Ministry of Education and Science of Ukraine
Ternopil Ivan Puluj National Technical University

Faculty of Computer Information System and Software Engineering
(full name of faculty)
Department of Computer Science
(full name of department)

QUALIFYING PAPER

For the degree of

Master's thesis
(degree name)

topic: Heart Disease Prediction Using Machine Learning Alghorithms

Submitted by: student IV course, group ICAm-62
specialty 124 System analysis
(шифр і назва спеціальності)

		Ali Zulaikha Naadu
	(signature)	(surname and initials)
Supervisor		Zolotyi R.Z.
	(signature)	(surname and initials)
Standards verified by		
	(signature)	(surname and initials)
Head of Department		Bodnarchuk I.O.
	(signature)	(surname and initials)
Reviewer		
	(signature)	(surname and initials)

Ternopil
2024

Ministry of Education and Science of Ukraine

Ternopil Ivan Puluj National Technical University

Faculty Faculty of Computer Information System and Software Engineering
(full name of faculty)

Department Department of Computer Science
(full name of department)

APPROVED BY

Head of Department

_______________ Bodnarchuk I.O.
(signature) (surname and initials)

« » 20___

ASSIGNMENT

for QUALIFYING PAPER

for the degree of Master's theis
(degree name)

specialty 124 System analysis
(code and name of the specialty)

student Ali Zulaikha Naadu
(surname, name, patronymic)

1. Paper topic Heart Disease Prediction Using Machine Learning Alghorithms

Paper supervisor Zolotyi R.Z., PhD
(surname, name, patronymic, scientific degree, academic rank)

Approved by university order as of « 20 » 12 2024 № 4/7-1201

2. Student's paper submission deadline 20.01.2025

3. Initial data for the paper Literature sources about architecture, principles of operation and development of information systems.

4. Paper contents (list of issues to be developed)

5. List of graphic material (with exact number of required drawings, slides)

6. Advisors of paper chapters

Chapter	Advisor's surname, initials and position	Signature, date	
		assignment was given by	assignment was received by
Occupational health and			
emergency safety			

7. Date of receiving the assignment 20.01.2025

TIME SCHEDULE

LN	Paper stages	Paper stages deadlines	Notes
1	Analysis of the task for qualifying work. Selection		*Completed*
	and work with literary sources.		
2	Writing chapter 1		*Completed*
3	Writing chapter 2		*Completed*
4	Writing chapter 3		*Completed*
5	Writing chapter 4		*Completed*
6	Standartization control		*Completed*
7	Plagiarism check		*Completed*
8	Preliminary defense of qualifying paper		*Completed*
9	Defense of qualifying paper		

Student ______________ Ali Zulaikha Naadu
(signature) (surname and initials)

Paper supervisor ______________ Zolotyi R.Z.
(signature) (surname and initials)

ANNOTATION

Heart Disease Prediction Using Machine Learning Alghorithms // Diploma thesis Master's degree // Ali Zulaikha Naadu // Ternopil' Ivan Puluj National Technical University, Faculty of Computer Information System and Software Engineering, Department of Computer Science // Ternopil', 2025 // P. __, Fig. – __, Tables – __, Annexes – __, References – __.

Keywords: machine learning, ssz, pytorch, program, code.

The purpose of the qualification work is to research and develop machine learning methods that can be effectively applied to predict heart disease and correct behavioral risk factors in the prevention of cardiovascular disease (CVD).

During the qualification work, the task was to predict the likelihood of heart problems in people based on information about their health and accurately identify the problem, thanks to machine learning algorithms. The results of this study can be used by doctors to establish an accurate diagnosis of patients' hearts based on the analyzed data, which will predict the possibility of developing heart disease.

CONTENT

INTRODUCTION

Cardiovascular disease (CVD) represents a global health crisis, placing a significant burden on healthcare systems and accounting for a significant proportion of mortality worldwide. The prevalence of CVD and its severe consequences have made this problem particularly relevant in our time. It is of utmost importance to address this problem in a comprehensive manner, aiming not only to mitigate the devastating impact of CVD, but also to strengthen the overall level of public health and reduce global mortality rates. To achieve this goal, machine learning has become a key tool for predictive analysis and intervention, offering new ways to detect, prevent and treat CVD. This comprehensive study delves into the formulation of the problem under consideration and the key role of machine learning in the prediction and prevention of CVD.

The global prevalence of cardiovascular disease, including conditions such as heart disease and stroke, has reached alarming levels, with consequences that transcend geographical boundaries and socioeconomic differences. This cardiovascular disease pandemic has ushered in a new era of health challenges that require innovative approaches to alleviate suffering and reduce staggering mortality rates. It is against this backdrop that problem formulation becomes a central issue [4].

The problem is multidimensional and multifaceted, affecting the lives of millions of people and affecting their quality of life. The urgency of solving this problem is due to the urgent need to develop predictive tools that could identify individuals at high risk and provide timely intervention to prevent or effectively treat cardiovascular diseases. Thus, the role of machine learning

is viewed in a significant light, offering hope and potential solutions in the search for improving public health and extending human life.

