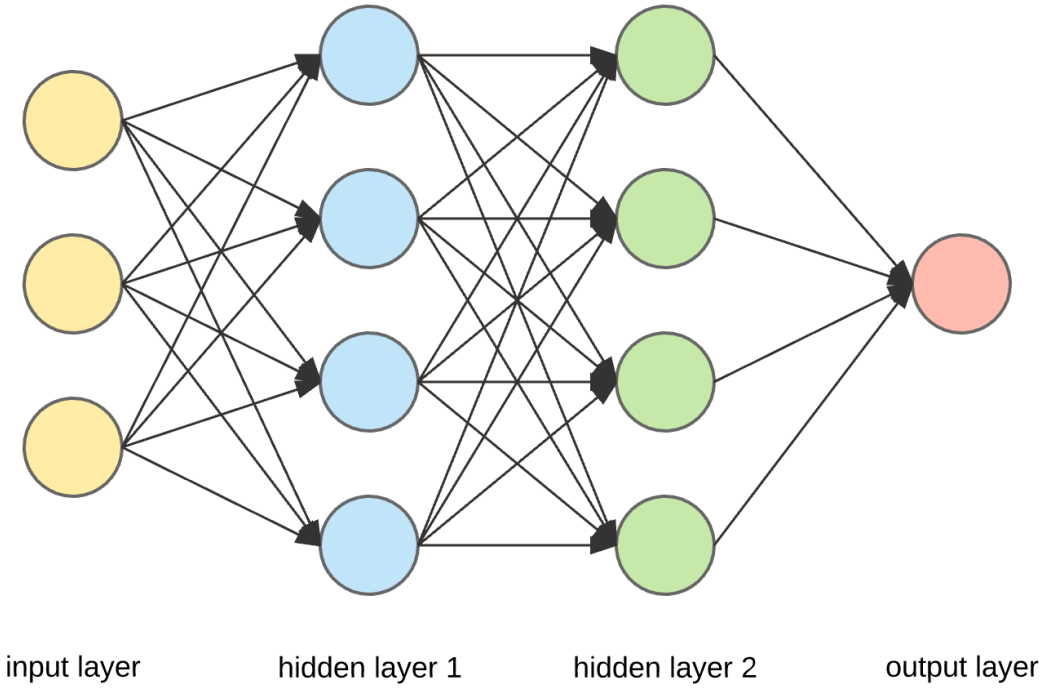


Figure 8: Artificial Neural Network Typical Structure

The ability of the neural network to provide appropriate data handling lies in the correct selection of weights. This feature differentiates it from other forms of data processing. It includes a weight adjustment mechanism and in conjunction with the structure chosen, determines the relationship of the input data to its output. The simplest structure is one in which modules are distributed over two layers: An input layer and an output layer. Each unit at the input layer has a single input and an output that is equal to the input. The output module has all the modules of the input layer connected to its input, with a combination function and a transfer function. By adding 1 or more hidden layers between the input and output layers and the units in that layer, the predictive power of a neural network increases. The number of hidden layers should be as small as possible in order for the neural network to avoid overfitting its output to the specific input elements of the training set. This can happen when the weights make the system learn details of the learning set instead of discovering structures and patterns in it. This may also be due to the fact that the size of the training set is too small relative to the complexity of the model[115].

3 AI in Healthcare

Artificial intelligence and particularly machine and deep learning, subsets of AI, are becoming part of medicine in rapid rhythm. With new technology novelties, the increase in the availability of biological and clinical data is unprecedented and the nature of the data multivariant. Data that represent information at various levels of biological complexity such as biomedical imaging, health records, multiomics, and data from wearable sensors and implantable electronic devices provide a gold mine for researchers to test AI techniques to improve diagnostic and prognostic procedures. This task requires close collaboration between several sciences (engineers, computer scientists, biologists, healthcare professionals) to achieve results free of limitations and to utilize machine and deep learning aspects at its fullest. Providing such tools to specialized healthcare professionals promotes precision, efficiency in the daily workflow and an effective path into how hospitals operate.

In recent years cardiovascular medicine is at the forefront of many AI applications and several comprehensive reviews[1-8,11,12] have been published from 2019 until now. The subject attracts researchers since the benefits and potential impacts are high in volume – cardiovascular diseases are the leading cause of death worldwide, while in 2015 caused 18 million deaths[85] - but they also detect many challenges. Specifically in the field of cardiac electrophysiology, on which this thesis is focused, ML applications have seen a rapid growth in popularity. The use of AI in the interpretation of pulse irregularities and cardiac arrhythmias through the electrocardiogram, a major diagnostic tool, has been approached from various angles which are presented in later chapters.

Even though computer-generated interpretations have been used for many years, they are based on predefined rules. However, AI methods in form of deep neural networks or other type of deep learning algorithms, which are mimicking human-like interpretation of the ECG, provide automated feature, pattern and signature recognition and can detect several details on an ECG which are not always visible by the human eye. The ECG is a low-cost tool and widely available. Because of their digital format, ECGs easy to store and transfer. In conjunction with the advances in computational capabilities, the interpretation of the ECG makes an ideal candidate for deep learning AI applications.[5,17,18]

The AI-enhanced ECG, an outcome of combining the standard ECG with AI, is nowadays the subject of many reviews and studies. While many reviews[1,5,8] depict the characteristics of machine and deep learning and presenting separate studies and novelties in the field, other reviews[6,11] focus on the central tasks of AI-ECG and discuss the possible solutions to key challenges and deficiencies AI-ECG is facing and provide some key insights.

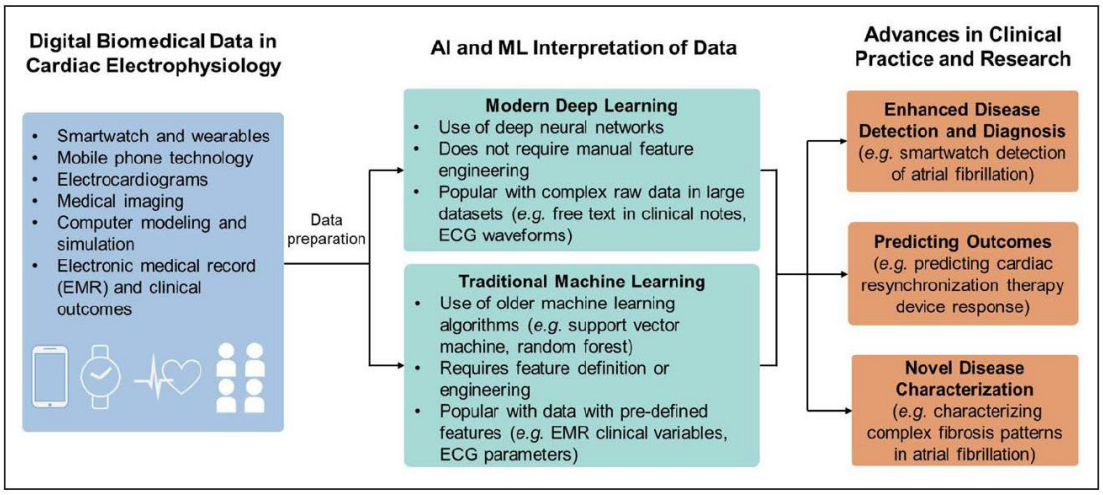


Figure 9 – Overview of artificial intelligence and machine learning in cardiac electrophysiology (source: [8])

Overall, the huge influx of electronic health records and various digitalized medical data, alongside with new techniques to analyze large amounts of data in an efficient way have re-ignited the interest in the field of machine learning in healthcare innovation[2]. Moreover, hardware improvements like the development of GPUs and the availability of powerful computing platforms as cloud computing and new algorithms for tasks with high complexity have developed new technologies in artificial intelligence[14-16].

3.1 ECG and AI-ECG

The electrocardiogram (ECG) was invented in 1901[19] by Willem Einthoven and over the next years, it has become a vital tool for disease identification, risk stratification, and cardiovascular management[9,10]. It also has an established value in the diagnosis, prognosis, and therapeutic monitoring of several cardiovascular diseases. To further extent, the ECG is a ubiquitous tool in the healthcare field, and it's being used by cardiologists and non-cardiologists. It is a low-cost, and simple test that is available even in the most resource-scarce settings. Although the acquisition of the ECG recording is standardized and reproducible in all settings, the manual interpretation of

the ECG, when done by a human, depends greatly on the level of experience and expertise.

From a technical point, the test gives access into the structural condition of the heart and the overall electrical activity but can also provide valuable diagnostic clues for several diseases. Typically, the ECG consists of 12 voltage versus time traces (electrodes) which are collected from the body surface over the heart. A normal ECG tracing shows the P wave, QRS complex, and T wave continuously and regularly repeated in sequence. A sample of a standard 12-lead ECG is illustrated below.

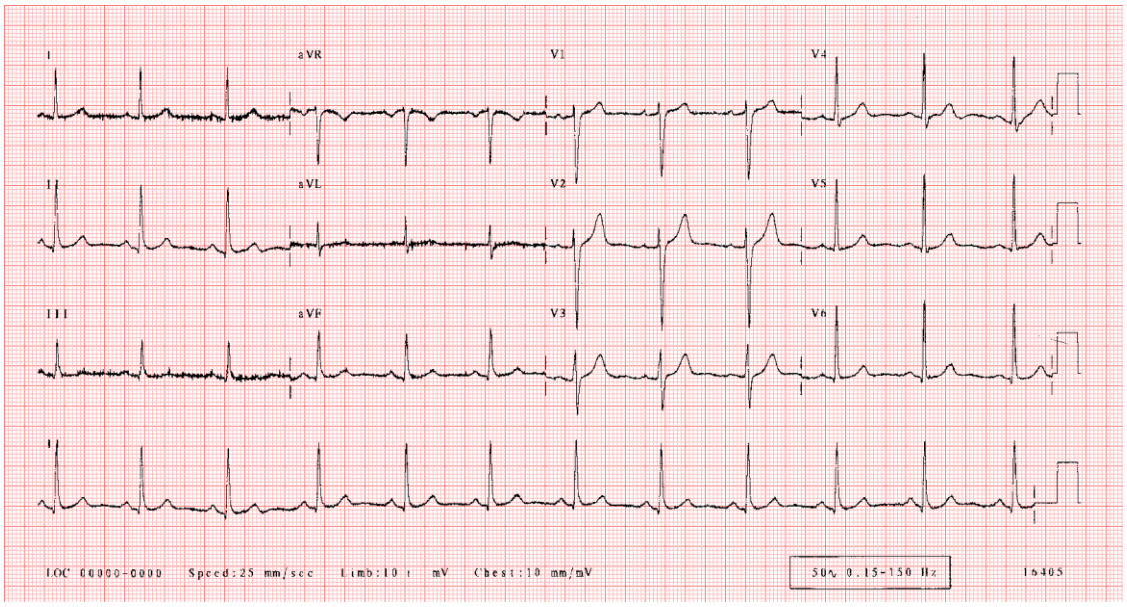


Figure 10 - Normal 12-lead ECG (source: ecglibrary.com)

Computer-generated interpretations are limited on predefined rules and manual pattern or feature recognition algorithms. However, artificial intelligence (AI) has now the ability to analyze the 12-lead ECG and single lead tracings from wearable devices. The ECG interpretation supported by AI has proven to have greater diagnostic accuracy and efficiency than previous traditional interpretations.

AI tools have shown promise in automating and assisting disease diagnosis, and tools are now being developed to enhance prediction of disease prognosis and response to therapeutics and provide novel characterization of health and disease. This advance on automated ECG interpretation is often named as AI-ECG.

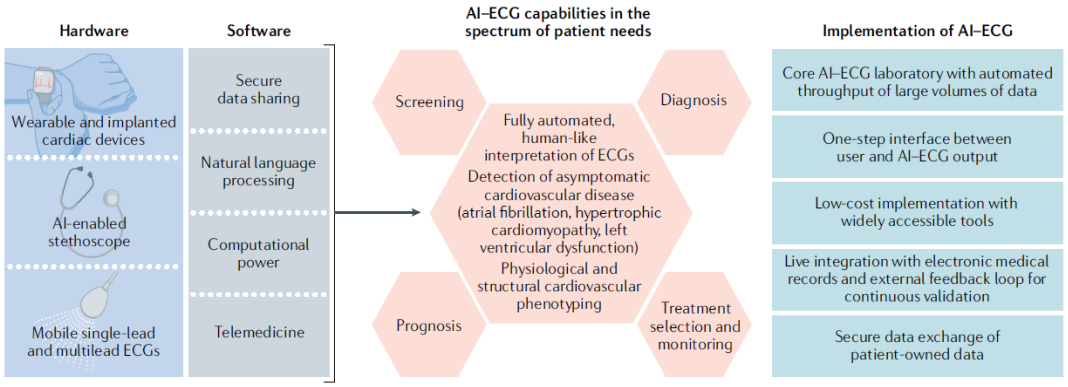


Figure 11 – Framework for AI-ECG in clinical practice (source: [5])

3.1.1 Types of the ECG

There are three main types of electrocardiograms (ECGs), the standard 12-lead ECG, also known as resting ECG, the exercise ECG and the Holter monitor. The types are further discussed on this chapter with the addition of single-lead ECGs which are popular in recent studies for several diseases and arrhythmia detection.

Resting 12-lead ECG

An ECG test is typically performed using 12-lead ECG, which is standard for hospital usage. Also called as "the standard 12-lead ECG", this system simultaneously records 12 different signals. The standard 12-lead ECG system simultaneously captures the electrical signals of the heart from frontal plane (Limb leads), horizontal plane (precordial leads), respectively, from different vectors, so 12 different shapes of P-wave, QRS complex, and T-wave are observed. Further analysis of the innerworkings of a 12-lead ECG will be provided on chapter 3.1.2.

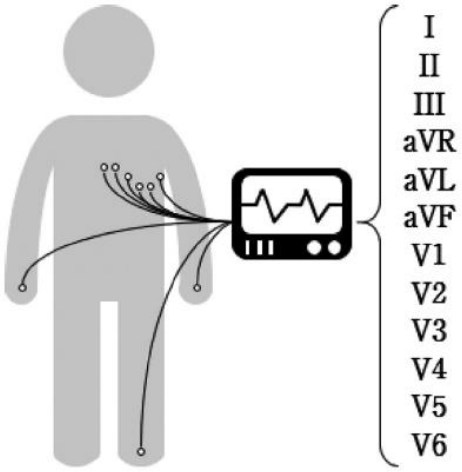


Figure 12 - Illustration of the spatial angles of the 12 single-lead ECG system (source: [21])

Single-Lead ECG

In recent years, wearable devices such as the Apple Watch[22] and Zio Patch cardiac monitor[17,23] have been introduced in the market to detect irregular heartbeats. The main difference is that these types of devices utilize only one single-lead ECG (like lead II) data instead of employing the conventional standard 12-lead ECG system. Thus, in most cases, rather than using all the ECG information, newer systems are being developed to perform ECG signal classification based on single-lead ECG output.

Exercise ECG

Also called a stress test, this ECG monitors the heart's capabilities and activity under physically demanding conditions, such as exercise. Usually, this test is done in controlled environments with the patient hooked up to an ECG and asked to walk or pedal in training equipment (treadmill, stationary bike). For about 10 to 20 minutes, the intensity of the exercise is gradually increased. In order for this test to provide clear and unbiased results, patients may be asked to alter their medication routine beforehand. This allows the ECG to record the heart without any external factors that may otherwise impede or improve its base performance.

Holter Monitor

Holter monitoring is a non-invasive means of continuously recording an electrocardiogram. Using adhesive-backed electrodes connected to a monitor (illustrated in figure 5), the device will record any irregularities that may not be picked up during shorter ECG tests, since the symptoms of a cardiac patient are often transient. Holter monitors are capable of disseminating any heat built up and can continuously record ECG, allowing the assessment of the heart rate and rhythm during normal activities such as rest or exercise. Indications for its use include the detection and assessment of arrhythmias, investigation of intermittent collapse and exercise intolerance, while helping cardiologists to determine the efficacy of medicines on patients. Since Holter monitoring requires ECG recording for 12 hours to 48 hours continuously, the issue of efficient storage and transmission of ECG data is of great concern[25,26]. An issue which was tackled in a 2021 study[27].

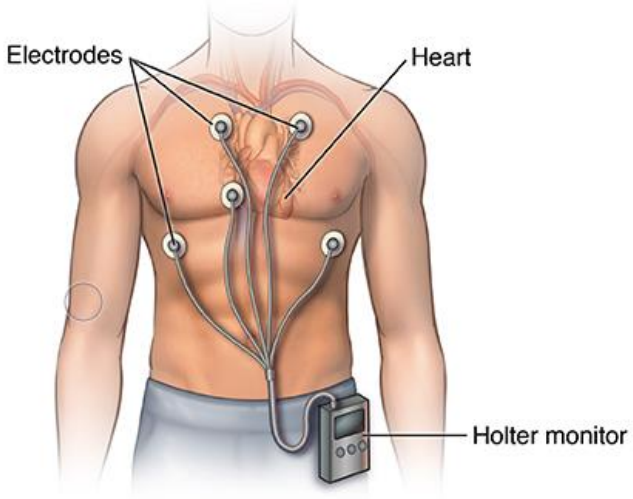


Figure 13 - Holter monitor (source: hopkinsmedicine.org)

3.1.2 Annotation, Modern Tools, and Formats

To systematically approach a 12-lead ECG which a common ECG type used in the research papers and in diagnostics, it's crucial to understand the parts of an ECG and what each segment refers to. In figure 6, there is a visualization of the basic structure of the heart and alongside the figure 2 (ECG sample), this chapter will be a guide into understanding how the annotation procedure works and the meaning behind which segment in the ECG.

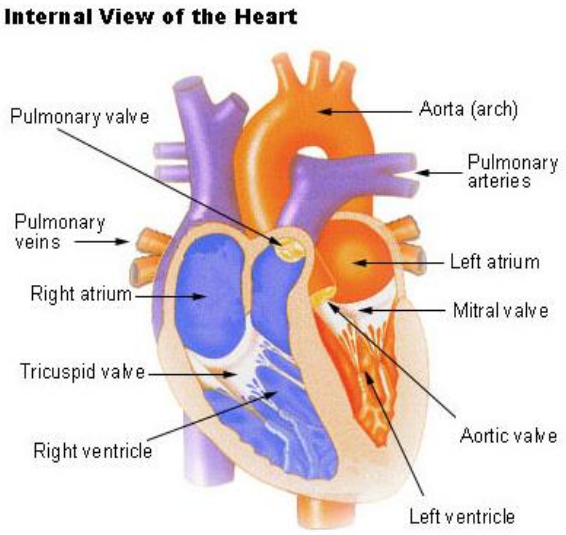


Figure 14 – Internal structure of the Heart (source: https://training.seer.cancer.gov)

An ECG strip is a waveform which consists of different components and together reveal the electrical activity of the heart. The figure 7 is a short example of a normal waveform. The components are the p-wave, pr segment, QRS complex, T-wave and the ST-segment.

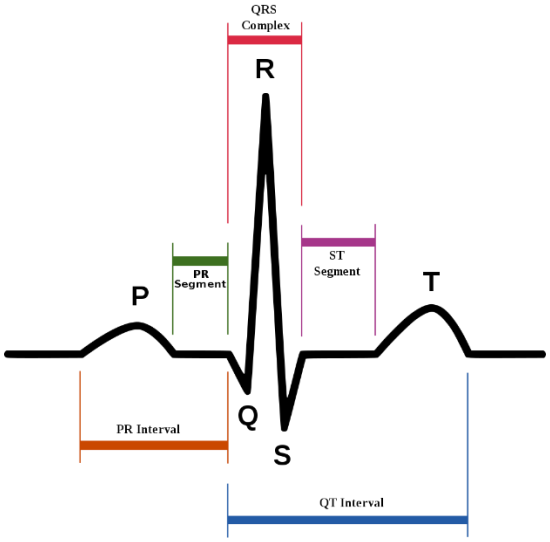


Figure 15 – Sinus Rhythm Waveform (source: https://www.wikipedia.org/)

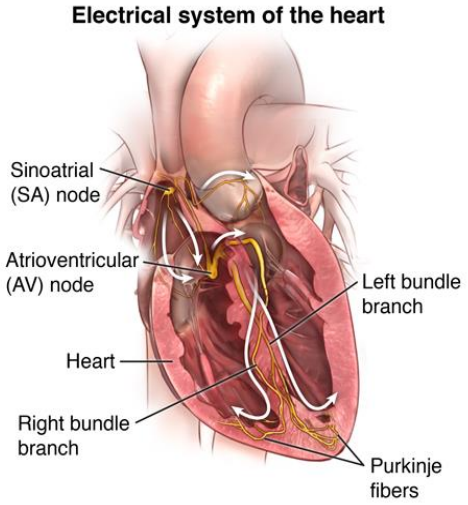


Figure 16 – Parts of the electrical system of the Heart (source: https://www.hopkinsmedicine.org/)

P-wave is the first positive deflection in the ECG. The P-wave is indicative of atrial depolarization. The structure in figure 8 called the SA node in the upper part of the atria intrinsically depolarizes and generates action potentials that spread throughout the atria, that electrical activity that's spreading throughout the atrium from the SA node is represented by the P-wave.

The next part is at the end of the P-wave until the QRS complex, a flat line that's called the PR segment which is the time period where all the electrical activity that is coming from the SA node throughout the atria all converge and come on to the AV node. The AV node gets hit with that depolarizing wave and holds on to the electrical activity. Thus, PR segment is indicative of AV node depolarization.