This comprehensive study will cover the multifaceted landscape of cardiovascular disease prediction and prevention through the lens of machine learning. The underlying premise is that harnessing the power of machine learning can uncover important insights from patient data, revolutionizing the healthcare landscape and saving countless lives. The following sections provide a detailed analysis of the problem statement, the importance of CVD prediction, and the methods by which machine learning algorithms are being applied to address this global healthcare challenge [5].

1. ANALYSIS OF DEVELOPMENTS ON THE TOPIC OF THE WORK

1.1. Justification of the relevance of the chosen topic

The relevance of the topic of using machine learning as a core technology for predicting cardiovascular disease based on patient data is vast and multifaceted. This relevance stems from the acute global health problems associated with cardiovascular disease and the transformative potential that machine learning offers to address these problems. The following are a few key points that highlight the importance and relevance of this topic.

Global health crisis: cardiovascular disease represents a global health crisis with a significant impact on mortality rates, health systems and economies worldwide. The urgency of addressing this crisis is underscored by the high prevalence of cardiovascular disease, making this topic of utmost relevance in the field of public health and public health.

Rising incidence: The incidence of cardiovascular disease is increasing, affecting people across age groups and demographics. This upward trend requires innovative and scalable solutions to identify and manage people at risk. Machine learning provides a data-driven approach to address this growing problem.

Big Data: With the proliferation of electronic medical records, wearable devices, and health-related apps, there is a vast amount of patient data available for analysis. Machine learning is uniquely suited to process and extract valuable insights from these large data sets, making it a highly relevant technology for cardiovascular disease prediction [4].

Precision Medicine: Machine learning enables the customization of healthcare interventions, a concept known as precision medicine. By tailoring predictions and interventions to individual patients based on their unique data profiles, machine learning facilitates a more effective and personalized approach to the prevention and treatment of CVDs [6].

Cost-effective healthcare. Effective predictive models built using machine learning can help allocate healthcare resources more efficiently, reducing unnecessary costs and optimizing resource utilization. This is especially relevant in the context of limited healthcare budgets and the need to provide quality care to a growing population [5].

Public Health Impact: Accurate prediction of cardiovascular disease using machine learning has profound public health implications. This enables the development of targeted health strategies and interventions, reducing the overall burden of cardiovascular disease on communities and countries. This relevance extends to public health policy and planning.

Improving quality of life: The focus of cardiovascular disease prediction and prevention is not only on prolonging life, but also on improving the quality of life of people at risk. By identifying people at an early stage, machine learning can help them make informed decisions and change their lifestyles, promoting healthier and more fulfilling lives [5]. Challenges of scale: The global scale of the cardiovascular disease problem requires solutions that can be applied at scale. Machine learning, with its capabilities for automated analysis and scalability, is a relevant technology to effectively address this problem.

Research Advances: Continuous advances in machine learning research and technology are constantly expanding capabilities forecasting of the SSR. The relevance of this topic is further emphasized by the potential for continuous innovation and improvement.

Interdisciplinary approach: Machine learning for cardiovascular disease prediction involves collaboration between healthcare professionals, data scientists, and researchers. This interdisciplinary approach is essential for developing comprehensive solutions, highlighting the relevance of this topic to foster collaboration across multiple domains [4].

Thus, the relevance of using machine learning as a core technology for predicting cardiovascular disease based on patient data is deeply rooted in the urgency of addressing the global health crisis. The transformative potential of machine learning, its ability to analyze large amounts of patient data, and its ability to provide personalized and data-driven solutions make it a central and extremely relevant technology in the quest to strengthen public health, reduce mortality, and improve the quality of life of people at risk of cardiovascular disease.

1.2. Areas of application

The use of learning machine for, as rule, predict cardiovascular disease on patient data based is becoming increasingly common in healthcare and research. Here are some examples of machine learning applications in this context:

Risk assessment and early detection. Machine learning technique algorithms are used to develop risk assessment models that can predict the likelihood of developing cardiovascular disease in an individual. These models take into account factors such as age, gender, family history, lifestyle choices, and biomarkers to assess risk [2].

Medical imaging: machine learning is used to analyze images of medical content, such as CT scans, echocardiograms, MRIs to detect structural abnormalities, assess heart function, and early detect signs disease of heart.

Smart Wearables: Wearable devices equipped with continuously data collect sensors on heart rate, blood pressure, and physical activity. Machine learning algorithms process this data to track changes over time, detect anomalies, and provide real-time feedback to users and healthcare providers [3].

Electronic Health Records (EHR): Intelligent Data Analysis. Machine learning is used to analyze electronic health records to extract valuable information. It can identify patterns and correlations in patient data, helping clinicians make more informed decisions about treatment and prevention.

Genomic data – genomic risk prediction: Machine learning techniques are applied to genomic data to identify genetic markers associated with cardiovascular disease. This allows for the development of more accurate predictive models based on an individual's genetic profile.

Patient stratification: Machine learning algorithms can classify patients into different risk groups based on their health data. This helps identify high-risk individuals who need more intensive monitoring and intervention.

Adherence to treatment: Machine learning models can predict adherence to treatment based on patient history and behavior patterns, which is crucial to ensuring patients stick to their prescribed treatment plan.

Telemedicine and Remote Monitoring – Remote Health Monitoring: Machine learning is used in telemedicine solutions to remotely monitor patients' vital signs and detecting anomalies, ensuring timely intervention and reducing visits for in-person needs.

Patient education and behavioral interventions: Machine learning can analyze patient data to personalized provide for lifestyle recommendations changes, such as diet and exercise, to reduce the risk of cardiovascular disease [7].

Clinical Decision Support - Risk Prediction Tools: Machine learning-based clinical decision support systems help healthcare providers make decisions about treatments and interventions by providing risk assessments and evidence-based recommendations.

Drug research and development: Machine learning is used to analyze large data sets to identify potential drug candidates for the treatment of cardiovascular diseases. It can also predict a patient's response to certain medications, creating more personalized treatment plans [11].

Population health management: Machine learning is applied to population-level health data to identify trends, clusters, and high-risk regions, allowing health authorities to implement targeted interventions and policies.

The technology's ability to process and analyze large data sets, identify subtle patterns, and provide personalized insights holds great promise for improving prevention, treatment, and overall outcomes for people at risk of or living with cardiovascular disease.

1.3. Justification of the feasibility of improving existing solutions

The importance of predicting cardiovascular disease cannot be overstated. Timely and accurate prediction of cardiovascular disease offers numerous benefits both individually and collectively.

Some of the key factors that highlight the importance of CVD prediction include:

Early intervention – identifying individuals at risk for cardiovascular disease allows for targeted interventions, such as lifestyle changes or treatment, that can significantly reduce disease severity and improve outcomes [4].

Preventive healthcare predictive models allow for the implementation of preventive healthcare interventions tailored to the specific individual, focusing on risk factors and behaviors that can be modified to reduce the likelihood of CVD.

Resource allocation – Healthcare resources are often limited. Predictive tools can help allocate these resources more effectively by identifying those at highest risk, ensuring they receive the attention and perspective they need.

Public health planning. Public health institutions can benefit from predictive models to develop and implement targeted health care and policies aimed at reducing the overall burden of cardiovascular disease.

Reduced healthcare costs – Timely prevention and intervention can lead to reduced healthcare costs associated with the treatment of CVD. This, in turn, can reduce the financial burden on individuals and healthcare systems.

Improved quality of life – Predicting and preventing cardiovascular disease not only saves lives, but also improves the quality of life of people at risk. People can make informed decisions about their health, which leads to better well-being.

Long-term health management – CVD prediction is not limited to initial risk assessment; it is also good for long-term health management. Continuous monitoring and adjustment based on predictive models can help people maintain cardiovascular health over time.

In a world of these factors, it is clear that cardiovascular disease prediction is a transformative approach to managing this global health problem. Machine learning, capable of analyzing vast amounts of care data and identifying complex patterns, has the potential to revolutionize cardiovascular disease prediction and, as a result that main, improve public health and well-being on a global scale [7].

1.4. Problem statement

At the heart of the problem is the urgent need to address the high mortality rate associated with cardiovascular disease. This imperative stems from the sheer scale of the problem and its far-reaching consequences affecting individuals, communities and nations. Cardiovascular disease is the leading cause of death worldwide, taking a significant toll on health systems and economies. The magnitude of the problem cannot be overstated, and it requires a proactive and strategic response.

The problem statement covers several critical aspects:

High mortality rate: cardiovascular diseases account for a significant proportion of global mortality, making them a major public health problem. They claim lives at an alarming rate, and their impact goes beyond mere statistics, leaving families devastated and societies devastated [4].

Prevalence: Cardiovascular disease is widespread, affecting people of all ages and backgrounds. This widespread impact highlights the need for a systemic and targeted approach to address the problem.

Quality of life: in addition to mortality, cardiovascular diseases significantly worsen the quality of life of patients. This can lead to deterioration of health and reduced productivity.

Economic Impact: Cardiovascular disease places a heavy economic burden on healthcare systems, governments and individuals. The cost of treatment, lost productivity and the need for long-term care further compound the problem.

Global impact: the problem is not limited to country or region specific, it always transcends borders and affects different populations. Therefore, it requires a coordinated global effort to effectively address the problem [5].

In light of these multifaceted challenges, problem formulation becomes a critical exercise in identifying the most effective prediction and prevention strategies. ML is a promising avenue for addressing these challenges and bringing about transformational changes in the approach to CVD. Leveraging the power of data and sophisticated algorithms, machine learning can provide insights that will enable healthcare professionals to predict and reduce cardiovascular disease risk with unprecedented accuracy and efficiency.