P-wave which is indicative of atrial depolarization in conjunction with the PR segment which is indicative of av node depolarization form the pr interval which is the time span from which the SA node fires depolarizes the atria, depolarizes the av node and it's just getting ready to send those action potentials down to the ventricles.

Following the PR interval is the QRS complex that is indicative of ventricular depolarization. In this case the ventricles receive the electrical activity then depolarize, forming the QRS complex.

The ST segment is still indicative of ventricular depolarization but there's no electrical activity in a particular net direction. The ventricles are depolarized and they're holding on to that positive charge getting ready to repolarize.

The next positive deflection is called the T-wave. The T-wave is indicative of ventricular repolarization so finally the ventricles go from positive to a negative charge and they create the upward deflection shown in figure 7.

The last part of this sample is the QT interval. The QT interval starts right before Q wave and is indicative of the time period where the ventricles are depolarized and repolarized, an important factor for prolonged qt intervals.

3.1.2.1 Annotation procedure

In order to register the electrical activity of the heart, physicians utilize electrodes placed in different parts of the body, which focus on a particular electrical activity in a very specific portion of the heart.

Overall, a 12-lead ECG has three (3) bipolar leads (I, II and III), three (3) augmented unipolar limb leads (aVF, aVL and aVR) and six (6) precordial chest leads (V_1, V_2, V_3, V_4, V_5 and V_6). These leads are categorized and used to determine the electrical activity for portions of the heart, which are visualized on figure 9.

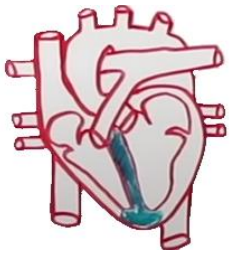
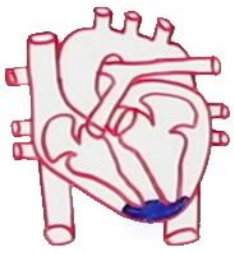
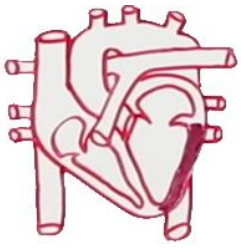
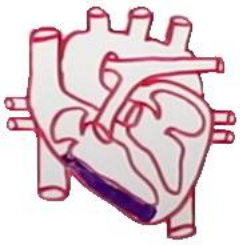
			
Anteroseptal	Inferior Wall	Left Ventricle	Right Ventricle
Lead $V_1/V_2/V_3/V_4$	Lead II/II/aVF	Lead I/aVL/V_5/V_6	Lead aVR/V_1/V_2

Figure 17 – Illustrated areas of the Heart and the leads which are used to determine the electrical activity of the area

Another crucial part for physicians to make a diagnosis from ECG strips is the grid. The grid refers to the background surface of the strip which consists of square boxes. A large box has 5 mm of height and 5 mm of width, containing 25 small square boxes with 1 mm height and 1 mm width. This is crucial because the height refers to amplitude (5 mm = 0.50 mV), meaning the volume in voltage of the detected electrical activity and the width refers to the time (5 mm = 0.20 sec) this electrical activity takes to go through different portions of the heart.

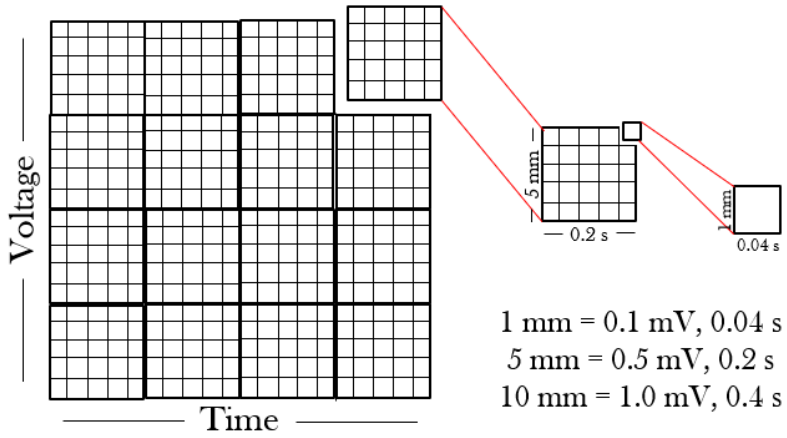


Figure 18 - Grid boxes on an ECG strip

A sample of an ECG with sinus rhythm is depicted in figure 2. The 12-lead ECG consists of all the aforementioned leads. An important exception is the lead II in the bottom of the ECG strip which continues through the duration of the ECG strip. The procedure of the annotation is based on five (5) checkpoints; rate, rhythm, QRS complex, sinus P-waves and P-R interval. The physician examines each point as follows:

1. Rate | Too fast, too slow or normal
    - Beats per minute over 100 means tachycardia
    - Beats per minute under 100 means bradycardia
    - Beats per minute in range of 60-100 means normal

    The rate is calculated by the ECG machine, box method or/and R-waves (x6).

2. Rhythm | Regular or irregular
    - If the R-R interval is constant, then rhythm is regular
    - If the R-R interval is not constant, then rhythm is irregular

3. QRS Complex | Narrow or wide
    - If the duration of the QRS is over 0.12 seconds then is labelled as wide
    - If the duration of the QRS is under 0.12 seconds then is labelled as narrow
4. Sinus P-waves | Present or not present
    - Upright P-wave in lead II
    - Inverted P-wave in aVR
    - Every P-wave followed by QRS in lead II (rhythm strip)

    If all of the above conditions are met, then the rhythm is labelled as sinus.

5. P-R interval | Labelled as:
    - Normal
    - Prolonged
    - Constant
    - Variable
    - Progressively longer

Based on the above systematic approach, the physician will label the ECG strip and determine the diagnosis.

3.1.2.2 Modern tools for ECG annotation

Modern ambulatory ECG monitors allow for up to 30 days of continuous monitoring, producing far too much data for physicians to comprehensively analyze. For this reason, service providers are commonly used to annotate ECG recordings and create reports that summarize and highlight ectopic activity. These reports provide clinical decision support for prescribing physicians. Service providers rely on certified technicians and supporting algorithms to process and annotate the data from ECG monitoring studies. Historically, supporting algorithms have achieved levels of performance well below that of humans[29] but high enough to be used for prioritization and conservative filtering of ECG as it is queued for human interpretation.[30]

<u>BeatLogic platform</u>

The Preventice BeatLogic platform is a comprehensive ECG annotation platform that leverages Deep Learning for beat and rhythm detection and classification. Annotations were made in accordance with standard practice by a dedicated team of Certified

Cardiographic Technician (CCT)-certified ECG technicians having experience ranging from 9–30 years.

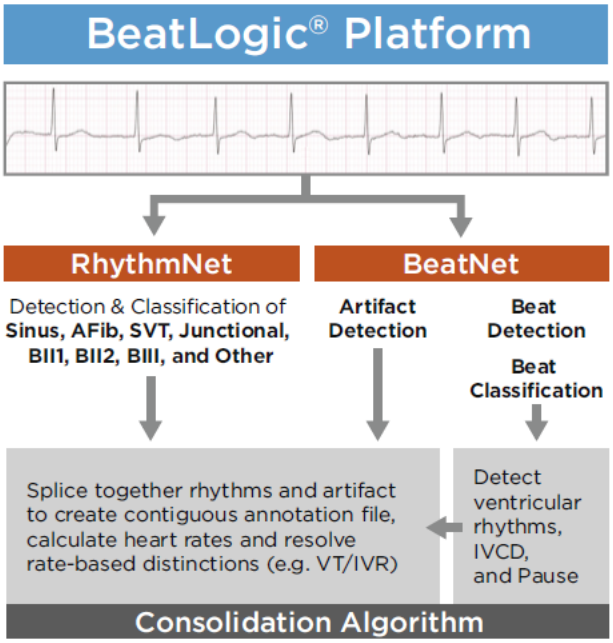


Figure 19 – BeatLogic platform flowchart (source: [28])

Human-machine integration ECG intelligent annotation system

A group of researchers in 2021 developed an intelligent ECG-assisted annotation system, that not only supplements labelled data, but also significantly reduces the workload compared with manual annotation. Since beat annotation is the most basic and important part, a GAN-based generation model that can generate 14 types of simulation beats and a CNN-based beat pre-annotation model were proposed. The experimental results showed that the simulation beat has high similarity to real beat and the accuracy of the pre-annotation model on the test set of 14 classes of beats is 99.28%. The proposed ECG intelligent annotation system's self-learning mechanism could improve pre-annotation performance and annotation efficiency by generating more labelled data.

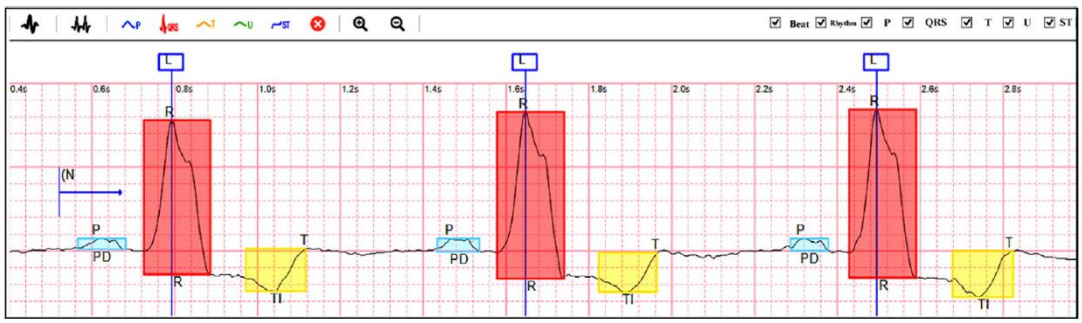


Figure 20 – Annotation interface of the annotation system developed by this research paper (source: [31])

WaveformECG

In a study originated from the Johns Hopkins University published a research paper[32] in 2016, researchers developed an open source platform as a tool for analyzing, visualizing and annotating the ECGs. At that time, researchers detected a lack of open, noncommercial platforms for managing and analysis ECG data. As an outcome, they developed the WaveformECG platform. WaveformECG is a Web-based tool developed as part of the CardioVascular Research Grid (CVRG).[86] The platform can import ECG data in several different formats, including Philips XML 1.03/1.04, HL7aECG, Schiller XML, GE Muse/Muse XML 7+, Norav Raw Data (RDT), and the WaveForm DataBase (WFDB) format used in the Physionet Project.[87]

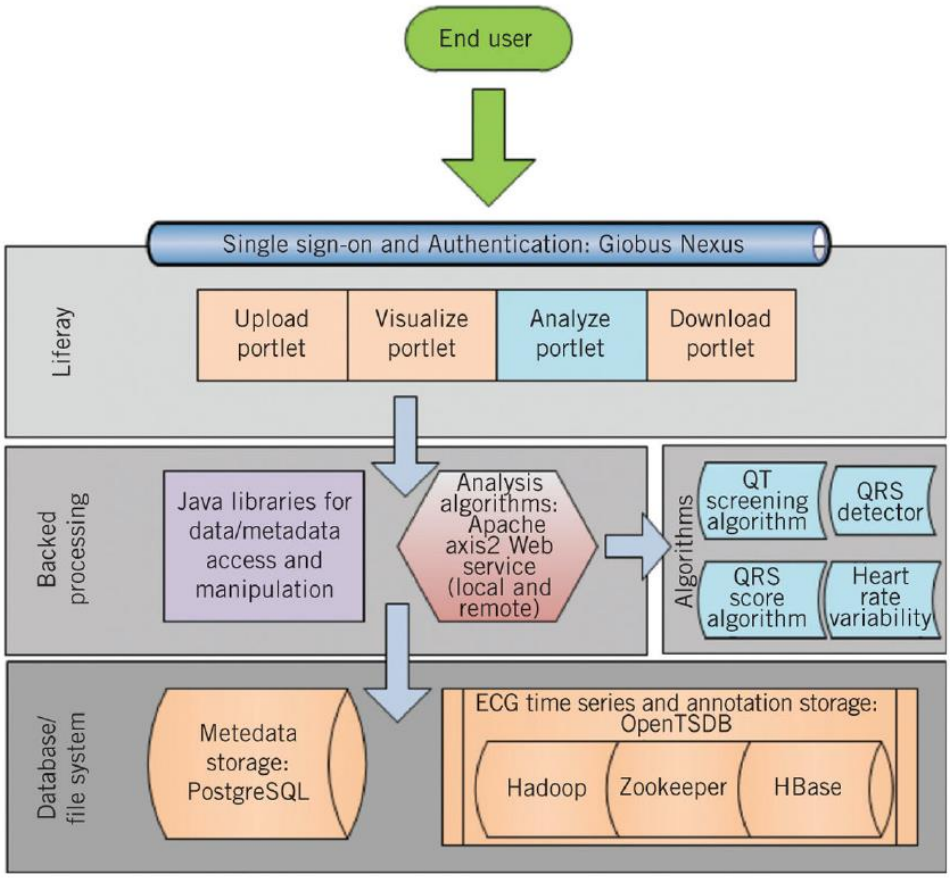


Figure 21 – Waveform's architecture (source: [32])

3.1.2.3 ECG Formats

In many cases, the ECGs are presented and saved as an ECG file with .ecg extension. Opening this type of file, requires specialized software to visualize the context of the file and many of those provide the tools to annotate the ECG with markers. The markers are then saved in json format which includes, ECG_ID, win_start_time, reviewer_id, start_time, episodes with number of each episode, rhythm_name and rhythm_code and

lastly the onset and offset of the annotated rhythm. Files may vary based on the software which the cardiologists or specialized technicians use.

3.1.3 Available datasets

Researchers, while developing studies and testing their methods on applying AI on the ECG, often use public datasets or closed ones from organizations who ultimately fund the studies. Some of the datasets are presented below and can be referenced in two important reviews from Nature Reviews (Cardiology)[5] and the Journal of Electrocardiology (2019)[11].

Public datasets:

1 – MIT-BIH Arrhythmia (MITDB)[40,41]: this dataset includes 48 half-hour excerpts of two-channel ambulatory ECG recordings (sampling frequency: 360 samples per second) that was annotated by two or more cardiologist

2 – Creighton University Ventricular Tachyarrhythmia (CUDB)[40,42]: eight-minute ECG recordings of 35 patients who experienced episodes of sustained ventricular flutter, ventricular tachycardia, and ventricular fibrillation are provided in this dataset (sampling frequency: 250 samples per second).

3 – MIT-BIH Atrial Fibrillation (AFDB)[40,43]: this database contains 25 long-term two-channel ECG recordings (10 h) of patients with atrial fibrillation (sampling frequency: 250 samples per second).

4 – PhysioNet/CinC Challenge 2017[40,44]: single-channel short ECG recording and associate human annotations for 8528 and 3658 human subjects in public training set and blind test set is available for this dataset (sampling frequency: 300 samples per second)

Additional major contemporary ECG databases:

5 – Telehealth Network of Minas Gerais[45]: this database contains 12-lead ECGs from 1,676,384 patients from Brazil (2010-2018), aiming at automated ECG interpretation

6 – Mayo Clinic[18,46-50,52]: this database contains 12-lead ECGs from 449,380 patients from USA (1994-2017), aiming at automated ECG interpretation with studies on hypertrophic cardiomyopathy, silent atrial fibrillation, left ventricular dysfunction, age, sex, and race/ethnicity and serum potassium level detection

7 – Geisinger[53]: this database contains 12-lead ECGs from 253,397 patients from USA (1984-2019), aiming at overall survival

8 – Huazhong University, Wuhan[54]: this database contains 12-lead ECGs from 71,520 patients from China (2012-2019), aiming at automated ECG interpretation

9 – iRhythm Technologies/Stanford University[17,23]: this database contains single-lead, ambulatory ECG monitoring using a device called Zio Patch from 53,549 patients from USA (2013-2017), aiming at classification of 12 rhythm types

10 – University of California, San Francisco[55]: this database contains 36,186 12-lead ECGs from USA (2010-2017), aiming at detecting early diastolic mitral annulus velocity, pulmonary arterial hypertension, left ventricular mass, left atrial volume, amyloidosis, hypertrophic cardiomyopathy, and mitral valve prolapse

11 – Health eHeart Study[56]: this database contains single-lead, smartwatch-based ECGs from 9,750 patients with multinational background (2016-2017), aiming at detection of atrial fibrillation

12 – China Physiological Signal Challenge 2018[57]: this database contains 12-lead ECGs from 6,877 patients from China (2018), aiming at classification of 9 rhythm types

13 – Cleveland Clinic[58]: this database contains 12-lead ECGs from 946 patients from USA (2003-2012 and 2017-2018), aiming at the response to cardiac resynchronization therapy

3.1.4 AI applications using ECG

Hyperkalemia and hypokalemia are clinically silent, common in patients with renal or cardiac disease and are life threatening. Even though treatments for hyperkalemia are effective and available[25], the diagnosis particularly outside hospitals or clinic, proves to be challenging because patients are often asymptomatic. Moreover, guideline-directed blood potassium monitoring which can detect hyperkalemia is severely underperformed. In this chapter there are studies which tackle the issue by developing a noninvasive method for tracking potassium through the interpretation of ECG, an important step into clinical advance. A sample of hyperkalemia kai hypokalemia are illustrated in figures 14 and 15.

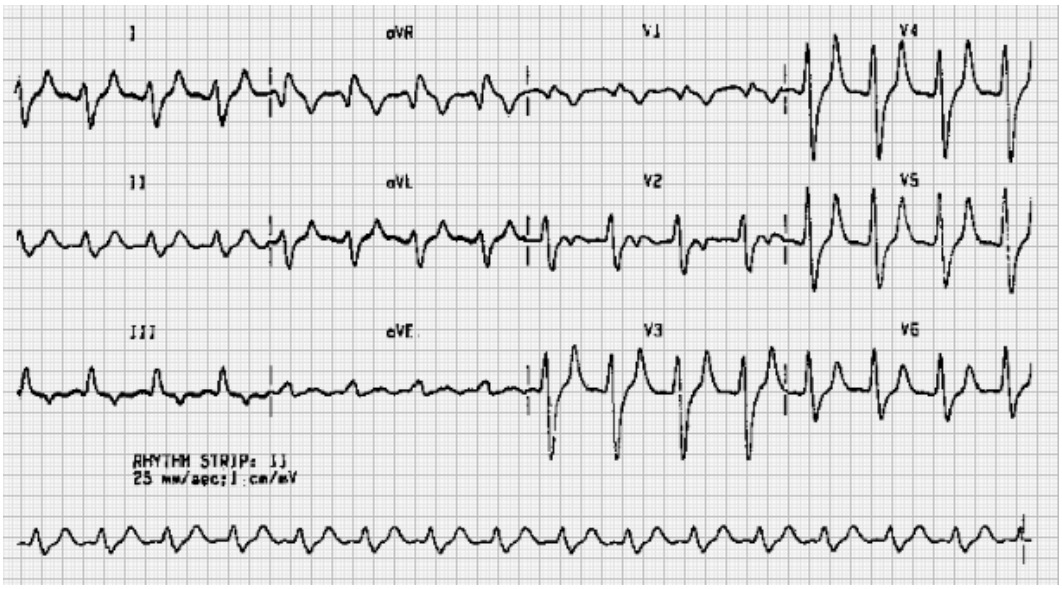


Figure 22 - Sample of hyperkalemia (source: https://ecglibrary.com/)

For this issue, a study which was published in 2016[59], developed a novel bloodless **potassium** determination using single-lead ECGs. The researchers utilized two groups of hemodialysis patients (development group, n=26; validation group, n=19) and collected processed single-channel ECGs from each patient, based on which they made accurate predictions (absolute error of 0.44±0.47 mmol/L for the training group and 0.5±0.42 for the validation set) on potassium levels for each patient.

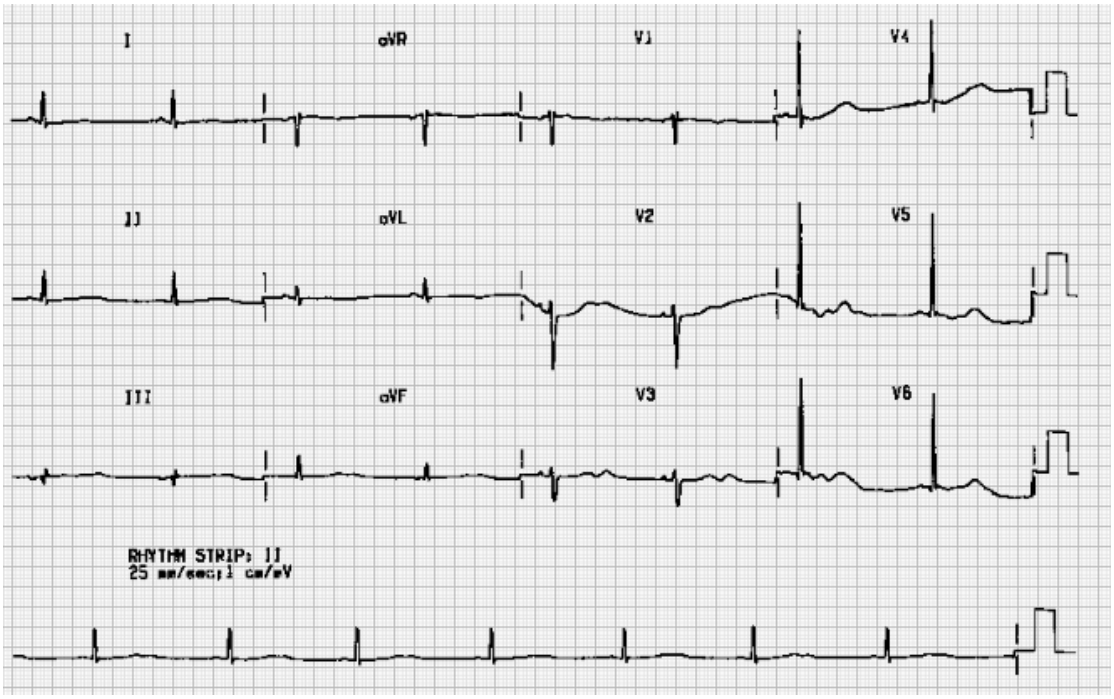


Figure 23 - Sample of Hypokalemia (source: https:/ecglibrary.com)

In 2019 another study[50] on **hyperkalemia** was published in connection to patients with Chronic Kidney Disease (CKD)[60,61]. With more than 1.6 million ECG tracings, the researchers developed a deep learning model with a high AUC of 0.853 to 0.901 for identifying hyperkalemia among patients with CKD from 2 or 4 ECG leads.