The main objective of this strategy is to predict the probability of heart problems in people based on their health information (cholesterol, blood pressure, age, etc.), which allows for quick and accurate identification of the problem. Ten machine learning algorithms were used in this study, including SVM, Random Forest, Gradient Boost, Decision Tree, eXtreme Gradient Boost, KNN, LR, and MLP. The proposed study analyzes performance data and evaluates several classification methods to predict heart disease. Based on the results of the examination, an accurate diagnosis of the patient's heart condition will be made. The patient's health report is used by the doctor as input. The model is built using the data as input and predicts the possibility of developing heart disease.

2. PRINCIPLES OF MACHINE LEARNING AND DATA PROCESSING ALGORITHMS

2.1. History of machine learning

Machine learning term was implemented in 1959 by Samuel Arthur, an pioneer and employee IBM in the sector of computer games and intelligence artificial. Synonym for self-learning computers was also used during this period [8].

Raytheon in 1960s, created an experimental punched-tape "learning machine" called the Cybertron for signals sonar type analyze, for electrocardiograms, and patterns of human's speech using learning by reinforcement rudimentary. It was continuously "trained" by teacher a human specialist to patterns recognize and was equipped with a "fool" button to force it to re-evaluate incorrect decisions. A representative book on machine learning research in the 1960s was Nilsson's book on Learning Machines, which focused primarily on machine learning for pattern classification. Interest in pattern recognition continued into the 1970s, as described by Duda and Hart in 1973. In 1981, a report was made on the use of training strategies to teach a neural network to recognize 40 characters (26 letters, 10 numbers, and 4 special characters) from a computer terminal.

Tom M. Mitchell provided a widely cited, more formal definition of algorithms studied in the field of machine learning: "A computer program is said to learn from experience E on some class of tasks T and a performance measure P if its performance on tasks in T, as measured by P, improves with experience E." This definition of tasks that are the subject of machine learning offers a fundamentally operational definition, rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing and

Intelligence", in which the question "Can machines think?" is replaced by the question "Can machines do what we (as thinking entities) can do?".

Modern machine learning has two goals: one is to classify data based on developed models, and the other is to predict future outcomes based on these models. A hypothetical algorithm for classifying data might use computer vision of moles combined with supervised learning to train it to classify cancerous moles. A machine learning algorithm for stock trading might inform a trader about potential future predictions.

As a scientific discipline, machine learning emerged from the search for artificial intelligence (AI). In the early days of AI as an academic discipline, some researchers were interested in making machines learn from data. They tried to approach the problem with various symbolic methods, as well as with what were then called "neural networks"; these were mainly perceptrons and other models that later turned out to be reinventions of generalized linear statistical models. They also used probabilistic reasoning, especially in automated medical diagnostics [6].

However, the increasing emphasis on a logical, knowledge-based approach caused a rift between AI and machine learning. Probabilistic systems suffered from theoretical and practical problems in data collection and representation. By 1980, expert systems had come to dominate AI, and statistics was in the ascendancy. Work on symbolic/knowledge-based learning continued within AI, leading to inductive logic programming, but the more statistical line of research was now outside the realm of AI proper, in pattern recognition and information retrieval. Neural network research was abandoned by AI and computer science at about the same time. This line was also continued outside the realm of AI/CS, as "connectionism" by researchers from other disciplines, including Hopfield, Rumelhart, and Hinton. Their major success came in the mid-1980s with the reinvention of backpropagation.

Machine learning (ML), reorganized and recognized as a distinct field, began to flourish in the 1990s. The field shifted its focus from achieving artificial intelligence to solving practical problems. It shifted its focus from the symbolic approaches inherited from AI to methods and models borrowed from statistics, fuzzy logic, and probability theory [6].

Machine learning and data mining often use the same methods and overlap to a large extent, but while machine learning focuses on making predictions based on known properties obtained from training data, data mining focuses on discovering (previously) unknown properties in data (this is the knowledge discovery analysis stage in databases). Data mining uses many machine learning techniques, but with different goals; on the other hand, machine learning also uses data mining techniques as “unsupervised learning” or as a preprocessing stage to improve the accuracy of the learner. Much of the confusion between these two research communities (which often have separate conferences and separate journals, the main exception being ECML PKDD) stems from the underlying assumptions they work with: in machine learning, performance is usually evaluated in terms of the ability to reproduce known knowledge, while in knowledge discovery and data mining (KDD) the key task is to discover previously unknown knowledge. Evaluated against known knowledge, an uninformed (unsupervised) method will easily outperform other supervised methods, while in a typical KDD task, supervised methods cannot be used due to the lack of training data [8].

Machine learning also has a close relationship with optimization: many learning problems are formulated as minimizing some loss function on training set of examples. Loss functions express the discrepancy between the predictions of the trained model and the actual instances of the problem (for example, in classification, it is necessary to assign a label to instances, and models are trained to correctly predict the previously assigned labels of the set of examples) [13].

The difference between optimization and machine learning arises from the goal of generalization: while optimization algorithms can minimize loss on a training set, machine learning is concerned with minimizing loss on unseen samples. The generalization performance of different learning algorithms is an active topic of current research, especially for deep learning algorithms.

Machine learning and statistics are closely related fields in terms of methods, but differ in their primary goal: statistics makes inferences about populations based on samples, while machine learning finds generalizable predictive models. According to Michael I. Jordan, the ideas of machine learning, from methodological principles to theoretical tools, have a long history in statistics. He also proposed the term data science as a placeholder to name the general field.

Traditional statistical analysis requires a priori selection of the model that best fits the study data set. Furthermore, only significant or theoretically relevant variables are included in the analysis based on prior experience. Vice versa, machine teaching not built on the preliminary structured models; rather data form model, revealing underlying patterns. The more variables (inputs) used to train the model, the more accurate the final model will be. Leo Breiman distinguished between two paradigms of statistical modeling: the data model and the algorithmic model, where "algorithmic model" means more-less machine learning algorithms such as Random Forest.

Some statisticians have adopted machine learning methods, leading to a combined field they call statistical learning.

Analytical and computational methods derived from the deep-rooted physics of disordered systems can be extended to large-scale problems, including machine learning, for example, to analyze the weight space of deep neural networks. Thus, statistical physics finds applications in the field of medical diagnostics.

The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory using the probabilistic approximately correct learning (PAC) model. Because training sets are finite and the future is uncertain, learning theory typically does not provide guarantees on the performance of algorithms. Instead, probabilistic bounds on performance are quite common. The variance–variance decomposition is one way to quantify the generalization error.

For best performance in a generalization context, the complexity of the hypothesis should match the complexity of the function underlying the data. If the hypothesis is less complex than the function, the model does not fit the data well enough. If the model is made more complex in response, the training error decreases. But if the hypothesis is too complex, the model is overfitted and generalization will be worse [6].

In addition to performance bounds, learning theorists study the time complexity and feasibility of learning. In computational learning theory, a computation is considered feasible if it can be completed in polynomial time. There are two types of time complexity results: Positive results show that a certain class of functions can be learned in polynomial time. Negative results show that certain classes cannot be learned in polynomial time.

2.2. Application areas and frameworks

Machine learning relies on algorithms. These algorithms are very difficult to understand and work with unless you are a data scientist or machine learning expert.

Thus, the machine learning framework simplifies machine learning algorithms. An ML framework is any tool, interface, or library that allows you to easily develop ML models without understanding the underlying algorithms [8].

There are many machine learning frameworks designed for different purposes. TensorFlow, PyTorch, and scikit-learn are probably the most popular ML frameworks. However, the choice of which framework to use will depend on the work you are trying to do. These frameworks are focused on mathematics and statistical modeling (machine learning) as opposed to training neural networks (deep learning).

TensorFlow and PyTorch are direct competitors due to their similarities. They both provide a rich set of linear algebra tools and can perform regression analysis.

Scikit-learn has been around for a long time and is most familiar to R programmers, but it has a big caveat: it is not built to run on a cluster.

Spark ML is designed to run in a cluster, as that is the essence of Apache Spark. Now let's take a closer look at some specific frameworks.

TensorFlow was developed at Google Brain and later turned into an open source project. TensorFlow can: perform regression, classification, neural networks, etc.; runs on both CPUs and GPUs.

TensorFlow is one of the de facto machine learning frameworks in use today, and it's free.

TensorFlow is a full-fledged research and production tool for machine learning. It can be very complex, but it doesn't have to be.

Like an Excel spreadsheet, TensorFlow can be used simply or in a more advanced way:

TF is simple enough for a casual user who wants to return predictions on a given dataset. TF can also work for advanced users who want to set up multiple data pipelines, transform the data according to their model, configure all the layers and parameters of their model, and train on multiple machines while preserving user privacy [6].

TF requires a deep understanding of NumPy arrays. TensorFlow is built on tensors. It is a way of working with tensors; hence, NumPy is Python’s tool. NumPy is Python’s framework for working with n-dimensional arrays (a 1-dimensional array is a vector, a 2-dimensional array is a matrix, etc.). Instead of automatically converting arrays to one-shot vectors (true-false representation), a data scientist is expected to do this task.

But TensorFlow has a rich set of tools. For example, activation functions for neural networks can do all the heavy statistical work. If we define deep learning as the ability to build neural networks, then TensorFlow does that. But it can also solve more mundane problems, like regression.

PyTorch was developed by FAIR, Facebook AI Research. In early 2018, the FAIR team merged Caffe2, another machine learning framework, into PyTorch. It is a leading competitor to TensorFlow. When engineers decide to use an ML platform, their choice usually comes down to the question:

“Do we use TensorFlow or PyTorch?” Each serves its own purpose, but are fairly interchangeable.

Like TensorFlow, PyTorch: deals with regression, classification, neural networks, etc.; runs on both CPUs and GPUs; PyTorch is considered more Pythonic. Where TensorFlow can run a model faster and with some tweaking, PyTorch is considered more customizable, following a more traditional approach to object-oriented programming through the creation of classes.