In 2018 researchers published a study[63] aiming to test the application of a convolutional neural network (CNN) to assess morphological changes on the ECG in relation to **dofetilide plasma concentrations**. Dofetilide is an antiarrhythmic medication for rhythm control in atrial fibrillation but can potentially lead to significant risk of pro-arrhythmia and requires meticulous dosing and monitoring. To achieve their goal, researchers utilized publicly available ECGs and plasma drug concentrations from 42 healthy subjects who received dofetilide or placebo in a controlled-setting clinical trial. Furthermore, the CNN was developed to predict dofetilide plasma concentration in 30 subjects and then tested the model in the remaining 12 subjects. On the last stage of the study, they compared the deep learning approach to a linear model based only on QTc, which resulted in the deep learning model to achieve better correlation to dofetilide levels ($r = 0.85$), proving how valuable deep learning techniques can be.

AI models can also prove to be a tool for predicting **age and sex** of patients, using exclusively 12-lead ECGs. Age and sex are known to affect the ECG[64-66], and prior work was mainly based on statistical analysis and on specific feature extraction[66]. In a

published study[51] in 2019, researchers trained a CNN using 10-second samples of 12-lead ECG signals from 499,727 patients and then tested on a separate cohort of 275,056 patients. For sex classification, the model showed 90.4% accuracy with an area under the curve of 0.97 in the independent test data. Moreover, age was estimated as a continuous variable with an average error of 6.9±5.6 years (R-squared = 0.7).

A different approach, free of the concept of Neural Networks (NNs), is a method to automate ECG analysis for similar ECGs that already have a physician interpretation, a study[67] which was published in 2019. The method and the process are presented at Journal of Electrocardiology (2019)[67] and illustrate on the figure 16 below. The researchers developed an algorithm to automatically interpretate 12-lead ECGs by database comparison.

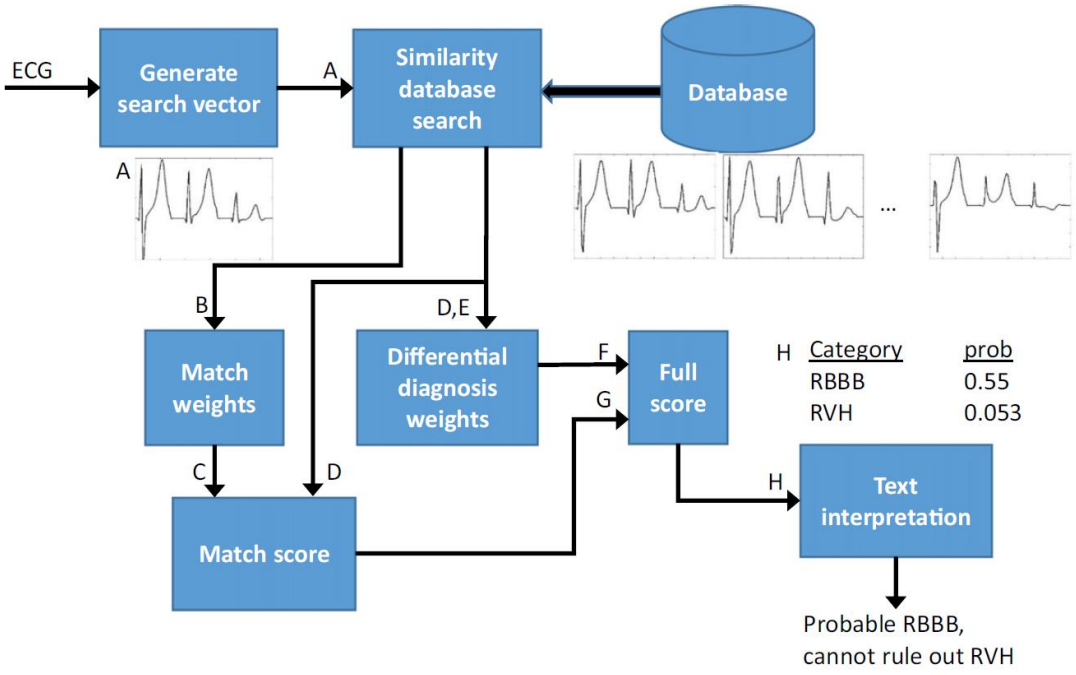


Figure 24 - Block diagram of algorithm for ECG interpretation by similar ECG (source: [X5])

Hypertrophic cardiomyopathy (HCM) is among the leading causes of sudden cardiac death among adolescents and young adults and is associated with significant morbidity in all age groups[68]. Over 90% of patients with HCM have electrocardiographic abnormalities[69] and 12-lead ECG may offer a noninvasive, low-cost, and rapid means of screening for the condition. During a study[49] in 2020, researchers trained and validated a CNN using digital 12-lead ECG from 2,448 patients with verified HCM diagnosis and 51,153 non-HCM control subjects. For testing the ability of the network to detect HCM a different dataset of 612 HCM and 12,788 control subjects was used.

The results are depicted in figure 17, proving that it's feasible for an AI-based network to detect HCM based on the standard, 12-lead ECG with high diagnostic performance.

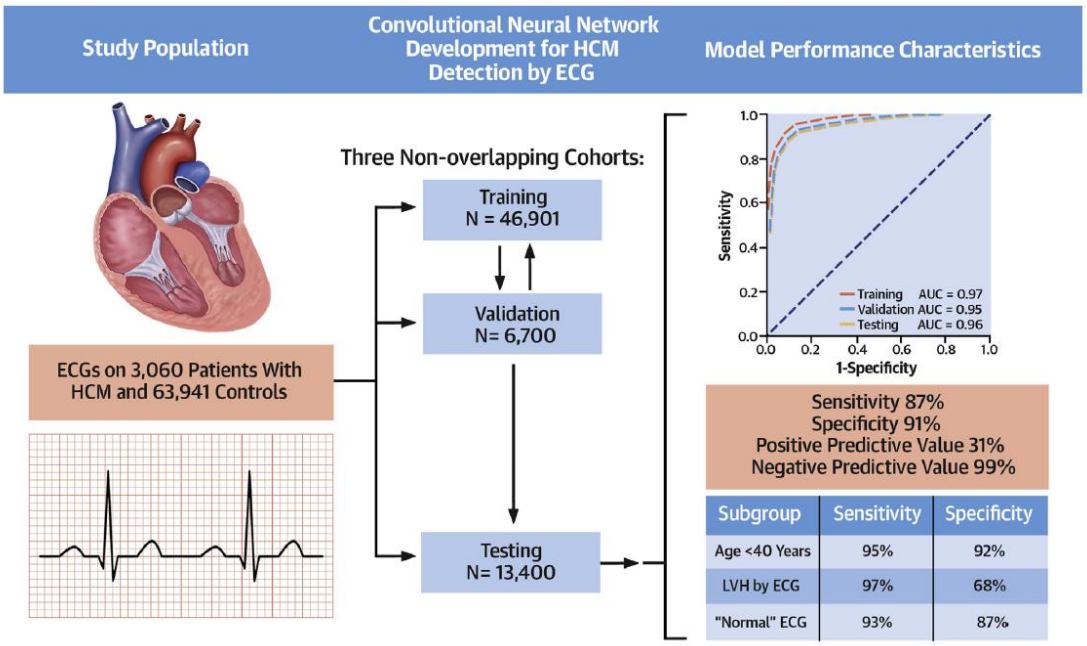


Figure 25 – Training, Validation, and Testing of an AI-Based Electrocardiography Screen for Hypertrophic Cardiomyopathy developed on study 'Detection of Hypertrophic Cardiomyopathy using a CNN-enabled ECG [49]

For **cardiac amyloidosis** (CA) – a life threatening disease - detection, researchers developed an AI-based tool to screen standard 12-lead electrocardiograms (ECGs), in a recent study[70] (2021). They collected 12-lead ECG data from 2,541 patients at Mayo Clinic between 2000 and 2019. After training a developed deep neural network to predict the presence of CA they also experimented using single-lead and 6-lead ECG subsets. Among other metrics, the AUC was 0.91 (CI, 0.90 to 0.93), with a positive predictive value for detecting either type of CA of 0.86. This also proves that the developed AI-driven ECG model can effectively detect CA leading to an early diagnosis.

In 2021 a study[71] on **Left Ventricular Systolic Dysfunction** (LVSD) was presented by Mayo Foundation researchers, who randomly selected a cohort of 2,041 subjects from Olmsted County, Minnesota. They acquired ECGs from the patients and assessed the performance of the AI-ECG to identify LVSD, which resulted in AUC = 0.97 (sensitivity 90%, specificity 92%) while proving that AI-augmented ECG can identify preclinical LVSD and warrants further study as a screening tool.

Studying the economic advantages of AI-driven ECG for detecting **Asymptomatic Left Ventricular Dysfunction** (ALVD), researchers at Mayo Foundation, evaluated an algorithm under various clinical and cost scenarios when used for screening patients at the age of 65.[72] They concluded that the novel AI-ECG which was developed appears to be cost-effective under most of the clinical scenarios but the study needs further examination and external validation of the AI-ECG.

In 2021 researchers tested a developed AI-ECG as a tool for screening patients with **dilated cardiomyopathy (DC)**,[73] which can be asymptomatic or present as sudden cardiac death. They used standard 12-lead ECGs as input data from 421 DC cases and 16,025 control subjects. As an outcome, the AI-ECG demonstrated high sensitivity and negative predictive value for detection of DC, resulting in a simple and cost-effective screening tool.

In 2021 another study[74] was published focused on detecting aortic stenosis using an AI-ECG developed on a convolutional neural network. The **aortic valve stenosis (AS)** occurs when the heart's aortic valve narrows, which then limits or blocks blood flow from the heart into aorta and to the rest of the body, which may lead to heart failure. During the study, 258,607 adults underwent through examination specifically with an echocardiography and an ECG performed within 180 days and stored in the Mayo Clinic database (1989-2019). Ultimately, the researchers underlined the usefulness of an AI-ECG as a screening tool for identifying patients with moderate or severe AS.

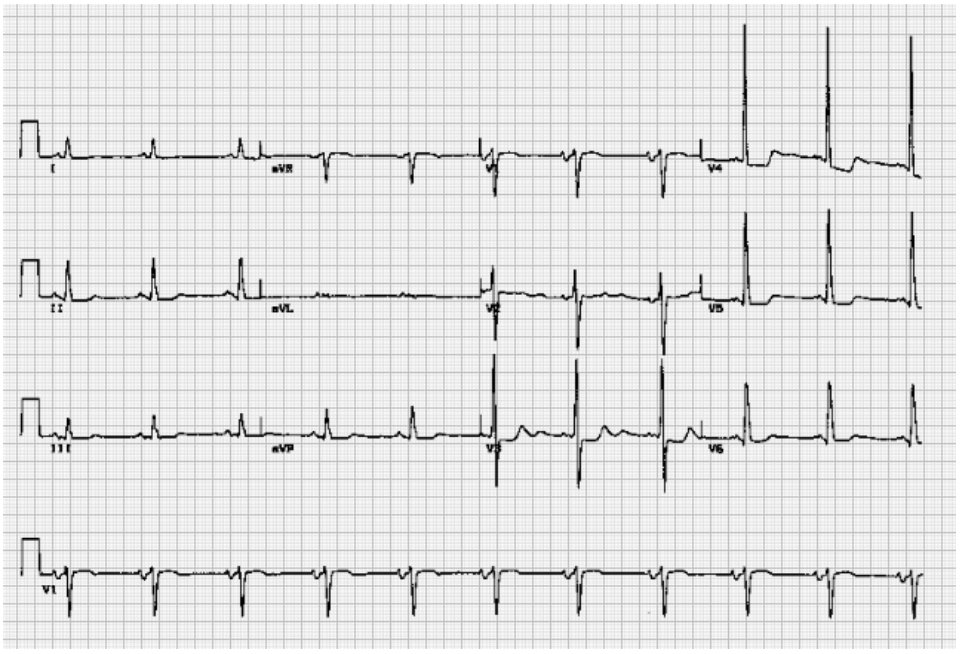


Figure 26 - ECG from an 83 year old man with aortic stenosis (source: https://ecglibrary.com)

3.2 Arrhythmia

Cardiac arrhythmias refer to an irregular or abnormal heartbeat. Heart rhythm issues, heart/cardiac arrhythmias, occur when the electrical signals that coordinate the heart's beats have irregular behavior. The faulty signaling causes the heart to beat too fast (tachycardia), too slow (bradycardia) or irregularly. Although cardiac arrhythmias fluctuations are recorded daily due to exercise or during sleep, some cardiac arrhythmias may cause life-threatening symptoms or be linked to life-threatening diseases.

In US alone, more than four million people are affected by cardiac arrhythmias summing up an annual healthcare cost of up to $67.4 billion, a substantial economic burden[88].

3.2.1 Types of arrhythmias

In general, cardiac arrhythmias are grouped by the speed of the heart rate:

- Tachycardia is a fast heartbeat, with resting heart rate greater than 100 beats a minute. A sample of sinus tachycardia is illustrated in figure 19.

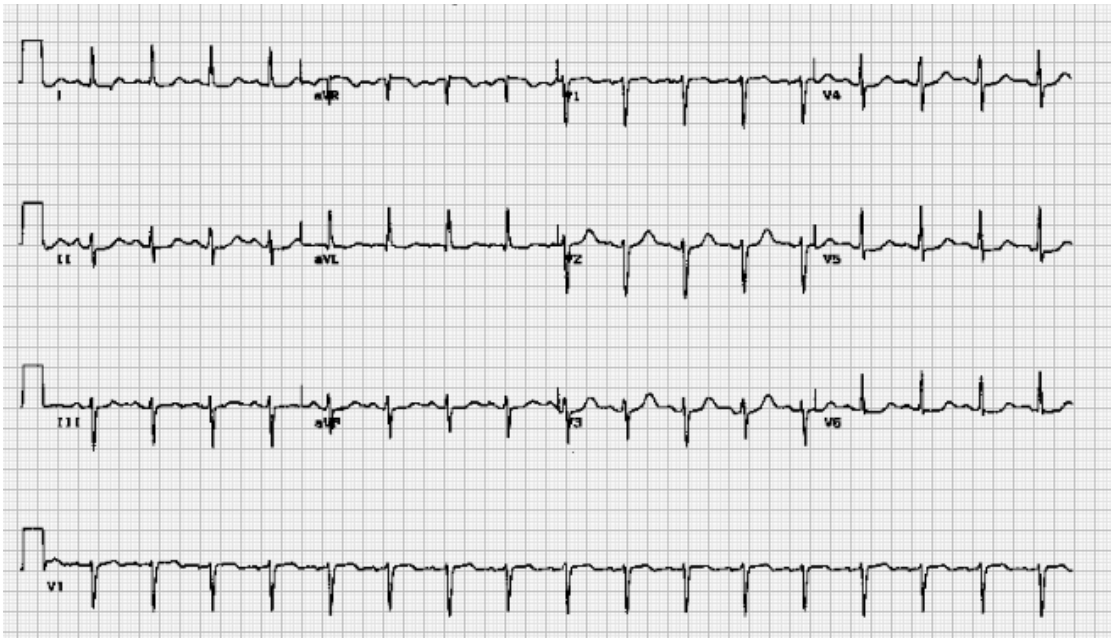


Figure 27 - Sinus tachycardia (source: https://ecglibrary.com)

- Bradycardia is a slow heartbeat, with resting heart rate less than 60 beats a minute. Despite that, a low resting heart rate doesn't always signal a problem. Factors like physical condition, slow heart rate and how much blood is the heart pumping through the body in combination can lead to bradycardia. A sample of sinus bradycardia is illustrated in figure 20.

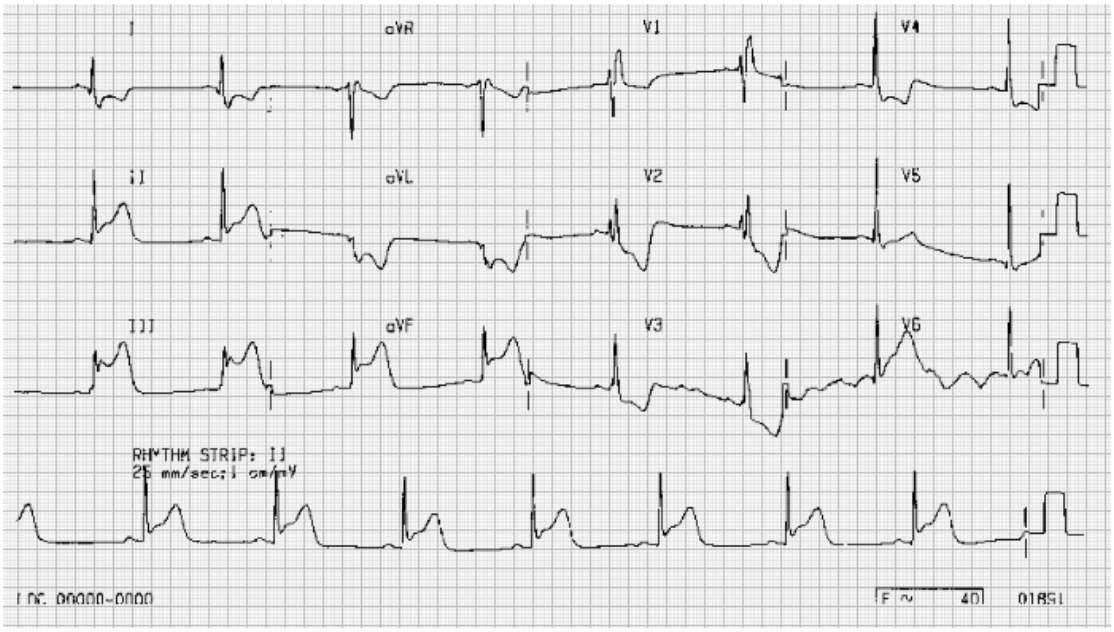


Figure 28 - Sinus bradycardia (source: https://ecglibrary.com)

Tachycardias include:

Atrial Fibrillation (AF or A-Fib)

AF refers to a chaotic heart signaling which causes a rapid and uncoordinated heart rate. AF is a common, frequently asymptomatic, and thus undetected arrhythmia or even clinically silent[34]. It is associated with an increased risk of stroke[35], heart failure, and mortality[34,35]. In the United States alone, AF is the most diagnosed clinically significant cardiac arrhythmia[36], with a lifetime risk as high as 1 in 3[37]. It has become one of the most important public health problems and a significant factor in increasing healthcare costs worldwide[38]. Detecting this cardiac arrhythmia is a challenge since it is often fleeting while the existing screening methods require prolonged monitoring. The monitoring is mostly relied on the electrocardiogram (ECG) as it offers a simple and low cost solution, but due to its nature there are diagnostic delays since the ECG can appear normal between episodes. As a result, predicting AF has been a longstanding research and clinical priority.

Atrial Flutter

A cardiac arrhythmia similar to AF, but the heartbeats are more organized and similarly to AF is also linked to risk of stroke.

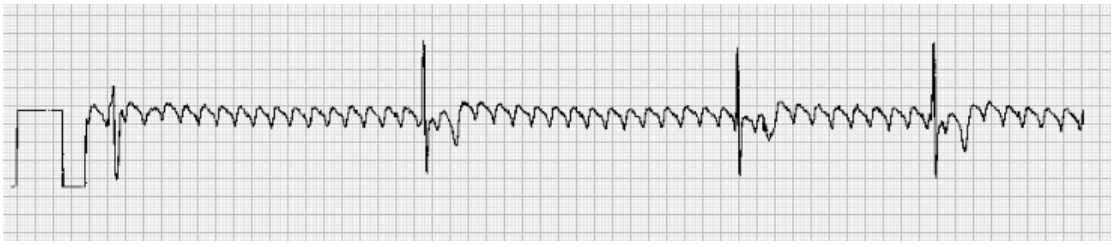


Figure 29 - Atrial Flutter (source: https://ecglibrary.com)

Supraventricular tachycardia (ST)

According to Mayo Clinic[S1], ST refers to arrhythmias that start above the lower heart chambers (ventricles) and causes episodes of a pounding heartbeat (palpitations) that begin and end abruptly.

Ventricular Fibrillation (VF)

According to Mayo Clinic[S1], this type of arrhythmia occurs when rapid, chaotic electrical signals cause the lower heart chambers (ventricles) to quiver instead of contacting in a coordinated way that pumps blood to the rest of the body. This serious issue can potentially lead to death if a normal heart rhythm isn't restored within minutes. Most people who have VF have an underlying heart disease or have experienced serious trauma.

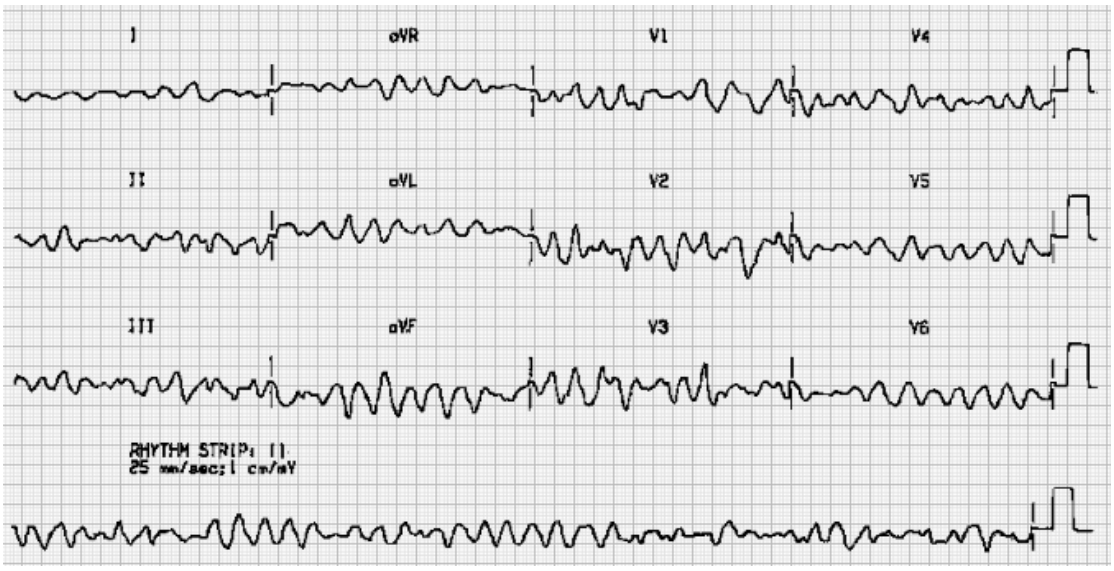


Figure 30 - Ventricular Fibrillation (source: https://ecglibrary.com)

Ventricular tachycardia (VT)

According to Mayo Clinic[S1], this rapid, regular heart rate starts with faulty electrical signals in the lower heart chambers (ventricles). The rapid heart rate doesn't allow the ventricles to properly fill with blood. As a result, the heart can't pump enough blood to

the body. VT may not cause serious problems in people with an otherwise healthy heart. In those with heart disease, ventricular tachycardia can be a medical emergency that requires immediate medical treatment.

Long QT

While this is not an arrhythmia, it can predispose someone to have one. According to Cleveland Clinic[S2], the QT interval is the area on the ECG that represents the time it takes for the heart muscle to contract and then recover, or for the electrical impulse to fire and then recharge. When the QT interval is longer than normal, it increases the risk for 'torsade de pointes', a life-threating form of ventricular tachycardia. In addition, a linked disease which is the subject of several research papers is Long QT syndrome (LQTS), which is characterized by prolongation of the QT interval 1 and is associated with an increased risk of sudden cardiac death. However, although QT interval prolongation is the disease's hallmark feature, approximately 40% of patients with genetically confirmed LQTS have a normal QTc at rest (concealed LQTS)[39]. A sample of Long QT interval is illustrated in figure 23.

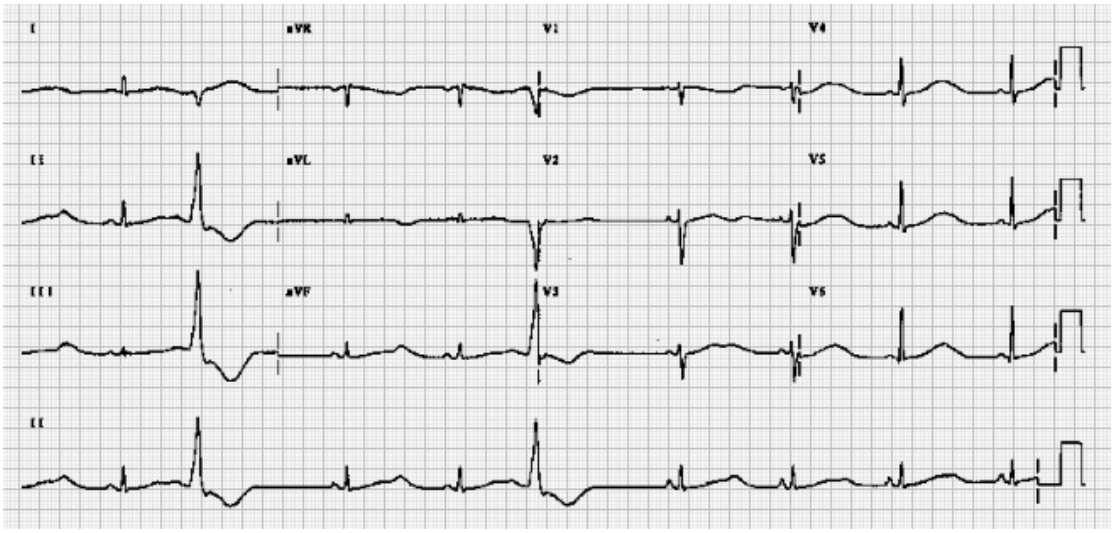


Figure 31 - Long QT interval (source: https://ecglibrary.com)

Bradycardias include:

Sick sinus syndrome

According to Mayo Clinic[S1], this sinus node is responsible for setting the pace of the heart. If it doesn't work properly, the heart rate may alternate between too slow (bradycardia) and too fast (tachycardia). Sick sinus syndrome can be caused by scarring

near the sinus node that's slowing, disrupting or block the travel of impulses. Sick sinus syndrome is most common among older adults.

Conduction block

According to Mayo Clinic[S1], a block of the heart's electrical pathways can cause the signals that trigger the heartbeats to slow down or stop. Some blocks may cause no signs or symptoms, and others may cause skipped beats or bradycardia.

3.2.2 AI applications on arrhythmia detection

In 2019 a team of researchers developed an AI–enabled electrocardiograph (AI-ECG) using a deep learning model to detect the electrocardiographic signature of **atrial fibrillation** present during normal sinus rhythm, using standard 10-second, 12-lead ECG.[48] The dataset consisted of a large cohort of patients (180,922 patients with 649,931 normal sinus rhythm ECGs and atrial fibrillation detected by trained personnel under cardiologists' supervision) from the Mayo Clinic Digital Data Vault, while the researchers allocated ECGs to the training, internal validation, and testing datasets in a 7:1:2 ratio. The AI model was based on a convolutional neural network (CNN) using the Keras Framework with a TensorFlow backend and Python[89]. This study resulted in the ability of the AI-enhanced ECG to identify patients with atrial fibrillation with success based on metrics, which were used to mathematically assess the performance of the model, area under the curve (AUC=0.90) of the receiver operating characteristic (ROC) curve, and other metrics like sensitivity (82.3%), specificity (83.4%), F1 score (45.4%) and overall accuracy (83.3%).

	AUC	Sensitivity	Specificity	F1 score	Accuracy
Main analysis	0.87	79.0%	79.5%	39.2%	79.4%
	(0.86-0.88)	(79.0-79.9)	(79.0-79.9)	(38.1-40.3)	(79.0-79.9)

Table 1 - Model performance (source: [48])

A study[75] in 2019 was focused on participants who used the **smartwatch device** which is produced by Apple. The purpose was to show that a non-on site study can provide a foundation for large-scale studies in which outcomes can be reliably assessed with user-owned devices. The participants consented to monitoring by using their smartphone (Apple iPhone) and then if the smartwatch-based irregular pulse notification algorithm identified possible **atrial fibrillation**, the researchers would send an electrocardiography (ECG) patch to be worn for up to 7 days. The number of

participants mounted up to 419,297 people over 8 months. As an outcome, over a median of 117 days of monitoring, of the 419,297 participants enrolled, only 0.52% received notifications of irregular pulse, and among those with an initial notification who returned an ECG patch, 84% of their subsequent notifications were confirmed to be atrial fibrillation. Overall, 34% among patients who received notifications were diagnosed with atrial fibrillation.

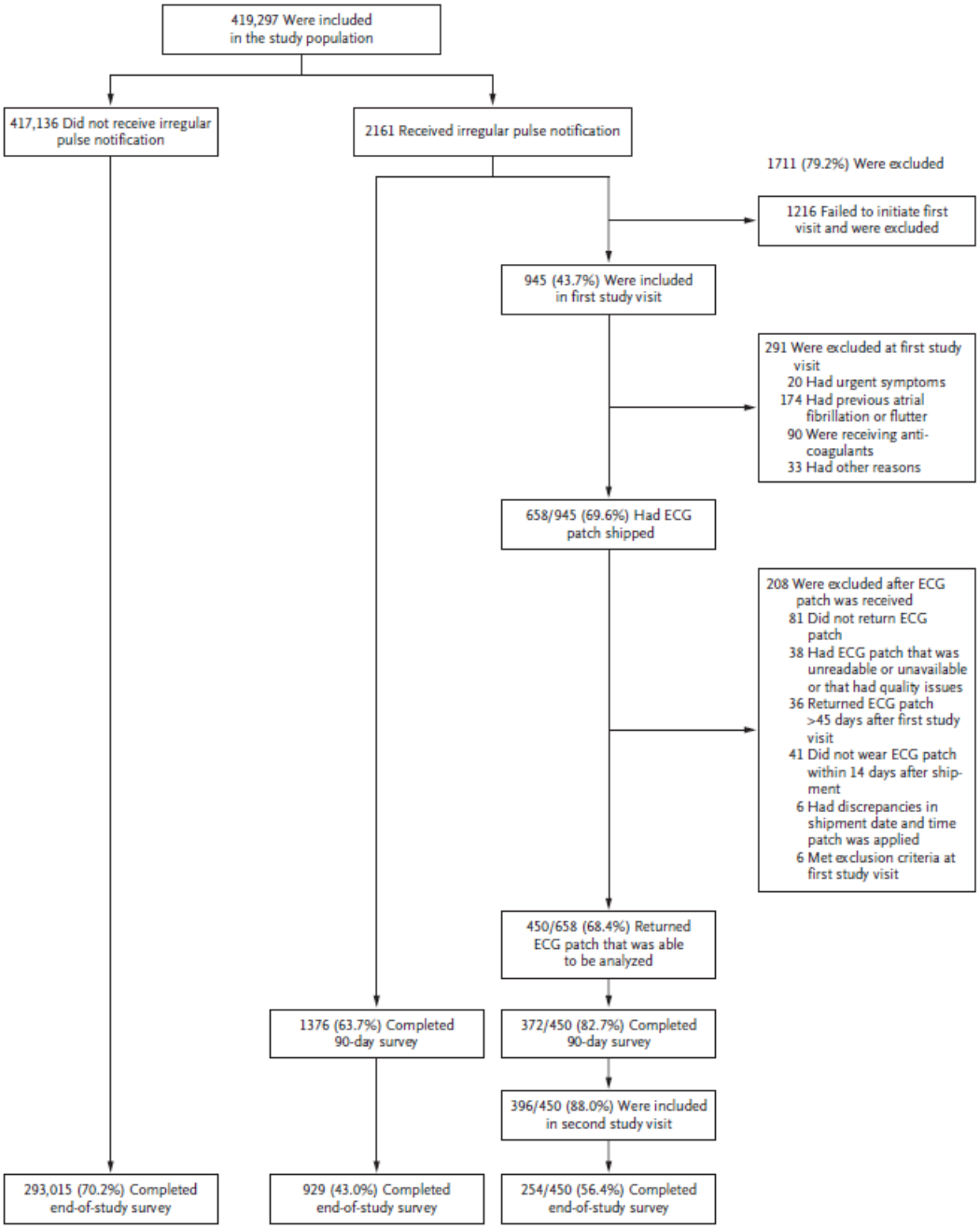


Figure 32 - Participant Selection (source: [75])

In this study[76], researchers developed a deep learning algorithm to identify **atrial fibrillation during normal sinus rhythm** (NSR) using 2,412 12-lead ECG findings (from Inha University Hospital, South Korea). The deep learning algorithm was structured on a recurrent neural network (RNN) using as training, validation, and test datasets normal sinus rhythm (NSR) ECGs in paroxysmal atrial fibrillation (PAF), as well as ECGs from healthy individuals. The developed AI model developed to estimate the probability of PAF during NSR was characterized as 'excellent' for identifying PAF (recall of 82%, specificity of 78%, F1 score of 75% and overall accuracy of 72.8%), while showing that the optimal interval to detect subtle changes of PAF was within 0.24 seconds before the QRS complex in a 12-lead ECG.

A 2021 study's[77] main purpose was to determine whether AI-based deep neural networks are better than the QTc alone in detecting patients with **concealed LQTS** from those with a normal QTc using a 10 seconds 12-lead ECG. The dataset consisted of 967 patients who had a definitive clinical and/or genetic diagnosis of type 1, 2 or 3 LQTS and 1092 patients who were seen because of an initial suspicion for LQTS but were discharged without this diagnosis. The patients were classified, based on a multilayer convolutional neural network while the dataset was parted in 60% training, 10% validating and 30% testing set. Generally, once the patient is fully evaluated and treated, the risk of cardiac events is low. Nevertheless, a small percentage of patients might experience potentially lethal, LQTS-triggered cardiac events[78]. The developed Deep Neural Network was able to distinguish the ECG of a patient with LQTS from the ECG of a patient who was evaluated for LQTS but discharged without this diagnosis with AUC equal to 0.900. It was also able to distinguish the three (3) main genotypic subgroups (LQT1, LQT2, and LQT3) with an AUC of 0.921 (95% Confidence interval, 0.890-0.951) for LQT1 compared with LQT2 and 3, 0.944 (95% CI, 0.918-0.970) for LQT2 compared with LQT1 and 3, and 0.863 (95% CI, 0.792-0.934) for LQT3 compared with LQT1 and 2.

During a study[79] published also in 2021, the researchers developed an algorithm using ECG tracings acquired from a **mobile ECG** (mECG) prototype device equivalent to the commercially-available AliveCor KardiaMobile 6L, to determine the **QTc**. Heart rate-corrected QT interval (QTc) prolongation can predispose to ventricular arrhythmias and sudden cardiac death[80]. QTc can be triggered by drugs, genetics including congenital long QT and/or systemic diseases including SARS-CoV-2 (COVID-19). The utilized

data were 12-lead ECGs (>1.6 million) from 538,200 patients, which were partitioned in 250,767 patients for training, 107,920 patients for testing and 179,513 patients for validation. The data were applied on a Deep Neural Network loosely based on concepts from ResNet[81]. Overall, researchers concluded that QTc estimation from an AI-enabled mECG device may provide a cost-effective method of screening for congenital long QT syndrome in a variety of clinical settings where standard 12-lead electrocardiography is not accessible or cost-effective.

The application of deep learning techniques on the ECGs for arrhythmia detection was mentioned early in 2018[82], in which the researchers contemplated on if there is a **need for the raw samples**. The researchers describe the procedure of applying raw tracings on a Deep Neural Network to automate the interpretation of the ambulatory ECG. The dataset used, was originated from the 2017 PhysioNet AF Challenge (PAFC) as the training data and results from several approaches were publicly available. As a model they also used the at that time state-of-the-art arrhythmia detection 34-layer convolutional neural network (CNN) with residual connections between layers, developed by researchers at Stanford University[17], a novelty in the field which will be mentioned later on.

The main hypothesis of a study[90] in 2019, was to test a deep neural network on predicting an important future clinical event, **one-year all-cause mortality**, from ECG traces. The performance of the algorithm on that event showed an average AUC of 0.85 from a model cross-validated on 1,775,927 12-lead resting ECGs, collected from ~400,000 patients over a period of 34 years. To further evaluate the results of the algorithm, the researchers conducted a blinded survey given to three cardiologists, who after examining 240 paired examples of labeled true positives (dead) and true negatives

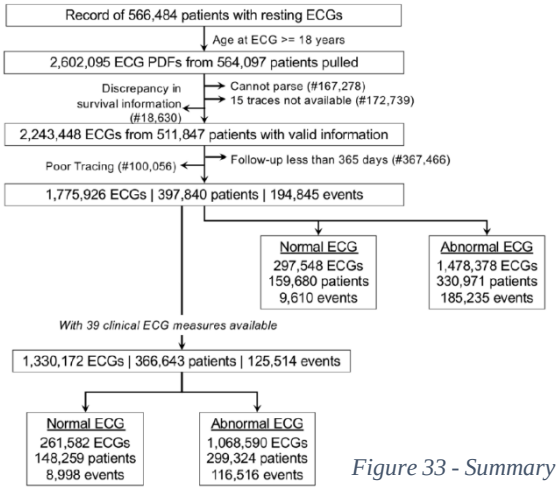


Figure 33 - Summary of data used in the study (source: [90])

(alive) suggested that the patterns captured by the model were generally not visually apparent. Overall, the study concluded that deep learning can add prognostic value to the interpretation of 12-lead resting ECGs, even in cases that are labeled as 'normal' by physicians.

The main purpose of this study[83] was to compare an already existing process called **CHARGE-AF** (Cohorts for Heart and Aging Research in Genomic Epidemiology model for atrial fibrillation) with another method developed by the researchers using AI-ECG. The sample which was used originated from the population-based Mayo Clinic Study of Aging with patients who had no history of AF at the time of the study visit.

To further the knowledge on **AF**, scientists examined the use of AI-enabled ECG, which already has the ability to identify patients with unrecognized AF, for determining if the model differentiates between patients with embolic strokes of unknown source (ESUS) and those with known causes of stroke.[84] ESUS are common and often caused by unrecognized paroxysmal atrial fibrillation. The study concludes with evidence that AI-ECG may become a useful tool in monitoring patients with ESUS to identify those who might benefit from anticoagulation.

4 Application and comparison of algorithms on case study

4.1 ECG Description

An electrocardiogram is the medium for recording the electrical activity of the heart. In order to perform an ECG session, a set of electrodes are placed on human chest and limps, from different angles. Electrical signals are collected and processed to identify the way the heart works.

An electrocardiogram consists of the following parts:

- P waves: This kind of wave corresponds to atrial depolarization. ECG derived from healthy humans contains a P wave before each QRS complex.
- PR interval: It is the time segment from the beginning of the P wave to the beginning of a Q wave. It corresponds to the time interval required for a full electrical activity between the atria and the ventricles.
- QRS complex: The QRS complex corresponds to the ventricle's depolarization. On the ECG it is represented as three related waves on the ECG (the Q, R and S wave).
- ST segment: The ST segment is the time interval between the end the S wave and the beginning of the T wave. It consists of an isoelectric line representing the time needed to pass from depolarization to repolarization of the ventricles.
- T wave: Each T wave corresponds to ventricular repolarization. It is formed as a small wave after the QRS complex.
- RR interval: It is the time interval between the peak of two consequent R waves. It also corresponds to time elapsed between two QRS complexes.
- QT interval: It is the time interval between the beginning of the QRS complex and the end of the T wave. It corresponds to the time needed for the ventricles to depolarize and then repolarize.

The following picture shows schematically the different parts of an ECG and their relationship.

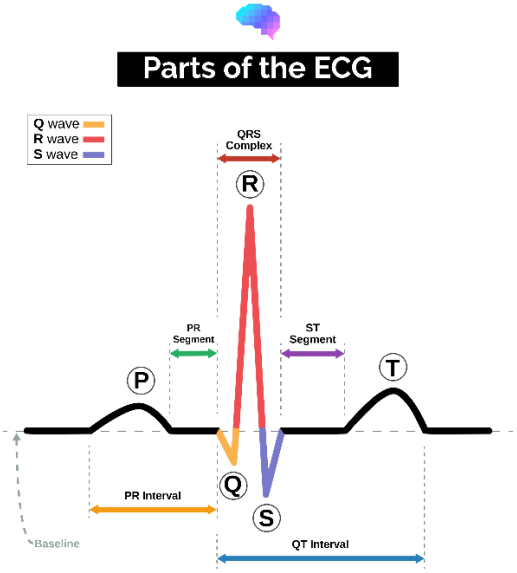


Figure 34: ECG's parts

The ECG is recorded on a special form with the following characteristics:

- It is divided into small and large squares
- Each small square represents 0.04 seconds
- Each large square represents 0.2 seconds (it contains 25 small squares)

The following picture shows an ECG sample paper.

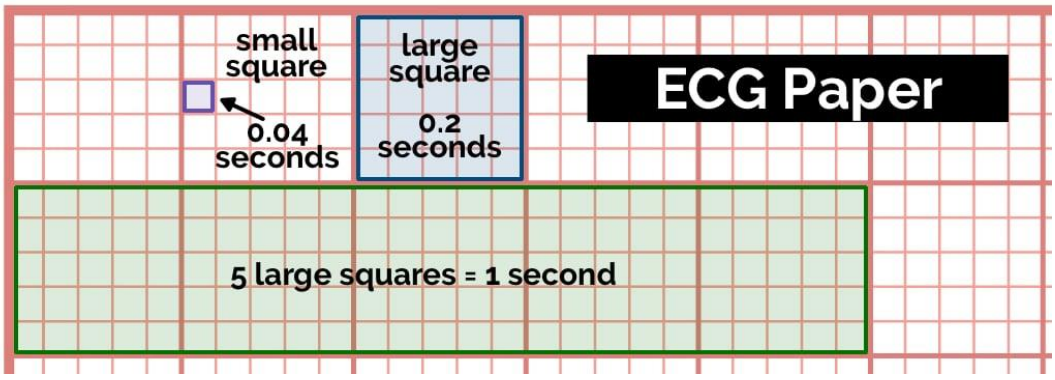


Figure 35: ECG paper

Each electrode used to produce the ECG is attached to the person's skin and records electrical activity. The collected data is used to calculate 12-lead ECG. By the term ECG lead, a graphical representation of the heart's electrical activity is described. Data needed for this representation is derived by ECG electrodes. Each ECG lead creates a graph on a piece of ECG paper. 10 physical electrodes are used for producing 12 leads.