PyTorch has been found to have faster training times. This speedup is negligible for many users, but may be significant for large projects. PyTorch and TensorFlow are both under active development, so a speed comparison is unlikely to be different.

Sometimes all you need is a quick test to gauge the likely success of a hypothesis. Scikit-learn is an old standard in the data science world, and it can be useful to run quick sketches of an ML model to see if the model can be interpreted in a certain way.

Scikit is another Python package that can perform many useful machine learning tasks: linear regression, decision tree regression, random forest regression, K-nearest neighbor, SVM, stochastic gradient descent models.

Scikit provides model analysis tools like confusion matrix to evaluate how well a model performs. Many times you can start ML in scikit-learn and then move on to another framework. For example, scikit-learn has great data preprocessing tools for quickly encoding categorical data. After preprocessing your data with Scikit, you can move it to TensorFlow or PyTorch.

Spark ML – can run in clusters. In other words, it can handle really large matrix multiplications by taking pieces of the matrix and running that calculation on different servers. (Matrix multiplication is one of the most important operations in machine learning.) This requires a distributed architecture so that your computer doesn’t run out of memory and doesn’t run too long when working with large amounts of data.

Spark ML is complex, but instead of working with NumPy arrays, it allows you to work with Spark RDD data structures, which anyone who uses Spark in a big data role will understand. And you can use Spark ML to work with Spark SQL data frames, which anyone who knows most Python programmers. Thus, it creates dense and sparkly feature label vectors for you, removing some of the complexity of preparing data for input into machine learning algorithms.

Torch claims to be the simplest ML framework. It’s an old machine learning library, first released in 2002. Previously, in PyTorch, Python was the method of choice for accessing the underlying tables that Torch does its calculations on. Torch itself can be used with Lua with the LuaRocks package manager. Torch’s relative simplicity is due to its interface to the Lua programming language. Lua is really simple. There are no floats or integers, just numbers. And all objects in Lua are tables. So it’s easy to create data structures. And it provides a rich set of easy-to-understand functions for partitioning tables and appending to them.

Like TensorFlow, the basic data element in Torch is a tensor. You create one by typing torch.Tensor. The CLI (command line interface) provides built-in help and helps with indentation. People who have used Python will be relieved because it means you can type functions in place without having to start over when you make a mistake. And for those who like complexity and sparse code, Torch supports functional programming. Other new frameworks for the future include huggingface.co, which is one of the best machine learning libraries that creates good basic models for researchers, built on TensorFlow and PyTorch, and Keras, a neural network library built on TensorFlow to make machine learning modeling easy [8].

2.3. Working principles and methods

The basic function of machine learning involves two main steps: studying data, identifying patterns and creating a model based on them, and making decisions and predictions based on the models.

It does not involve explicit programming; instead, machine learning (ML) algorithms train software to become capable of learning and improving. These algorithms require vast amounts of information, and their careful study, parsing, and analysis allow them to identify patterns and develop instructions for making decisions and predictions.

ML is often considered a part of artificial intelligence; others say that only the “intelligent” part of it is a subfield of AI. In “The Book of Why,” Jude Pearl explains the difference between machine learning and artificial intelligence: machine learning relies on passive observation, while AI actively interacts with the environment. In any case, ML shares many common approaches and principles with AI.

ML-based software development relies heavily on model training. Key steps in the process include:

Selecting and preparing a training dataset. Training data is a dataset that represents other data that the ML model will process and process when solving its task [13].

Depending on the case, the training data may be labeled, meaning that it has a specific feature or belongs to a specific class that the model should identify. If the data is unlabeled, the model labels the features or assigns classes on its own.

A mandatory requirement for training data is that it should be prepared in a way that can be easily processed by the ML model. Hence, it should be analyzed, normalized, and divided into several subsets that will be used for training, testing, and evaluation.

Data preparation is very important. It is said that the work of a data scientist is 90% data preparation and only 10% training the model on the prepared data.

Choosing a machine learning method. The specifics of machine learning software development projects depend on the machine learning method the team chooses: supervised, unsupervised, reinforcement, and semi-supervised learning. The method is determined by the way the algorithms are trained[10].

Supervised learning – As the name suggests, this type of learning is done under the supervision of a data scientist. The data scientist labels the input data, sets the necessary variables, determines the desired outcomes, and determines the correctness of the estimation. This type of learning works well in the following cases:

- regression modeling;
- ensemble;
- binary classification;
- multiclass classification;
- object detection and segmentation;
- learning without supervision.

As the name suggests, this type of ML-based software development does not require a human to monitor how the algorithms are learning. Unsupervised learning relies on predefined datasets and predictions.

In this case, algorithms are trained using unlabeled training data and examine the data sets for patterns that can help select meaningful relationships and subsets of the data. This type of training is used to:

- anomaly detection;
- data clustering;
- weight loss;
- mining association.