The electrodes used and the area they collect data from are listed in the table below.

No	Electrode	Position	Lead
1	V1	4th intercostal space at the right sternal edge	Lead V1
2	V2	4th intercostal space at the left sternal edge	Lead V2
3	V3	Midway between the V2 and V4 electrodes	Lead V3
4	V4	5th intercostal space in the midclavicular line	Lead V4
5	V5	Left anterior axillary line at the same horizontal level as V4	Lead V5
6	V6	Left mid-axillary line at the same horizontal level as V4 and V5	Lead V6
7	Red (RA)	Ulnar styloid process of the right arm	Lead I, Lead II, aVR, AVL, aVF
8	Yellow (LA)	Ulnar styloid process of the left arm	Lead I, Lead III, aVR, AVL, aVF
9	Green (LL)	Medial or lateral malleolus of the left leg	Lead II, Lead III, aVR, AVL, aVF
10	Black (RL)	Medial or lateral malleolus of the right leg	

Leads description is listed below.

Lead	View of the heart
V1	Septal view of the heart
V2	Septal view of the heart
V3	Anterior view of the heart
V4	Anterior view of the heart
V5	Lateral view of the heart

Lead	View of the heart
V6	Lateral view of the heart
Lead I	Lateral view (calculated by analysing activity between the RA and LA electrodes)
Lead II	Inferior view (calculated by analysing activity between the RA and LL electrodes)
Lead III	Inferior view (calculated by analysing activity between the LA and LL electrodes)
aVR	Lateral view (calculated by analysing activity between LA+LL -> RA)
AVL	Lateral view (calculated by analysing activity between RA+LL -> LA)
aVF	Inferior view (calculated by analysing activity between RA+LA -> LL)

Each lead is of different shape because data used are derived from a different heart view. Electrical activity from heart to electrode is recorded as a positive deflection, while electrical activity from electrodes is recorded as negative reflection. As in fact electrical activity flows to many directions, each deflection represents the average orientation of it. The lead that has the maximum positive deflection value is supposed to be closer to the direction of hearts electrical activity. A 12-lead ECG consists of 3 limb leads (leads I, II, and III), 3 augmented limb leads (leads aVR, aVL, and aVF), and 6 atrial leads (V1 to V6). The data obtained from each lead corresponds to the electrical activity of the heart from views at different angles[108].

The following schema shows the angle of heart view that each lead corresponds.

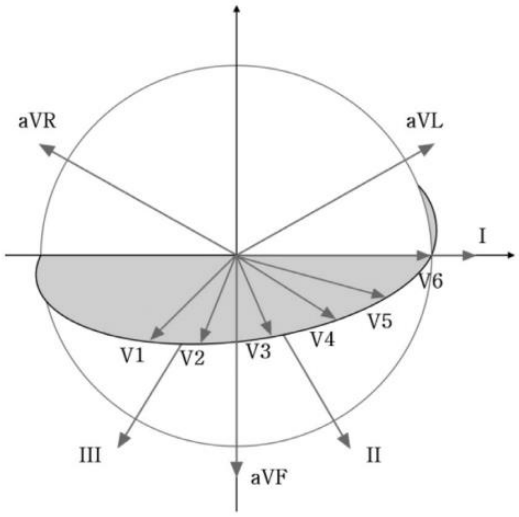


Figure 36: Leads and heart view angle correspondence.

Depending on the type of electrical activity, located in specific areas of the heart, pathologies can be determined in specific parts of it[110].

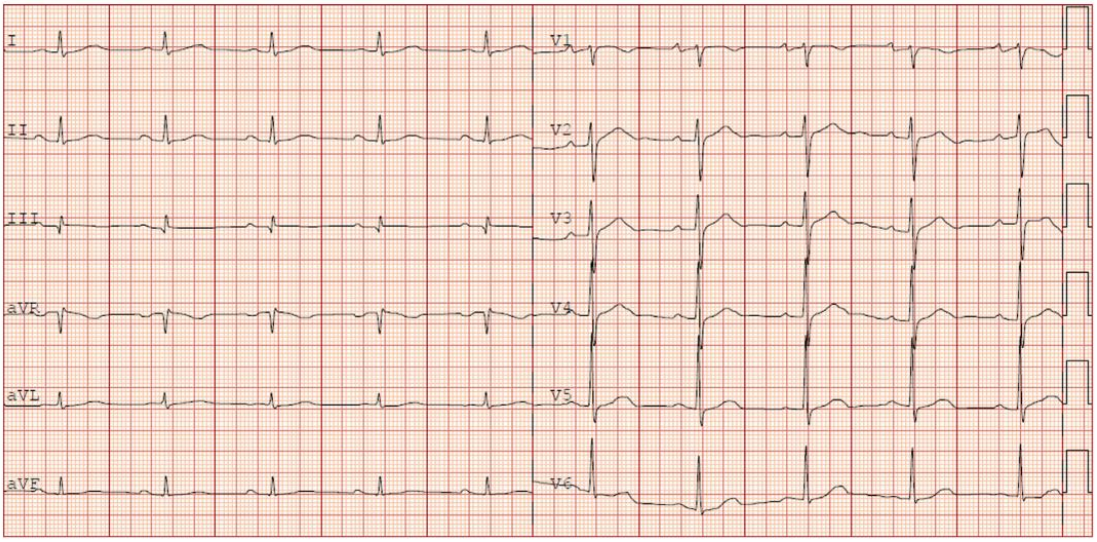


Figure 37: A 12 lead ECG example

An ECG corresponding to a healthy person shows P-waves, QRS-complexes, and T-waves continuously and repeatedly in a sequence. The pattern of this sequence is

referred to as sinus rhythm*. Abnormal heart functions are identified with the help of ECGs and are referred to as arrhythmia. A key role in the study to identify these cases is played by the rhythm and form of the P-wave, the rhythm of the QRS-complex and the correlation between the P-wave and the QRS-complex. The images below show a comparison of an ECG of a healthy person and a person with a form of arrhythmia. In the second case, the P-wave and T-wave are not clearly distinguishable, and the rhythm of the QRS-complex is not stable[112].

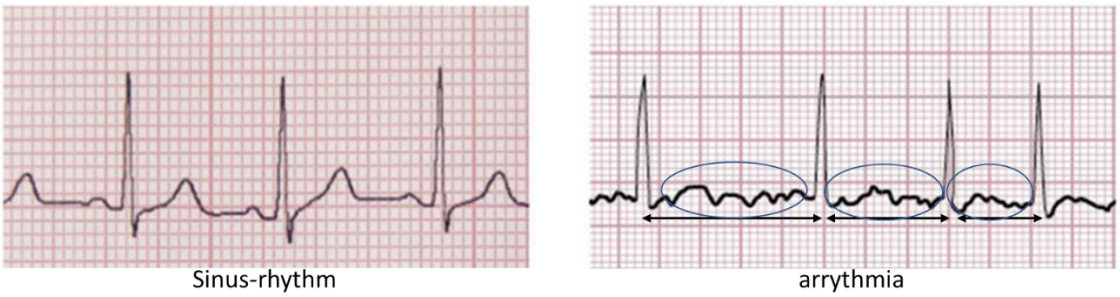


Figure 38: Sinus-rhythm and arrythmia comparison

For the automated detection of arrhythmia in an ECG it is necessary to recognize the distinct types of waves and to decode the relationships between them. The difficulty of this process has to do with the variability of the wave morphology between samples.

4.2 ECG Data

The data used for the present study comes from the MIT-BIH Arrythmia. It is a collection of 48 fatigue test sessions, lasting 30 minutes. ECGs were performed on 47 subjects and examined between 1975 and 1979. 23 of these sessions were randomly selected from a total of 4000 one-hour recordings, from a mixed population of inpatients (approximately 60%) and outpatients (approximately 40%) at Beth Israel Hospital, Boston. The remaining 25 records were selected from the same set to correspond to arrhythmias to ensure that rare forms of arrhythmia would also be included in the data set under consideration. Recordings were digitized at 360 samples per second per channel with 11-bit resolution at 10 mV range. More than one cardiologist was used to characterize each sample (if it was normal cardiac function or

* It is defined as a normal rhythm that includes a P-wave with a stable form, followed by a QRS complex and a positive vector.

to determine pathological function). The people to whom the ECGs are matched are 25 men aged 32 to 89 and 22 women aged 23 to 89. In most records, the upper signal is a modified limb lead II (derived by electrodes placed on the chest), while the lower one is a modified lead V1 or V2 or V5 or V4 (derived from electrodes also placed on the chest). Normal QRS complexes are presented in the upper signal and the lead axis for the lower signal may be nearly orthogonal to the mean cardiac electrical axis. It is difficult to discern normal beat in the lower signal[109].

4.3 ECG Annotations

Initially all detected events annotated as normal beats. Two different cardiologists added additional beat labels where they were missing, or false and changed the labels for all abnormal beats. They also added rhythm labels, signal quality labels, and comments. All disagreements were resolved by consensus. The annotations were analysed by an auditing program, which checked them for consistency, and which located the ten longest and shortest R-R intervals in each record, to identify possible missing or falsely detected beats[111]. Annotations are mainly located accurately at the R-wave peak. Thus, dataset is suitable for scientific studies. The database contains approximately 109,000 beat labels.

The following table shows the symbols used to annotate the signals[103].

Symbol	Meaning	Description
· or N	Normal beat	Normal Beat—N
L	Left bundle branch block beat	Normal Beat—N
R	Right bundle branch block beat	Normal Beat—N
A	Atrial premature beat	Supraventricular—SVEB
a	Aberrated atrial premature beat	Supraventricular—SVEB
J	Nodal (junctional) premature beat	Supraventricular—SVEB
S	Supraventricular premature beat	Supraventricular—SVEB

V	Premature ventricular contraction	Ventricular—VEB
F	Fusion of ventricular and normal beat	Fusion Beat—F
[	Start of ventricular flutter/fibrillation	Fibrillation
!	Ventricular flutter wave	Fibrillation
]	End of ventricular flutter/fibrillation	Fibrillation
e	Atrial escape beat	Normal Beat—N
j	Nodal (junctional) escape beat	Normal Beat—N
E	Ventricular escape beat	Ventricular—VEB
/	Paced beat	Unknown—Q
f	Fusion of paced and normal beat	Unknown—Q
x	Non-conducted P-wave (blocked APB)	Unknown—Q
Q	Unclassifiable beat	Unknown—Q
|	Isolated QRS-like artifact	Unknown—Q

4.4 Methodology

For this case study, the main purpose is the comparison of the performance of different algorithms on ECG data from Physionet challenge 2017. Specifically, in this case study, machine learning algorithm XGBoost and deep learning algorithm Neural Network with different types of structures will be compared based on accuracy with the same training dataset.

4.4.1 Getting data

Although 12-lead ECGs are the most commonly used for the diagnosis of heart-related diseases, single-lead ECGs are also used in various methodologies today, which show high yields. These methodologies usually involve deep learning techniques. Single lead ECGs from the relevant MIT database were used in the present study. Dataset is available for downloading from Physionet archives[†]. Four files correspond to each patient. The main one is the MIT Signal files (.dat) are binary files containing samples of digitized signals (waveforms). MIT Header files (.hea) are needed to describe signal files. There are also MIT Annotation files which are binary files containing annotations (.atr). Files having .xws extension contains records position url[109].

[†] https://www.physionet.org/content/mitdb/1.0.0/

4.4.2 Data Analysis

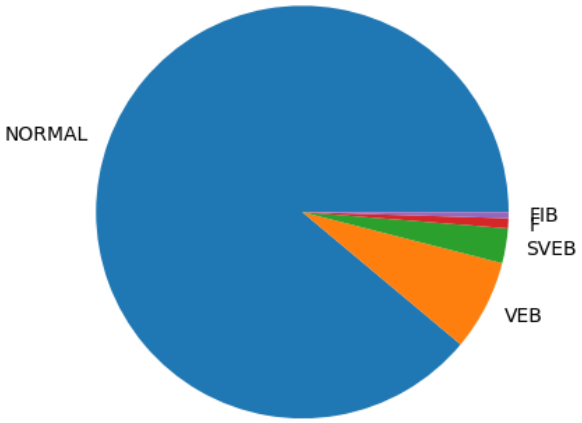
The data was read, and the characterizations assigned to the pulses recorded by the cardiologists were examined. The pulse peaks and their characterizations were examined. Each of these peaks was examined in time windows of 2 and 6 seconds (in the first case it was examined 1 sec before and 1 sec after the peak and in the second case 3 and 3 sec respectively).

The table below shows the number of pulses by type.

Symbol	frequency	Description
N	75052	Normal Beat—N
L	8075	Normal Beat—N
R	7259	Normal Beat—N
V	7130	Ventricular—VEB
/	7028	Unknown—Q
A	2546	Supraventricular—SVEB
+	1291	
f	982	Unknown—Q
F	803	Fusion Beat—F
~	616	
!	472	Fibrillation
"	437	
j	229	Normal Beat—N
x	193	Unknown—Q
a	150	Supraventricular—SVEB
|	132	Unknown—Q
E	106	Ventricular—VEB
J	83	Supraventricular—SVEB
Q	33	Unknown—Q
e	16	Normal Beat—N
[	6	Fibrillation
]	6	Fibrillation
S	2	Supraventricular—SVEB

Those records that could not be characterized were removed from the data set. The pulse occurrence frequencies corresponding to each of the arrhythmia types are shown in the table below.

DESCRIPTION	OCCURENCES
NORMAL	90730
VEB	7240
SVEB	2781
F	803
FIB	484



From the statistical analysis of the samples, it emerged that the number of those corresponding to normal cardiac function is much greater than the number of samples corresponding to arrhythmias. For this reason, virtual samples corresponding to arrhythmia were created. After arranging the number and type of samples a set of models were developed using dense neural networks. Convolutional neural networks were then developed. The models of the two classes were compared based on performance metrics.

4.4.3 Preprocess

The data had to go through a pre-processing stage in order to get an editable form. In order to read the files containing the ECG data, it was necessary to install the wfdb library. By using the library functions, binary files containing the descriptions and annotations of the ECGs were read. Descriptions corresponding to normal heart operation as well as heart operation associated with arrhythmias had to be retained. For this purpose, records whose annotations did not correspond to normal cardiac function or arrhythmia were removed.

4.4.4 Process

Although various methods have been proposed for ECGs processing in order to detect arrhythmia states (logistic regression, support vector machines, random forests, k-nearest neighbours, decision trees), the performance of deep learning techniques seems to be at much higher levels in relation to the rest. In the present study, the performances of different neural networks with different parameters were examined. Dense Neural Networks and Convolutional Neural Networks were examined. The aim of experiments carried out is defining the possible arrythmia type by the ECG format 1 or 3 seconds before and after beat peaks.

The following table shows the data of the experiments carried out.

No	Neural Network Type	Details
1	Dense	Window Size: 6 seconds Classes: 5 Layers: 3
2	Dense	Window Size: 6 seconds Classes: 5 Layers: 2 (initial size 1/36 of input size)
3	Dense	Window Size: 6 seconds Classes: 5 Layers: 3 (initial size 36)
4	Dense	Window Size: 2 seconds Classes: 5 Layers: 3
5	Dense	Window Size: 2 seconds

No	Neural Network Type	Details
		Classes: 5 Layers: 2 (initial size 1/36 of input size)
6	Dense	Window Size: 2 seconds Classes: 5 Layers: 3 (initial size 36)
7	Dense	Window Size: 6 seconds Classes: 2 Layers: 3
8	Dense	Window Size: 6 seconds Classes: 2 Layers: 2 (initial size 1/36 of input size)
9	Dense	Window Size: 6 seconds Classes: 2 Layers: 3 (initial size36)
10	Dense	Window Size: 2 seconds Classes: 2 Layers: 3
11	Dense	Window Size: 2 seconds Classes: 2 Layers: 2 (initial size 1/36 of input size)
12	Dense	Window Size: 2 seconds Classes: 2 Layers: 3 (initial size 36)
13	Convolutional	Window Size: 6 seconds Classes: 5 Layers: 2
14	Convolutional	Window Size: 6 seconds Classes: 5 Layers: 4
15	Convolutional	Window Size: 2 seconds Classes: 5 Layers: 2

No	Neural Network Type	Details
16	Convolutional	Window Size: 2 seconds Classes: 5 Layers: 4
17	XGBoost	Window Size: 3 seconds Classes: 5
18	XGBoost	Window Size: 3 seconds Classes: 2
19	XGBoost	Window Size: 1 seconds Classes: 5
20	XGBoost	Window Size: 1 seconds Classes: 2

4.5 Convolutional Neural Networks

Patrick Haffner and researchers from AT&T introduced the first Convolutional Neural Network (CNN) in 1996. It was the key component in a large-scale application for testing image recognition. This technology was positively evaluated 10 years later, when it was renamed deep learning and machine learning researchers began to focus on another technique developed by the same group: Support Vector Machines (SVM).

Convolutional Neural Networks (ConvNet) consist of neurons with inputs that include weights. These weights are related to the learning process and biases. Each neuron receives a set of inputs and generates a dot product. The entire network expresses a single differentiable score function. At the input of this network, the data enters in the form of a vector, and at its output, the estimate of the category to which you belong is obtained. Often convolutional neural networks are used in image categorization tasks, however they are also suitable for inputs that are in the form of large vectors. They also include a loss function at the last - fully connected - level. This allows certain properties to be encoded into their architecture, which make the forwarding function more efficient in implementation and significantly reduce the number of parameters in the network[101].

Neural networks take an input (a single vector) and transform it through a series of hidden layers. Each hidden layer consists of a set of neurons, where each neuron is fully connected to all neurons in the previous layer. Neurons in a single layer operate completely independently and do not share any connections. The last fully connected layer is called the output layer and in the classification settings it represents the scores of the available classes. The input data is described by vectors of large size[100].

Convolutional Neural Networks, unlike regular neural networks, are structured having on their layer's neurons arranged in 3 dimensions: width, height, and depth. Neurons in each conv layer are connected to only a small region of the layer before it, rather than all neurons in a fully connected manner. Additionally, the final output layer has dimensions that can be adjusted to the number of categories. Links between neurons of neighbouring layers can be represented by sparse graphs. The following image shows the difference between neurons linkage on dense and convolutional neural networks[106].

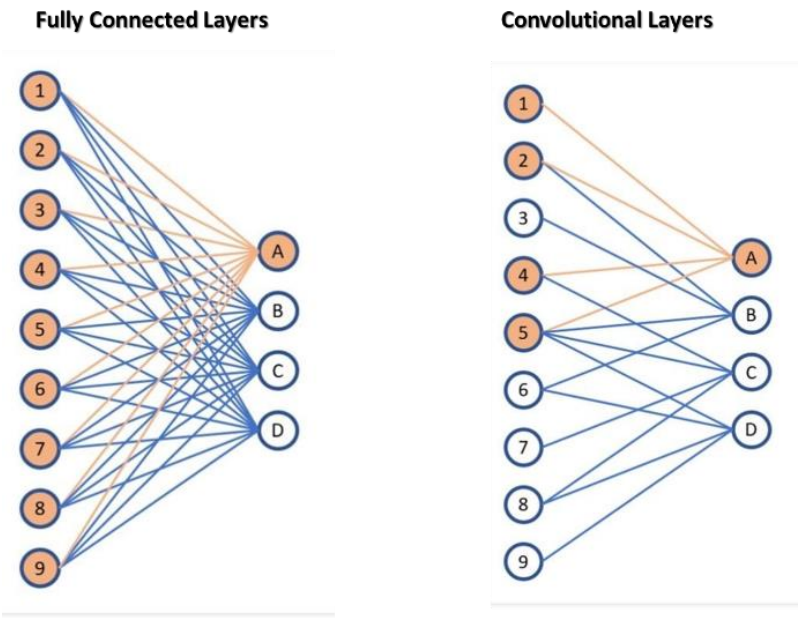


Figure 39: Comparison between Fully Connected and Convolutional Layers

ConvNets consist of a sequence of layers. Each layer transforms one set of activations into another through a differentiable function. Three types of layers are used to build ConvNet architectures:

- Convolutional Layer
- Pooling Layer
- Fully Connected Layer

The composition of such a neural network usually includes the following levels:

- Input level: In this level, the original raw vectors will be entered.
- Convolutional Layer (Conv): In this layer the computations performed for the output of the neurons connected to local regions at the input. Each computation calculates a dot factor between their weights and a small region they are connected to input vectors. This leads to a number of computations involved in the input vectors and the filters selected.
- Relu level: In this level an activation function of the form max(0,x) is applied which rejects negative values.
- Pooling Level (Pooling): At this level a subsampling operation is performed along the dimensions of the generated vectors resulting in the limitation of the output limitations.

- Fully Connected Level: This level calculates the compatibility evaluation of the generated vector with the classes. A characteristic of this level is the fact that each neuron is connected to all the neurons of the previous level.