Unsupervised learning is the basis of deep learning and neural networks.

Semi-supervised learning is a combination of supervised and unsupervised learning, meaning that data scientists provide labeled data to algorithms, but the algorithms can explore the data and develop their own understanding without human intervention. Based on the labeled data provided by humans, the algorithms identify features in the data and apply them to the unlabeled data.

Why is this approach chosen? It is easier to manage than unsupervised learning, but is more cost-effective and takes less time than supervised learning. The most prominent examples of its application in software development projects are fraud detection systems and machine translation.

Reinforcement Learning – This type of ML is used when the process that the machine needs to learn to perform consists of many steps, and each of them has well-defined rules. During the training process, data scientists provide the algorithm with feedback and cues about the quality of its performance. This will encourage the ML algorithm to seek positive rewards (reinforcements).

Good examples of using reinforcement learning are video games, robotics-based automation software, and resource management.

Choosing an algorithm to run on training data. A machine learning algorithm is a combination of data processing procedures embedded in program code.

The factors that determine the type of ML algorithms to be used are the business problem to be solved, the available resources, and the specifics of the data (whether it is labeled or unlabeled, and its quantity).

Typically, algorithms are divided into the following categories depending on the problem they are used to solve:

Classification: They choose between several classes and assign probabilities to them: logistic regression, naive Bayesian, support vector machine, K-nearest neighbors, etc.

Regression: they allow a model to predict the quantity of a variable: Naive Bayes, LARS Lasso, AdaBoost, XGBoost, Elastic Net, Random Forest, etc.

Clustering: they allow grouping similar data and labeling them according to the group they belong to: K-Means, Kohonen, TwoStep, etc. Dimensionality reduction: they allow combining or discarding insignificant data: analysis main components (PCA), linear discriminant analysis (LDA), direct feature selection, factor analysis, etc. There are two groups of ML algorithms depending on the type of data: ML algorithms, those used with labeled data (regression, instance-based algorithms, and decision trees), and those used with unlabeled data (clustering and association algorithms).

Training the algorithm to produce desired results. Training the algorithm is iterative, and each cycle includes:

- direct distribution;
- comparing results with the desired result;
- adjusting the algorithm to obtain the best results and repeating these steps until the result required for a given probability norm is obtained.

The resulting solution is a machine learning model. Once an ML model is trained, it can be used with new data to solve new business problems and gradually improved in terms of efficiency and accuracy.

Machine learning is relevant in many industries and fields and has the ability to evolve over time. Here are six real-world examples of machine learning being used.

Image recognition is a well-known and widely used example of machine learning in the real world. It can identify an object as a digital image based on pixel intensity in black and white or color images [12].

Real-world examples include labeling X-ray images as cancerous or not, assigning a name to a photographed face (aka "tagging" on social media), and recognizing handwriting by dividing a single letter into smaller images.

Machine learning is also often used to recognize faces in images. Using a database of people, the system can identify common features and match them to faces. This is often used in law enforcement.

Also speech recognition, machine learning can translate speech into text. Some software programs can convert live voice and recorded speech into a text file. Speech can also be segmented by intensity in time-frequency bands.

Real-world examples of speech recognition: voice search, voice dialing, device control.

Speech recognition software is most commonly used on devices like Google Home or Amazon Alexa.

Machine learning can help diagnose diseases. Many doctors use chatbots with speech recognition capabilities to discern patterns in symptoms.

Real-world examples of medical diagnostics: making a diagnosis or recommending a treatment option, oncology and pathology using machine learning to recognize cancerous tissue, body fluid analysis.

In the case of rare diseases, the combined use of facial recognition software and machine learning helps scan patient photos and identify phenotypes that correlate with rare genetic diseases.

Statistical arbitrage is another example of a use case. Arbitrage is an automated trading strategy used in finance to manage a large number of securities. The strategy uses a trading algorithm to analyze a set of securities using economic variables and correlations.

Real-world examples of statistical arbitrage: algorithmic trading that analyzes market microstructure, analyzing large data sets, identifying arbitrage opportunities in real time, optimizing arbitrage strategies to improve results.

Predictive analytics – Machine learning can classify available data into groups, which are then determined by rules set by analysts. Once the classification is complete, analysts can calculate the probability of error.

Real-world examples of predictive analytics: predicting whether a transaction is fraudulent or legitimate, improving a prediction system to calculate the probability of failure.

Predictive analytics is one of the most promising examples of machine learning applied to everything from product development to real estate pricing.

Extraction – Machine learning can extract structured information from unstructured data. Organizations accumulate huge amounts of data from customers. A machine learning algorithm automates the process of annotating datasets for predictive analytics tools.

Real-world mining examples: creating a model to predict vocal cord disorders, developing methods for preventing, diagnosing, and treating diseases, helping doctors quickly diagnose and treat problems

Typically, these processes are tedious. But machine learning can track and extract information to obtain billions of data samples.