The image below shows their architecture, schematically:

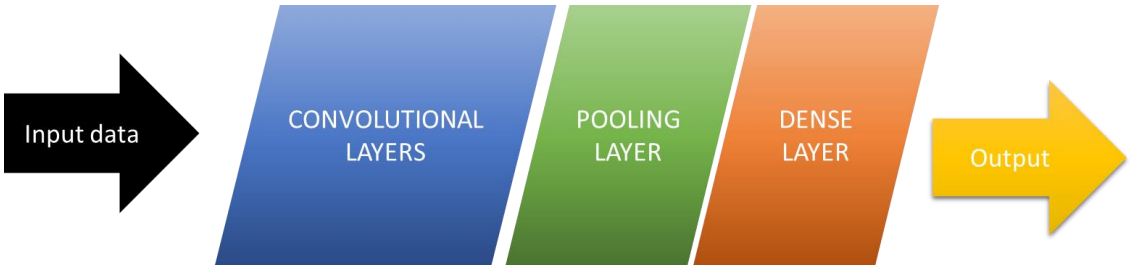


Figure 40: Typical Convolutional Neural Network Architecture

In this way, Convolutional NN transform the level corresponding to the original vector into the final class scores. Some levels contain parameters and some do not. Specifically, the Convolutional – Fully Connected Layers perform transformations that are depended to both the activations in the input and the parameters used (neuron weights and biases). On the other hand, the Relu – Pooling layers perform fixed functions. Parameters at Convolutional – Fully Connected Layers are being trained by gradient descent methodology[‡] so that the class scores computed by the Convolutional Network are consistent with the labels in the training set for each vector.

The differentiation of convolutional networks are:

- In the functions used
- In the parameters of the structure of each level

4.5.1 Convolutional Layers

The Conv layer is the basic building block of a Convolutional Network, as it performs most of the neurons' function weights calculations. The parameters of the this layer consist of a set of learnable filters. Each filter spans the entire depth. During the forward pass, each filter is dragged by entire input and the dot factors between the filter

[‡] This is the convergence algorithm that runs by decreasing the derivative. The downscaling algorithm is initialized with values θ_0 and θ_1. Changing these values causes the convergence to values that minimize the cost function $J(\theta_0,\theta_1)$

inputs and the input at any position are calculated. As the filter is dragged across the input vector, an activation map is created that gives the responses of that filter at each location. Intuitively, the network will learn filters that are activated when they detect some feature, such as entire patterns in higher layers of the network. In this way, a set of filters is created at each convolutional layer and each of them will produce a separate activation map. These are stacked along the depth dimension to produce the output. Each entry of the output can also be interpreted as the result of operations on a neuron corresponding to a limited region of the input as well as its correlation with its neighbouring neurons (corresponding to neighbouring regions).

As the inputs of the initial layer are high dimensional, it is preferred to connect each neuron to a limited area. The extent of this region is a hyperparameter of the neural network and is of the same size as the filter applied. The extent of connectivity along the depth axis is always equal to the depth of the input. If the input size is 32x32x1 and filter size id 5X5, then each neuron in the Pooling layer would weight in an area of 5X5X1.

The size of the output is determined by three hyperparameters, as described below:

- The depth of the output: It is a hyperparameter corresponding to the number of filters used. Sometimes each filter may be used for different purposes (e.g. recognising different objects in an image).
- The step at which the filter is dragged: This value determines how far the filter is moved each time. The smaller this value, the larger outputs are produced.
- Padding: This is the application of a technique in which the edges of the image are filled with zeros. The size of this zero padding is a hyperparameter. The advantage of using zero padding is that it allows control over the spatial size of the output volumes. Used to maintain the spatial size of input size by height and width. The size of the output is calculated as a function of the size of input size W, the size of the receptive field of the neurons in the Pooling layer F, the step with which S is applied and the size of the zero padding P. Under these conditions, the number of neurons is done based on the formula (W+F+2P)/(S+1)

4.5.2 Pooling Layer

Pooling Layers are placed between successive Convolutional layers. They are used for progressively reduce the size of the spatial representation and the elimination of parameters used for the calculations performed in the neural network.

They operate independently on each depth section of input size and resize them spatially, using mainly the max or the average function (the max or the average value of a number of values are being replaced). In general, the Pooling level accepts input of size WXYXD, requiring hyperparameters: the spatial expansion (F) and the shift (S). The result is the output of size W'XH'XD, where:

W'=(W-F)/(S+1), H'=(H-F)/(S+1)

4.5.3 Fully Connected Layer

A Convolutional architecture includes near the output one or more Fully Connected Layers. Neurons in a fully connected layer have full connections to all activations in the previous layer. They are mainly used to adapt the output results of Convolutional Layers to the requirements of the class scores.

4.5.4 General Operation

ConvNet architectures make the explicit assumption that for inputs, their neighbouring features are strongly correlated with each other. This fact makes the forwarding function more efficient, as the number of parameters in the network is considerably reduced. Deep convolutional neural networks have greatly improved the classification processes of unstructured and large-volume data, such as images, objects within images, or texts. Because of this complexity, conventional machine learning approaches are generally slow to process this kind of data. The complexity arises as their encoding usually produces large-volume vectors. Processing them requires a lot of time and processing power. Convolutional Neural Networks have the ability to produce highly accurate results by performing a relatively small number of calculations, exploiting the correlation of neighbouring regions of the input.

4.6 XGBoost Algorithm

The XGBoost algorithm is based on the operation of decision trees. Its characteristic is that it produces reliable prediction models in a short time. It is actually an implementation of gradient-boosting decision trees. Gradient boosting algorithm aims to minimizing the loss function (mean squared error, cross-entropy etc) of the previous model using gradient descent. At every step computes the gradient of the loss function in combination with the predictions of the current sample. After that it trains a weak model, trying to minimize this gradient. The predictions of the derived model are added to the sample. This process is being repeated since a criterion to be met[107]. Gradient boosted trees are a kind of gradient boosting algorithm where the learning process is based on classification, and regression trees. Each sample contains a set of trees. At each algorithm step (let be the i-th), a model (tree) is being trained using a set of features X and a set of labels y_i. The model is being used to predict the labels of X (y_{i+1}). The distance between y_i and y_{i+1} is the loss value. At the next step X and y_{i+1} is being used for constructing the new model (tree). The process is being repeated while loss value is below a threshold. The next figure shows these gradient boosted trees algorithm schematically[104].

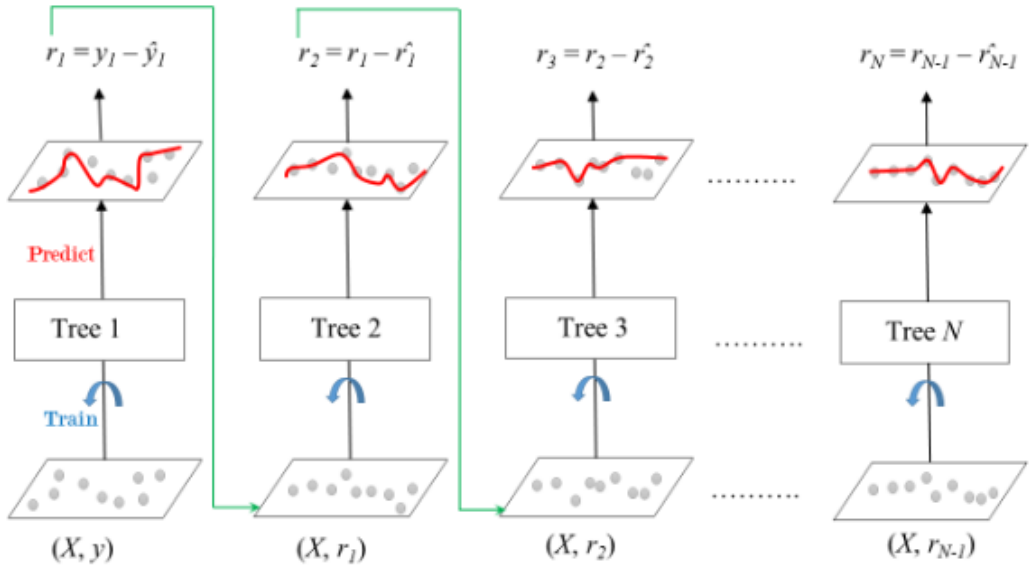


Figure 41:Gradient boosted trees algorithm

Its main advantage is that it can provide reliable results without the need to optimize its parameters. This fact makes it easy to use and quick to deliver results. The algorithm's main elements are:

- It handles overfitting problem by introducing L1/L2 penalties on the weights and biases of developed decision trees.
- It uses weighted quantile sketch algorithm for handling sparse data sets.
- It can perform parallel execution as it is block structured.
- It uses cache awareness for eliminating memory requirements (especially for cases of training large datasets).
- It can perform out-of-core computing using disk-based data structures as well as in-memory, during the computation phase[102].

The XGBoost weak learner basic model, might be decision tree, linear model, or DART model. Concerning trees, the most important hyperparameters considered are:

- Learning rate which describes the threshold in loss values for stopping the learning process (in order to prevent overfitting)
- Gamma which is the desired loss difference value for stopping learning process.
- Lasso – L1 and Ridge - L2 regularization term on weights: Increasing L1 value makes the model more conservative and less likely to overfit. Increasing L2 value makes the model more conservative and less likely to overfit. These parameters' objective is to control overfitting by lowering variance and increasing some bias. L1 adds the sum of the absolute beta coefficients while Ridge(L2) adds the sum of the beta coefficients squared. The loss equitation from the above form

$$loss = \sum_{i=0}^{n} (y_{i-\widetilde{y)}}{}^2 = \sum_{i=0}^{n} (y_{i-X_i b)}{}^2$$

Transformed to the next ones:

$$L1 = loss = \sum_{j=0}^{m} |b_i \ | + \sum_{i=0}^{n} (y_{i-X_i b)}{}^2$$

$$L2 = loss = \sum_{j=0}^{m} b_i{}^2 + \sum_{i=0}^{n} (y_{i-X_i b)}{}^2$$

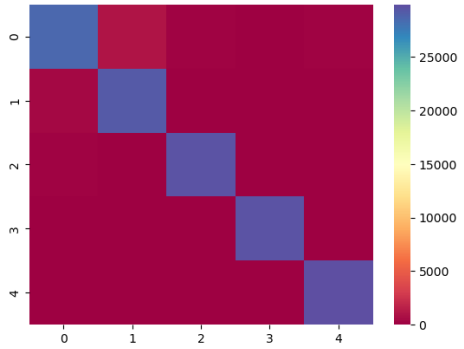
where $\check{y}$ is the predicted value, y the actual value and X the independent variable

- Maximum number of basic models (trees). Increasing this value makes the model more complex and more likely to overfit.

- Objective function used.
- The maximum depth of each tree produced at each step[102].

4.7 Results

Experiment 1

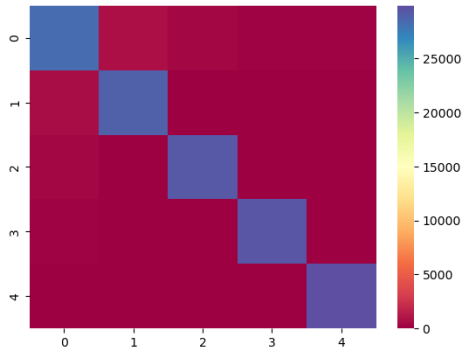


Confusion Matrix

28565	957	120	81	120
380	29271	17	5	
217	79	29561	42	11
64	6	0	29540	0
114	0	1	0	29887

Accuracy: 0.985

Experiment 2

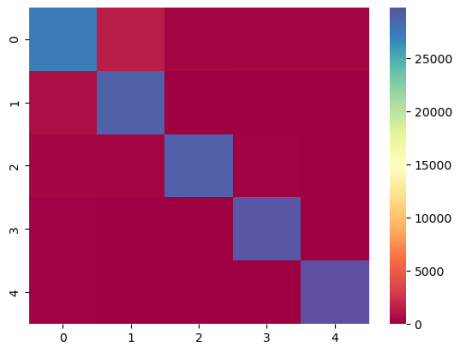


Confusion Matrix

28338	726	448	181	150
677	28902	78	13	5
354	83	29404	59	10
151	0	13	29446	0
106	1	21	0	29874

Accuracy: 0.979

Experiment 3

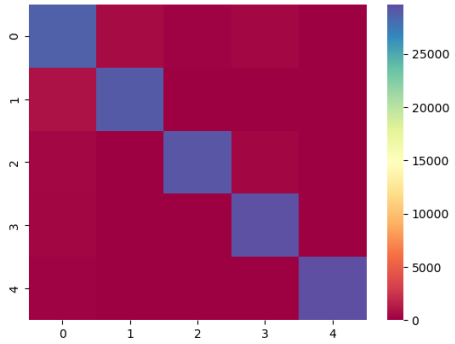


Confusion Matrix

27385	1424	432	335	267
740	28825	50	11	49
436	235	28993	184	62
168	30	34	29378	0
143	4	33	5	29817

Accuracy: 0.968

Experiment 4

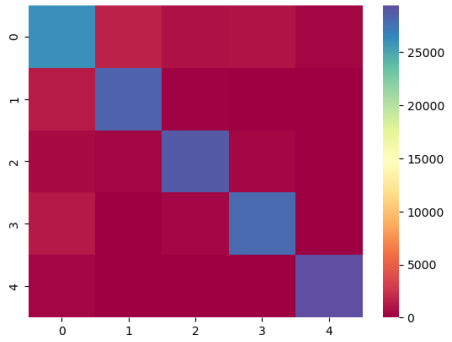


Confusion Matrix

28685	511	159	459	48
818	28993	54	9	0
358	98	29248	333	2
292	0	5	29448	0
192	0	9	0	29649

Accuracy: 0.977

Experiment 5

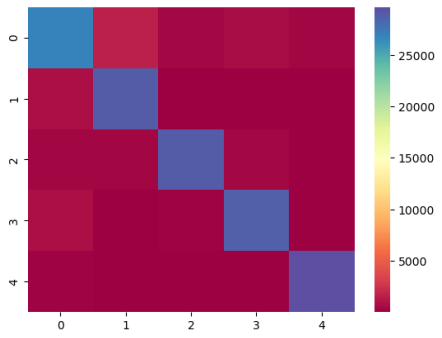


Confusion Matrix

26088	1664	811	972	327
1327	28307	176	42	22
468	279	28859	377	56
1183	107	382	28051	22
324	17	63	0	29446

Accuracy: 0.942

Experiment 6

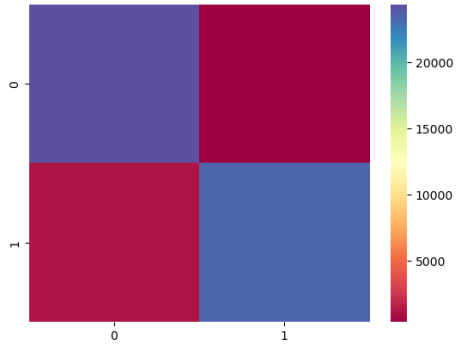


Confusion Matrix

26898	1551	419	693	301
769	28939	107	48	11
321	309	28906	448	55
803	11	145	28777	9
141	3	31	5	29670

Accuracy: 0.958

Experiment 7

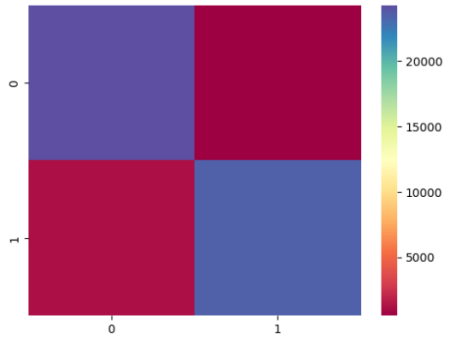


Confusion Matrix

24387	390
1163	23425

Accuracy: 0.968

Experiment 8

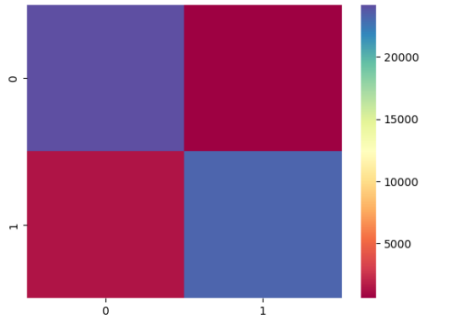


Confusion Matrix

24249	528
1141	23447

Accuracy: 0.966

Experiment 9

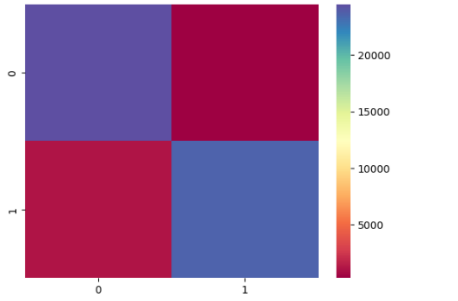


Confusion Matrix

24202	575
1340	23248

Accuracy: 0.961

Experiment 10

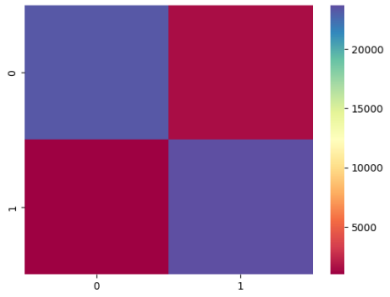


Confusion Matrix

24514	262
1070	23631

Accuracy: 0.973

Experiment 11

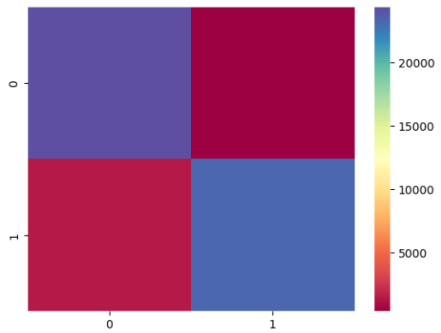


Confusion Matrix

23368	1408
963	23738

Accuracy: 0.952

Experiment 12

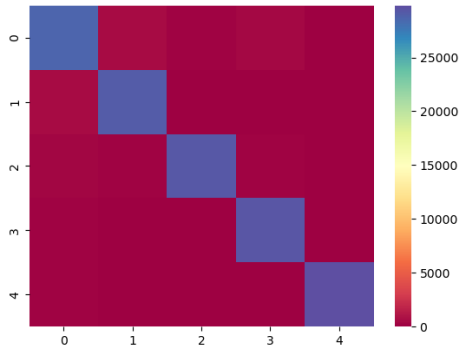


Confusion Matrix

24396	380
1404	23297

Accuracy: 0.963

Experiment 13

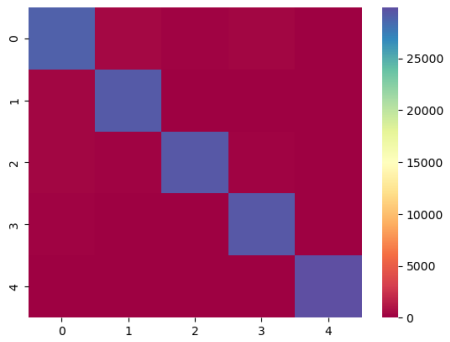


Confusion Matrix

28578	510	227	425	103
534	29032	85	22	2
251	188	29259	198	1
214	0	8	29388	0
140	0	30	0	29832

Accuracy: 0.980

Experiment 14

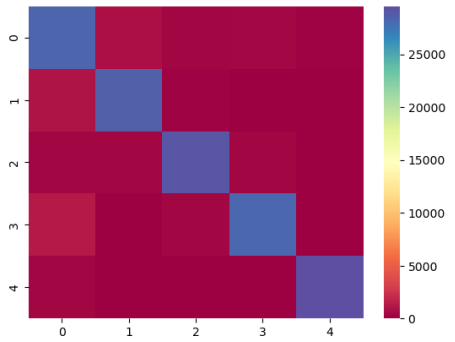


Confusion Matrix

28901	444	148	264	86
326	29264	75	10	0
239	130	29381	145	15
186	0	16	29408	0
66	0	6	0	29930

Accuracy: 0.985

Experiment 15

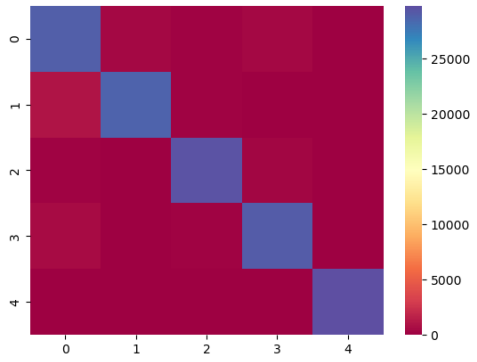


Confusion Matrix

28306	789	278	357	132
1032	28574	225	40	3
281	269	29131	296	62
1240	18	246	28241	0
[273	3	24	0	29550

Accuracy: 0.962

Experiment 16



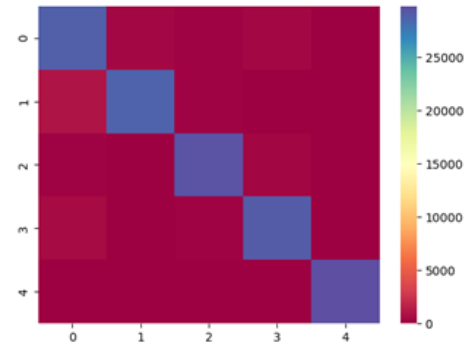
Confusion Matrix

28867	374	139	424	58
985	28659	170	60	0
222	89	29441	275	12
551	0	137	29057	0
73	1	12	0	29764

Accuracy: 0.976

Experiment 17

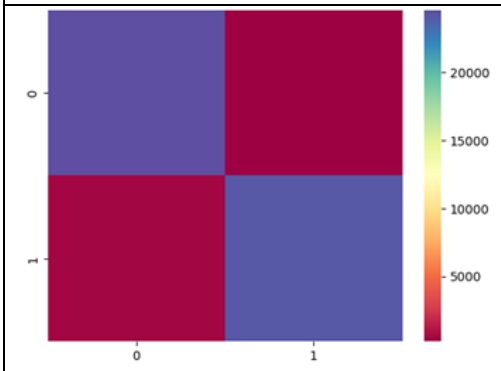
XGBoost

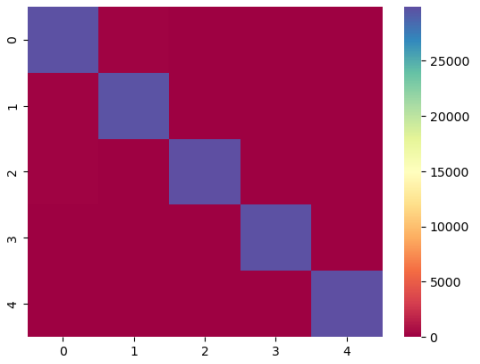


Confusion Matrix

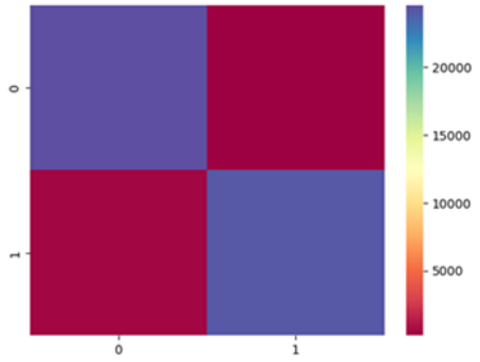
29683	87	38	27	8
150	29525	0	0	0
122	7	29779	2	0
12	0	0	29598	0
42	0	0	0	29960

Accuracy: 0.997

<table>
<tr><td>Experiment 18</td><td>XGBoost</td></tr>
<tr><td> 
</td><td>Confusion Matrix

<table><tr><td>24573</td><td>204</td></tr><tr><td>488</td><td>24100</td></tr></table>
Accuracy: 0.985</td></tr>
<tr><td>Experiment 19</td><td></td></tr>
<tr><td> 
</td><td>Confusion Matrix

<table><tr><td>29698</td><td>117</td><td>19</td><td>23</td><td>5</td></tr><tr><td>225</td><td>29649</td><td>0</td><td>0</td><td>0</td></tr><tr><td>135</td><td>5</td><td>29889</td><td>10</td><td>0</td></tr><tr><td>15</td><td>0</td><td>0</td><td>29730</td><td>0</td></tr><tr><td>31</td><td>0</td><td>0</td><td>0</td><td>29819</td></tr></table>
Accuracy: 0.996</td></tr>
<tr><td>Experiment 20

 
</td><td>Confusion Matrix

<table><tr><td>24543</td><td>233</td></tr><tr><td>429</td><td>24272</td></tr></table>
Accuracy: 0.987</td></tr>
</table>

The diagram below shows the comparison of the effectiveness of each experiment in relation to all the others.

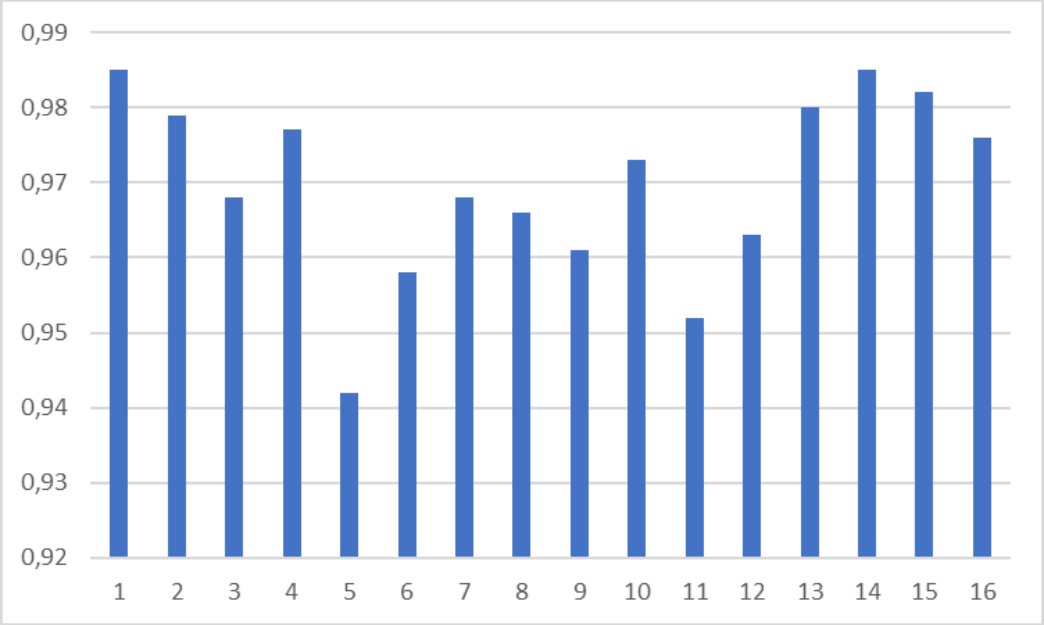


Figure 42: All experiments comparison diagram

The diagram below compares the efficiency of the classification in 5 or 2 classes, for dense neural networks.

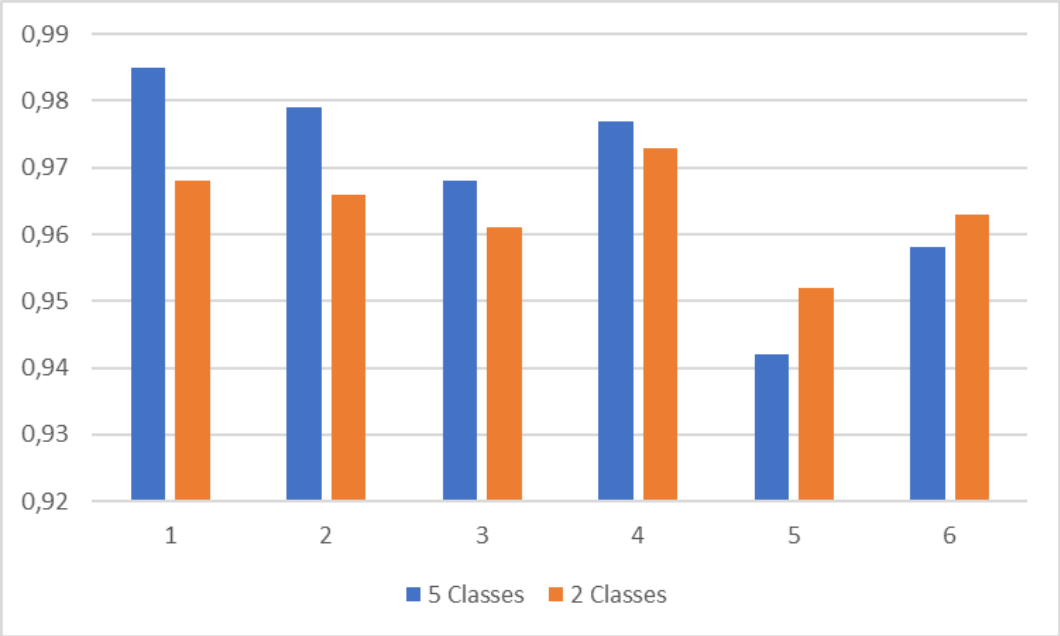


Figure 43: DNN 5 and 2 classes classification

The following plots show the comparison of the efficiency for a training data window of 6 or 2 sec and for 5 or 2 classes.

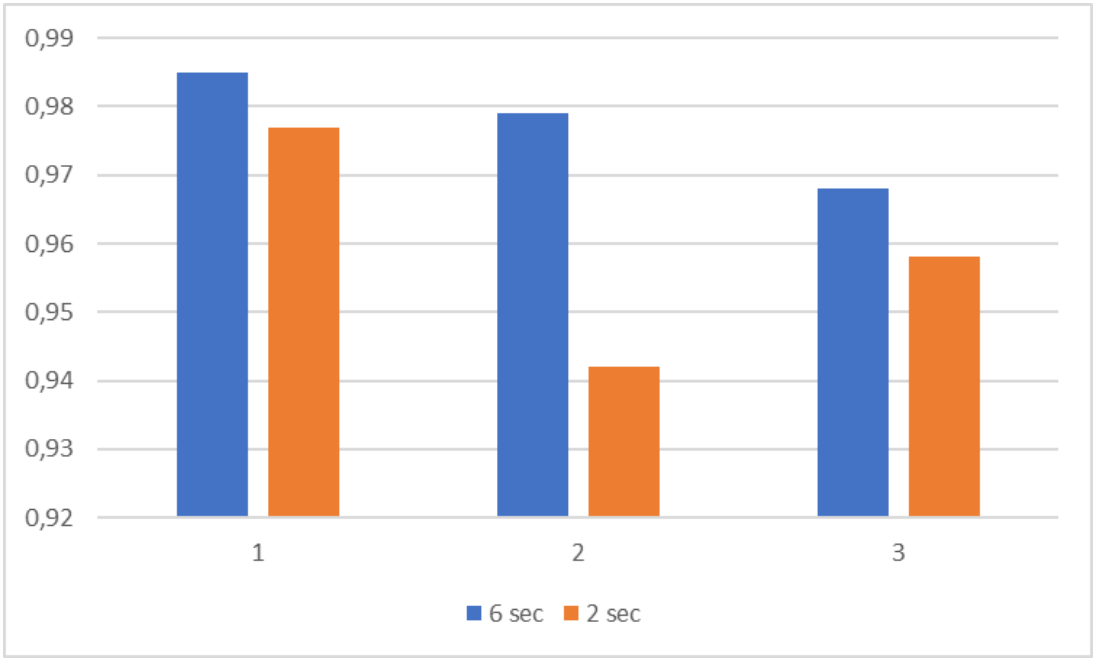


Figure 44: DNN 5 classes classification 6 and 2 secs window comparison

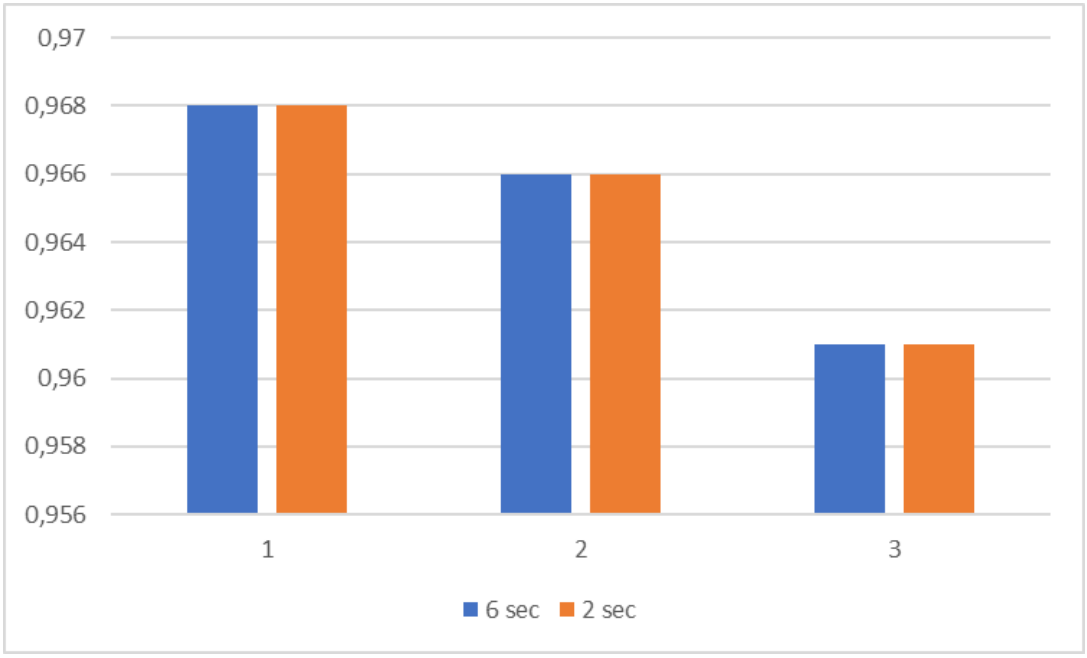


Figure 45: DNN 2 classes classification 6 and 2 secs window comparison

The next diagram compares the efficiency of Convolutional and Dense Neural Networks.

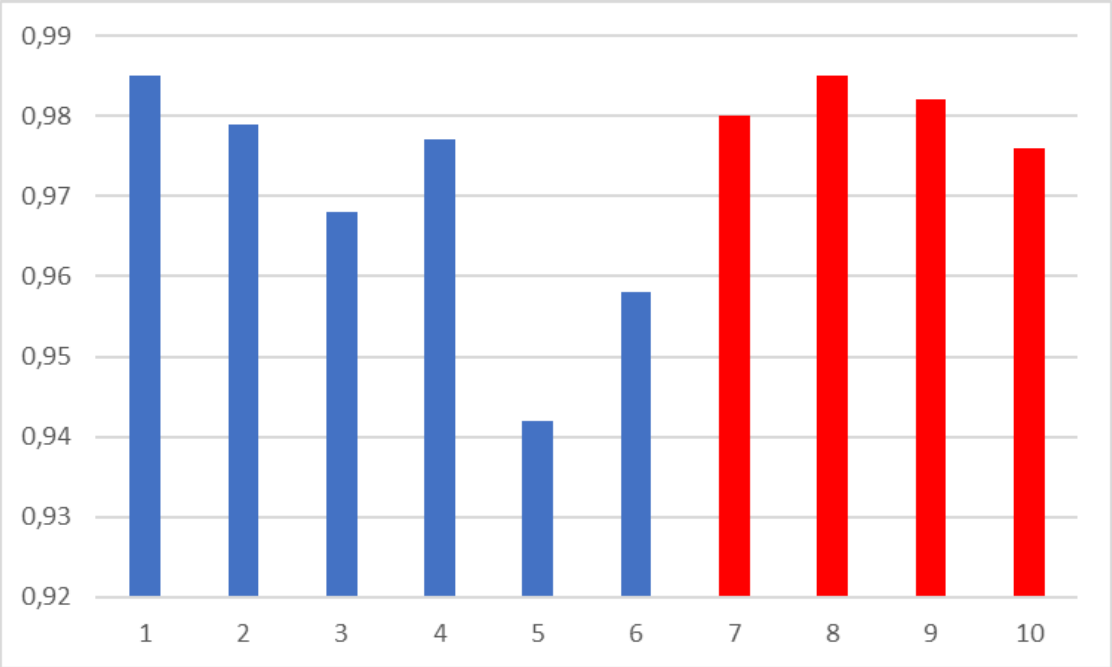


Figure 46: 5 classes DNN (blue) and CNN (red) comparison

XGBoost algorithm's results are being compared to the following diagram.

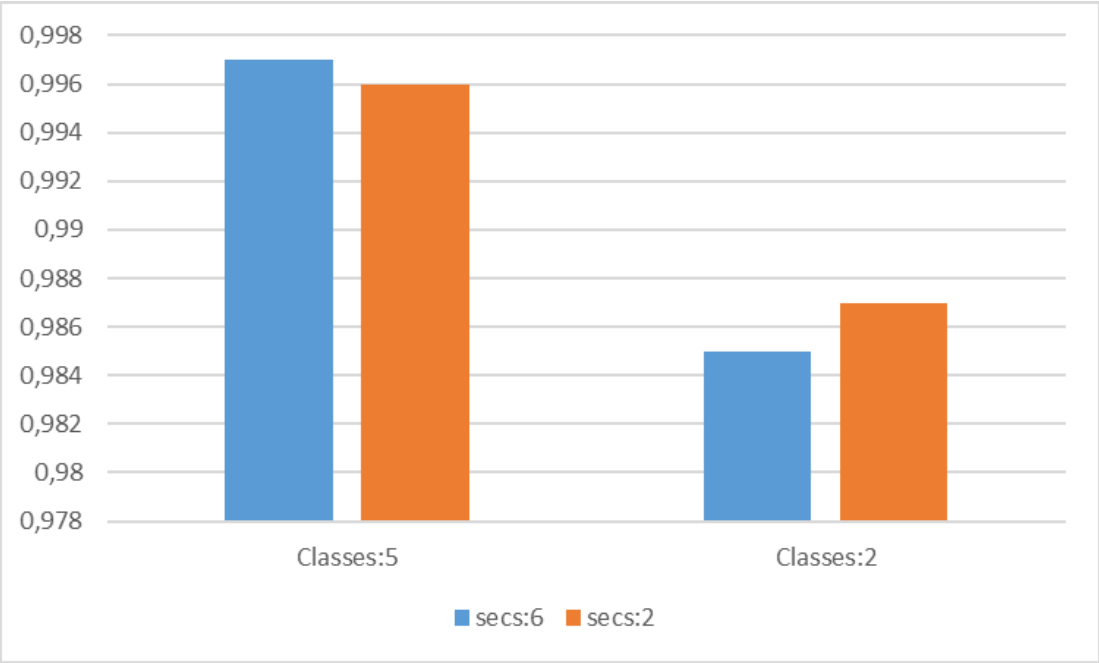


Figure 47: XGBoost Algorithm results

4.8 Discussion and Conclusions

The general conclusion derived by the current case study is that XGBoost algorithm is more reliable than the ones based on Neural Networks. Both categories of algorithms construct accurate models as their accuracy is more than 90%.

In general, this thesis explores the potential of applying machine-learning and deep-learning AI to the 12-lead ECG, highlighting its use in interpreting ECGs comprehensively and phenotyping cardiac health and diseases. While still in early stages, ongoing clinical investigations will determine the value and deployment of AI-ECG tools. Collaboration across fields is crucial for seamless integration, and proper vetting and clinician training are necessary for effective use[5].

Machine-Learning algorithms require substantial data for training and validation, posing challenges in data availability, quality, and organization. Specifically, supervised ML algorithms require labelled data, which can be labor-intensive and prone to biases. Deep learning models, while adept at analyzing complex data, lack transparency and interpretability (often called “black boxes”), making it susceptible to adversarial attacks[1].

In chapter 3.2.2 one can realize, that researchers despite the advantages and challenges of deep learning in analyzing complex medical data, recognise its potential in advancing cardiac electrophysiology research and clinical decision-making. Overall, it is important to comprehend and utilize the rich physiological information encoded in the ECG and the potential of AI to enhance its diagnostic capabilities, particularly, as a future-not-that-distant step, in enabling scalable point-of-care testing when integrated with smartphones[7].

5 References

Websites:

[S1]. https://www.mayoclinic.org

[S2]. https://my.clevelandclinic.org

Research papers:

[1]. Natalia A. Trayanova, Dan M. Popescu, Julie K. Shade
Machine Learning in Arrhythmia and Electrophysiology
Circulation Research. 2021;128:544–566. DOI: 10.1161/CIRCRESAHA.120.317872

[2]. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E, Dudley JT.
Artificial intelligence in cardiology.
J Am Coll Cardiol. 2018;71:2668−2679. doi: 10.1016/j.jacc.2018.03.521

[3]. Krittanawong C, Zhang HJ, Wang Z, Aydar M, Kitai T.
Artificial intelligence in precision cardiovascular medicine.
J Am Coll Cardiol. 2017;69:2657−2664.

[4]. Lyon A, Minchole A, Martinez JP, Laguna P, Rodriguez B.
Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. J R Soc Interface. 2018;15:20170821.

[5]. Konstantinos C. Siontis, Peter A. Noseworthy, Zachi I. Attia, Paul A. Friedman
Artificial intelligence-enhanced electrocardiography in cardiovascular disease management
Nature Reviews | Cardiology (2021). https://doi.org/10.1038/s41569-020-00503-2

[6]. Xue, J.; Yu, L.
Applications of Machine Learning in Ambulatory ECG.
Hearts 2021, 2, 472–494. https://doi.org/10.3390/hearts2040037

[7]. Zachi I. Attia, David M. Harmon, Elijah R. Behr, and Paul A. Friedman
Application of artificial intelligence to the electrocardiogram
European Heart Journal (2021) 00, 1–15. doi:10.1093/eurheartj/ehab649

[8]. Albert K. Feeny, Mina K. Chung, Anant Madabhushi, Zachi I. Attia, Maja Cikes, et. al
Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology
Circ Arrhythm Electrophysiol. 2020;13:e007952. DOI: 10.1161/CIRCEP.119.007952

[9]. Nygårds ME, Hulting J.
An automated system for ECG monitoring.
Comput Biomed Res. 1979;12:181–202. doi: 10.1016/0010-4809(79)90015-6 [PubMed: 371910]

[10]. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, Marwick TH.
Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review.
J Am Coll Cardiol. 2019;73:1317–1335. doi: 10.1016/j.jacc.2018.12.054 [PubMed: 30898208]

[11]. Saman Parvaneh, Jonathan Rubina, Saeed Babaeizadeh, Minnan Xu-Wilson
Cardiac arrhythmia detection using deep learning: A review
Journal of Electrocardiology 57 (2019) S70–S74. https://doi.org/10.1016/j.jelectrocard.2019.08.004

[12]. Ana Mincholé, Julià Camps, Aurore Lyon, Blanca Rodríguez
Machine learning in the electrocardiogram
Journal of Electrocardiology 57 (2019) S61–S64. https://doi.org/10.1016/j.jelectrocard.2019.08.008

[13] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank

resource with deep phenotyping and genomic data.
Nature. 2018;562:203. https://doi.org/10.1038/s41586-018-0579-z.

[14] LeCun Y, Bengio Y, Hinton G.
Deep learning.
Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.

[15] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

[16] Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al.
Automated cardiovascular magnetic resonance image analysis with fully convolutional networks.
J Cardiovasc Magn Reson. 2018;20:65. https://doi.org/10.1186/s12968-018-0471-x.

[17] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al.
Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms
using a deep neural network. Nat Med. 2019;25:65. https://doi.org/10.1038/s41591-018-0268-3.

[18] Attia ZI, Kapa S, Lopez-Jimenez F,McKie PM, Ladewig DJ, SatamG, et al.
Screening for cardiac contractile dysfunction using an artificial intelligence–enabled
electrocardiogram. Nat Med. 2019;25:70. https://doi.org/10.1038/s41591-018-0240-2.

[19]. Holmvang L, Lüscher MS, Clemmensen P, Thygesen K, Grande P.
Very early risk stratification using combined ECG and biochemical assessment in patients with
unstable coronary artery disease (A thrombin inhibition in myocardial ischemia [TRIM] substudy).
The TRIM Study Group. Circulation. 1998;98:2004–2009. doi: 10.1161/01.cir.98.19.2004

[20]. Voss A, Dietz R, Fiehring H, Kleiner HJ, Kurths J, Saparin P, Vossing HJ, Witt A.
High resolution ECG, heart rate variability and nonlinear dynamics: tools for high-risk stratification.
Proceedings of Computers in Cardiology Conference. 1993;261–264.

[21]. J. Park, J. An, J. Kim, S. Jung, Y. Gil, Y. Jang, K. Lee, Il-young Oh
Study on the use of standard 12-lead ECG data for rhythm-type ECG classification problems
Computer Methods and Programs in Biomedicine. 214 (2022) 106521

[22]. G.M. Marcus , The Apple Watch can detect atrial fibrillation: So what now?
Nat. Rev. Cardiol. 17 (3) (2020) 135–136 .

[23]. P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng.
Cardiologist- level arrhythmia detection with convolutional neural networks.
arXiv 2017, arXiv:1707.01836, (2011).

[24]. A.Y. Hannun , P. Rajpurkar , M. Haghpanahi , G.H. Tison , C. Bourn , M.P. Turakhia , A.Y. Ng.
Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a
deep neural network. Nat. Med. 25 (1) (2019) 65–69 .

[25]. Lee S, Kim J, Lee M. A real-time ECG data compression and transmission algorithm for an e-
health device. IEEE Trans Biomed Eng 2011;58(9):2448–55.

[26]. Jha CK, Kolekar MH. Classification and compression of ECG signal for holter device.
In: Biomedical signal and image processing in patient care. IGI Global; 2018. p. 46–63.

[27]. M.H. Kolekar, C.K. Jha and P. Kumar.
ECG Data Compression Using Modified Run Length Encoding of Wavelet Coefficients for Holter
Monitoring. IRBM, https://doi.org/10.1016/j.irbm.2021.10.001

[28]. Benjamin A. Teplitzky, Michael McRoberts, Hamid Ghanbari.
Deep learning for comprehensive ECG annotation.
Heart Rhythm 2020;17:881–888. https://doi.org/10.1016/j.hrthm.2020.02.015

[29]. Poon K, Okin PM, Kligfield P.
Diagnostic performance of a computer-based ECG rhythm algorithm.
J Electrocardiol 2005;38:235–238.

[30]. Schl€apfer J, Wellens HJ.
Computer-interpreted electrocardiograms: benefits and limitations.
J Am Coll Cardiol 2017;70:1183–1192.

[31]. Haiyan Wang, Yanjie Zhou, Bing Zhou, Xiangdong Niu, Hua Zhang.
Interactive ECG annotation: An artificial intelligence method for smart ECG manipulation.
Information Sciences 581 (2021) 42–59. https://doi.org/10.1016/j.ins.2021.08.095

[32]. Raimond L. Winslow, Stephen Granite, and Christian Jurado.
WaveformECG: A Platform for Visualizing, Annotating, and Analyzing ECG Data
Comput Sci Eng. 2016 ; 18(5): 36–46. doi:10.1109/MCSE.2016.91.

[33]. Khoo CW, Krishnamoorthy S ,Lim HS, Lip GY. Atrial fibrillation, arrhythmia burden and thrombogenesis. Int J Cardio l2012;157:318–23.

[34]. Sanna T, Diener HC, Passman RS, et al.
Cryptogenic stroke and underlying atrial fibrillation.
N Engl J Med 2014; 370: 2478–86.

[35]. Gladstone DJ, Spring M, Dorian P, et al.
Atrial fibrillation in patients with cryptogenic stroke.
N Engl J Med 2014; 370: 2467–77.

[36]. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X.
Estimates of current and future incidence and prevalence of atrial fibrillation in the U.S. adult population. Am J Cardiol 2013; 112: 1142-7.

[37]. Weng LC, Preis SR, Hulme OL, et al.
Genetic predisposition, clinical risk factor burden, and lifetime risk of atrial fibrillation.
Circulation 2018; 137: 1027-38.

[38]. Kim, D. et al. 10-year nationwide trends of the incidence, prevalence, and adverse outcomes of non-valvular atrial fibrillation nationwide health insurance data covering the entire Korean population. Am. Heart J. 202, 20–26 (2018).

[39]. Lane CM, Bos JM, Rohatgi RK, Ackerman MJ. Beyond the length and look of repolarization: defining the non-QTc electrocardiographic profiles of patients with congenital long QT syndrome. Heart Rhythm. 2018;15(9):1413-1419. doi:10.1016/j.hrthm.2018.04.033

[40]. Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P , Mark R, et al.
PhysioBank, PhysioToolkit, and PhysioNet: components of a newer search resource for complex physiologic signals. Circulation2000;101:e215.

[41]. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database.
IEEE Eng Med BiolMag2001;20:45–50.

[42]. Nolle F, Badura F, Catlett J, Bowser R, Sketch M. CREI-GARD.
A new concept in computerized arrhythmia monitoring systems.
ComputersinCardiology1986;13:515–8.

[43]. Moody G. A new method for detecting atrial fibrillation using RR intervals.
ComputersinCardiology1983:227–30.

[44]. Clifford G, Liu C, Moody B, Silva I, Li Q, Johnson A, et al.
AF classification from a short single lead ECG recording: the PhysioNet computing in cardiology challenge 2017 presented at the Computing in Cardiology Rennes-France;2017.

[45]. Ribeiro, A. H. et al.
Automatic diagnosis of the 12-lead ECG using a deep neural network.
Nat. Commun. 11,1760 (2020).

[46]. Attia, Z. I. et al. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. J. Cardiovasc. Electrophysiol. 30, 668–674 (2019).

[47]. Adedinsewo, D. et al. An artificial intelligence enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea.
Circ. Arrhythm. Electrophysiol. 13, e008437 (2020).

[48]. Attia, Z. I. et al. An artificial intelligence- enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction.
Lancet 394, 861–867 (2019). http://dx.doi.org/10.1016/S0140-6736(19)31721-0.

[49]. Wei-Yin Ko, Konstantinos C. Siontis, Zachi I. Attia, et al.
Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled ECG
Journal of the American College of Cardiology.vol.75, no.7, 2020 doi:10.1016/j.jacc.2019.12.030

[50]. Conner D. Galloway, Alexander V. Valys, Jacqueline B. Shreibati, Daniel L. Treiman, et al.
Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram
JAMA Cardiol. 2019;4(5):428-436. doi:10.1001/jamacardio.2019.0640

[51]. Zachi I. Attia, Paul A. Friedman, Peter A. Noseworthy, Francisco Lopez-Jimenez, et al. (2019)
Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs
Circ Arrhythm Electrophysiol. 2019;12:e007284. DOI: 10.1161/CIRCEP.119.007284

[52]. Noseworthy, P. A. et al. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis.
Circ. Arrhythm. Electrophysiol. 13, e007988 (2020).

[53]. Raghunath, S. et al.
Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network.
Nat. Med. 26, 886–891 (2020).

[54]. Zhu, H. et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study.
Lancet Digit. Health 2, E348–E357 (2020).

[55]. Tison, G. H., Zhang, J., Delling, F. N. & Deo, R. C.
Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery.
Circ. Cardiovasc. Qual. Outcomes 12, e005289 (2019).

[56]. Tison, G. H. et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol. 3, 409–416 (2018).

[57]. Chen, T. M., Huang, C. H., Shih, E. S. C., Hu, Y. F. & Hwang, M. J.
Detection and classification of cardiac arrhythmias by a challenge- best deep learning neural network model. iScience 23, 100886 (2020).

[58]. Feeny, A. K. et al. Machine learning of 12-lead QRS waveforms to identify cardiac resynchronization therapy patients with differential outcomes.
Circ. Arrhythm. Electrophysiol. 13, e008210 (2020).

[59]. Zachi I. Attia, C. V. DeSimone, John J. Dillon, Yehu Sapir, Virend K. Somers, et al.
Novel Bloodless Potassium Determination Using a Signal-Processed
Journal of American Heart Association. 2016;5:e002746 doi:10.1161/JAHA.115.002746.

[60]. Kovesdy CP. Management of hyperkalaemia in chronic kidney disease.
Nat Rev Nephrol. 2014;10 (11):653-662. doi:10.1038/nrneph.2014.168

[61]. Weiner ID, Wingo CS. Hyperkalemia: a potential silent killer.
J AmSoc Nephrol. 1998;9(8):1535-1543.

[62]. Sood MM, Sood AR, Richardson R.
Emergency management and commonly encountered outpatient scenarios in patients with hyperkalemia. Mayo Clin Proc. 2007;82(12):1553-1561. doi:10.1016/S0025-6196(11)61102-6

[63]. Attia ZI, Sugrue A, Asirvatham SJ, Ackerman MJ, Kapa S, Friedman PA, et al. (2018)
Noninvasive assessment of dofetilide plasma concentration using a deep learning (neural network) analysis of the surface electrocardiogram: A proof of concept study.
PLoS ONE 13(8): e0201059. https://doi.org/10.1371/journal.pone.0201059

[64]. Malik M, Hnatkova K, Kowalski D, Keirns JJ, van Gelderen EM.
QT/RR curvatures in healthy subjects: sex differences and covariates.
Am J Physiol Heart Circ Physiol. 2013;305:H1798–H1806. doi: 10.1152/ajpheart.00577.2013

[65]. Salama G, Bett GC.
Sex differences in the mechanisms underlying long QT syndrome.
Am J Physiol Heart Circ Physiol. 2014;307:H640–H648. doi:10.1152/ajpheart.00864.2013

[66]. Ball RL, Feiveson AH, Schlegel TT, Starc V, Dabney AR.
Predicting “heart age” using electrocardiography.
J Pers Med. 2014;4:65–78. doi:10.3390/jpm4010065

[67]. R.E. Gregg, S.W. Smith, S. Babaeizadeh.
12-Lead ECG interpretation by database comparison.
Journal of Electrocardiology 57 (2019) S79–S85

[68]. Maron BJ, Haas TS, Murphy CJ, Ahluwalia A, Rutten-Ramos S.
Incidence and causes of sudden death in U.S. college athletes.
J Am Coll Cardiol 2014;63:1636–43.

[69]. McLeod CJ, Ackerman MJ, Nishimura RA, Tajik AJ, Gersh BJ, Ommen SR.
Outcome of patients with hypertrophic cardiomyopathy and a normal electrocardiogram.
J Am Coll Cardiol 2009;54:229–33.

[70]**.** Martha Grogan, F. Lopez-Jimenez, M. Cohen-Shelly, Zachi I Attia, et al.
Artificial Intelligence-Enhanced Electrocardiogram for the Early Detection of Cardiac Amyloidosis
Mayo Clin Proc. 2021 Nov;96(11):2768-2778. doi: 10.1016/j.mayocp.2021.04.023. Epub 2021 Jul 2.

[71]. A. H Kashou, Jose R Medina-Inojosa, Peter A Noseworthy, R. J Rodeheffer, et al.
Artificial Intelligence-Augmented Electrocardiogram Detection of Left Ventricular Systolic Dysfunction in the General Population
Mayo Clin Proc. 2021 Oct;96(10):2576-2586. doi: 10.1016/j.mayocp.2021.02.029. Epub 2021 Jun 10.

[72]. A. S Tseng, V. Thao, B. J Borah, I. Z. Attia, J. M. Inojosa, et al.
Cost Effectiveness of an Electrocardiographic Deep Learning Algorithm to Detect Asymptomatic Left Ventricular Dysfunction
Mayo Clin Proc. 2021 Jul;96(7):1835-1844. doi: 10.1016/j.mayocp.2020.11.032. Epub 2021 Jun 9.

[73]. S. Shrivastava, M. Cohen-Shelly, Z. I Attia, A. N Rosenbaum, et al.
Artificial Intelligence-Enabled Electrocardiography to Screen Patients with Dilated Cardiomyopathy
Am J Cardiol. 2021 Sep 15;155:121-127. doi: 10.1016/j.amjcard.2021.06.021. Epub 2021 Jul 24.

[74]. M. Cohen-Shelly, Zachi I Attia, P. A Friedman, Saki Ito, et al.
Electrocardiogram screening for aortic valve stenosis using artificial intelligence
Eur Heart J. 2021 Aug 7;42(30):2885-2896. doi: 10.1093/eurheartj/ehab153.

[75]. Marco V. Perez, K. W. Mahaffey, H. Hedlin, J. S. Sumsfed, et al.
Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation
The New England Journal of Medicine. 381;20 nejm.org 2019.

[76]. Yong-Soo Baek, Sang-Chul Lee, Wonik Choi, Dae-Hyeok Kim
A new deep learning algorithm of 12-lead electrocardiogram for identifying atrial fibrillation during sinus rhythm
Scientific Reports. (2021) 11:12818. doi:10.1038/s41598-021-92172-5

[77]. J. Martijn Bos, Zachi I. Attia, David E. Albert, Peter A. Noseworthy, et al.
Use of Artificial Intelligence and Deep Neural Networks in Evaluation of Patients With Electrocardiographically Concealed Long QT Syndrome From the Surface 12-Lead ECG
JAMA Cardiol. 2021;6(5):532-538. doi:10.1001/jamacardio.2020.7422

[78]. Rohatgi RK, Sugrue A, Bos JM, et al.
Contemporary outcomes in patients with long QT syndrome.
J AmColl Cardiol. 2017;70(4):453-462. doi:10.1016/j.jacc.2017.05.046

[79]. John R. Giudicessi, Matthew Schram, J. Martijin Bos, et al.
Artificial Intelligence–Enabled Assessment of the Heart Rate Corrected QT Interval Using a Mobile Electrocardiogram Device
Circulation. 2021;143:1274–1286. DOI: 10.1161/CIRCULATIONAHA.120.050231

[80]. Giudicessi JR, Noseworthy PA, Ackerman MJ.
The QT Interval: an emerging vital sign in the precision medicine era?
Circulation. 2019;139:2711-2713.

[81]. He K, Zhang X, Ren S and Sun J.
Deep residual learning for image recognition.
In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:770–778. doi: 10.1109/cvpr.2016.90

[82]. Rob Brisk, Raymond Bond, Elizabeth Bank, Alicja Piadlo, et al.
Deep learning to automatically interpret of the electrocardiogram: Do we need the raw samples?
Journal of Electrocardiology. 57 (2019) S65-S69 doi:10.1016/j.jelectrocard.2019.09.018

[83]. G. Christopoulos, J. Graff-Radford, C. L. Lopez, X. Yao, et al.
Artificial Intelligence-Electrocardiography to Predict Incident Atrial Fibrillation
Circ Arrhythm Electrophysiol. 2020 December ; 13(12): e009355. doi:10.1161/CIRCEP.120.009355.

[84]. Alejandro A Rabinstein, Micah D Yost, Louis Faust, et al.
Artificial Intelligence-Enabled ECG to Identify Silent Atrial Fibrillation in Embolic Stroke of Unknown Source
J Stroke Cerebrovasc Dis. 2021 Sep;30(9):105998. doi: 10.1016/j.jstrokecerebrovasdis.2021.105998.

[85]. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al.
Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015.
J Am Coll Cardiol.2017;70(1):1–25.

[86]. Winslow R, et al.
The CardioVascular Research Grid (CVRG) Project.
Proc AMIA Summit on Translational Bioinformatics. 2011:77–81.

[87]. Moody GB, Mark RG, Goldberger AL.
PhysioNet: A Web-Based Resource for the Study of Physiologic Signals.
IEEE Eng Medicine and Biology Magazine. 2001; 20(3):70–75.

[88]. Tang DH, Gilligan AM, Romero K.
Economic burden and disparities in healthcare resource use among adult patients with cardiac arrhythmia. Appl Health Econ Health Policy2014;12:59–71.

[89]. van Rossum G. Python tutorial, technical report CS-R9526.
Amsterdam: Stichting Mathematisch Centrum, 1995.

[90]. S. Ragunath, A. E. Ulloa Cerna, L. Jing, D. P. vanMaanen, et al.
Deep neural networks can predict mortality from 12-lead electrocardiogram voltage data.
Journal of Electrocardiology. (2019)

[100]. Abade, A., Fereira, P. A., & Vidal, F. B. (2021). Plant diseases recognition on images using convolutional neural networks: A systematic review .

[101]. Bernal, J., Kushibar, K., Asfaw, D., Valverde, S., Oliver, A., Martí, R., & Lladó, X. (2019). Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review.

[102]. Drahokoupil, J. (2022). Application of the XGBoost algorithm and Bayesian optimization for the Bitcoin price prediction during the COVID-19 period. https://wp.ffu.vse.cz/pdfs/wps/2022/01/06.pdf

[103]. Farag, M. (2023, 1 17). A Tiny Matched Filter-Based CNN for Inter-Patient ECG Classification and Arrhythmia Detection at the Edge. mdpi: https://www.mdpi.com/1424-8220/23/3/1365

[104]. geeksforgeeks. (2023). Gradient Boosting in ML.
geeksforgeeks: https://www.geeksforgeeks.org/ml-gradient-boosting/

[105]. Girshick, R. (2016). Fast R-CNN. https://arxiv.org/pdf/1504.08083.pdf

[106]. Iglesias, L. L., Bellón, P. S., Barrio, A. P., Fernández-Miranda, P. M., González, D. R., Vega, J. A., . . . Blanco, J. A. (2021). A primer on deep learning and convolutional neural networks for clinicians.

[107]. Lomte, S., & Torambekar, S. G. (2021). Implementation on Uncertain Numerical Dataset Using XGBOOST Bagging and IndRNN.
https://www.milliyaresearchportal.org/sites/default/files/JAAST%20V8,3%20CS%20001.pdf

[108]. Parka, J., Ana, J., Kima, J., Junga, S., Gil, Y., Janga, Y., . . . Oh, I.-y. (2022). Study on the use of standard 12-lead ECG data for rhythm-type ECG classification problems.

[109]. PhysioNet. (2023). Frequently Asked Questions about PhysioNet.
https://archive.physionet.org/faq.shtml

[110]. Potter, L. (2023, 9 25). Understanding an ECG. https://geekymedics.com/understanding-an-ecg/

[111]. Teplitzky, B., McRoberts, M., & Ghanbari, H. (2020). Deep learning for comprehensive ECG annotation.

[112]. Winslow, R., Granite, S., & Jurado, C. (2016). WaveformECG: A Platform for Visualizing, Annotating, and Analyzing ECG Data.

[113]. Aloysius, N. (2017). A review on deep convolutional neural networks.
https://ieeexplore.ieee.org/abstract/document/8286426

[114]. Alzubaidi, L., Zhang, J., Humaidi, A., Al-Dujaili, A., Duan, Y., Al-Shamma, O., . . . Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. https://pubmed.ncbi.nlm.nih.gov/33816053/

[115]. Batista, K. V. (2015, 1 1). Neural Networks and Deep Learning.
https://d1wqtxts1xzle7.cloudfront.net/62971418/neuralnetworksanddeeplearning20200415-115041-1t7vxpc-with-cover-page-v2.pdf?Expires=1664988149&Signature=Z-SU9TasgTUpGz3D~gU1Od5Jb2pskjxXoNmLNN7vEHKKsslqTeASkFzVmtjkiTuYCk1F5cBA-Yi9yfhE072Z9WCLMH5llRtGsydF86k

[116]. Bhardwaj, A. (2020, 5 26). Silhouette Coefficient.
https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c

[117]. Brownlee, J. (2023). A Gentle Introduction to k-fold Cross-Validation.
https://machinelearningmastery.com/k-fold-cross-validation/

[118]. Chalkidis, M., & Vazirgiannis, M. (2005). Knowledge mining from web databases. Athens: Tipothito.

[119]. European Parliament. (2023). Artificial intelligence: threats and opportunities. https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities

[120]. Gurney, K. (2004). An introduction to neural networks. https://www.inf.ed.ac.uk/teaching/courses/nlu/assets/reading/Gurney_et_al.pdf

[121]. Iglesias, L., Bellón, P., Barrio, A., Fernández-Miranda, P. M., González, D. R., Vega, J., . . . Blanco, J. A. (2021). A primer on deep learning and convolutional neural networks for clinicians. https://pubmed.ncbi.nlm.nih.gov/34383173/

[122]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. https://www.nature.com/articles/nature14539

[123]. Louridas, A. (2017). Real-World Algorithms.

[124]. Mishra, A. (2018, 2 24). Metrics to Evaluate your Machine Learning Algorithm. https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

[125]. Pupale, R. (2018, 6 16). Support Vector Machines(SVM) — An Overview. https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989

[126]. Rebala, G., Ravi, A., & Churiwala, S. (2019, 1 1). An Introduction to Machine Learning. https://link.springer.com/content/pdf/10.1007/978-3-030-15729-6.pdf

[127]. Trayanova, N., Popescu, D., & Shade, J. (2021). Machine Learning in Arrhythmia and Electrophysiology. https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.120.317872