Machine learning offers a wide range of methods and techniques to combat cardiovascular disease (CVD). These methods use the power of data analytics and predictive modeling to aid in the early detection, risk assessment, prevention, and treatment of CVD. Here are some key methods and techniques through which machine learning contributes to this important cause:

Feature selection and development: Machine learning algorithms can identify the most relevant characteristics or variables in patient data that indicate cardiovascular disease risk. Feature selection and development help reduce the dimensionality of the data and improve the performance of predictive models.

Classification algorithms: Machine learning classification algorithms such as logistic regression, decision trees, random forests, and support vector machines are used to classify people into different risk groups. These models can predict whether a patient has a low, moderate, or high risk of developing cardiovascular disease.

Regression models: Regression methods such as linear regression and ridge regression are used to predict continuous outcomes such as blood pressure or cholesterol levels. These models can help assess the severity of CVD risk factors.

Deep learning: Deep neural networks, a subset of machine learning, are used for tasks such as image analysis in medical imaging, allowing the identification of structural abnormalities and accurate assessment of cardiovascular disease risk.

Methods ensemble: such complex methods, as packaging and reinforcement, combine multiple machine learning models to improve accuracy prognostication and reduction excessive equipment. They used to build robust risk prediction models for CVD. Natural Language Processing (NLP): used to extract valuable information from unstructured clinical notes, reports, and patient narratives in electronic medical records. This helps to obtain the patient's history and symptoms related to CVD.

Clustering algorithms: Clustering algorithms, such as k-means and hierarchical clustering, are used to segment patients into groups based on similar characteristics, aiding in patient stratification and individualized interventions.

Anomaly detection: Anomaly detection algorithms identify unusual patterns or outliers in patient data. These techniques are crucial for early detection of abnormalities in vital signs or biomarker levels [15].

Time series analysis: Time series models are used to analyze trends and changes in patient data over time, which is especially important for monitoring and treating chronic CVDs.

Imbalanced data processing: Techniques such as oversampling, undersampling, and synthetic data generation are used to eliminate unbalanced data sets, ensuring that machine learning models provide accurate predictions for both high- and low-risk individuals.

Dimensionality reduction: Principal component analysis and t-distributed stochastic neighbor embedding (t-sne) are used to reduce the dimensionality of high-dimensional patient data while preserving important information.

Model Clarity: Methods such as shap (shapley augmentation explanations) and lime (locally interpreted model-independent explanations) help explain decisions made by machine learning models, increasing their transparency and reliability in clinical settings.

Transfer learning: Transfer learning uses pre-trained models on large datasets and fine-tunes them for specific cardiovascular disease prediction tasks, reducing the need for large labeled data.

Continuous monitoring and alerting systems: Machine learning models can be integrated into continuous monitoring systems to detect sudden changes in patient data and trigger alerts for timely intervention.

Genomic analysis: Machine learning is used to analyze genomic data to identify genetic markers associated with cardiovascular disease risk, offering a personalized risk assessment based on genetic information.

Prognostic modeling affection to treatment: models Machine learning predicts patient adherence to prescribed medications, ensuring patients effectively adhere to their treatment plans.

Real-time data processing: Machine learning enables data from wearable devices to be processed in real time, providing immediate information and interventions to patients and healthcare providers.

Interdisciplinary collaboration: Machine learning facilitates collaboration between healthcare professionals, data scientists, and researchers, encouraging an interdisciplinary approach to solving complex cardiovascular disease problems.

These methods and techniques demonstrate the versatility of machine learning in the treatment of cardiovascular diseases. Using data-driven insights and predictive modeling, machine learning facilitates early detection, risk assessment, prevention strategies, and personalized interventions, ultimately improving the treatment of cardiovascular diseases and reducing their impact on the global community.

3. DEVELOPMENT OF PROGRAM STRUCTURE AND ALGORITHMS

3.1. Model

Prediction plays a crucial role in clinical practice. In particular, accuracy plays a vital role in predicting heart diseases. Recent contributions quantify the capabilities of machine learning algorithms to achieve the same. Machine learning algorithms and methods have been applied to various medical datasets to automate the analysis of large and complex data. In this context, an experimental study points to a proposed model that defines a machine learning approach, such as classification, to detect the presence of heart diseases.

Heart disease is considered one of the most complex and deadly world diseases for human. The diagnosis and treatment of heart disease are very difficult, especially in developing, as we see, countries, due to the rare availability, as rule, of diagnostic equipment and the shortage of doctors and other resources, which affects the correct prognosis and of heart patients treatment. Accurate and correct diagnosis of cardiovascular risk disease in many patients is necessary to reduce their associated risks of serious heart problems and improve cardiac safety. Invasive methods of diagnosing disease heart are on the analysis based of the patient's history medical, report examination physical, and relevant analysis of symptoms by experts of medical type. These all methodologies mostly cause inaccurate diagnosis and often delay diagnostic results due to errors by human.

To address these challenges in diagnosing heart disease, predictive models using artificial neural networks and machine learning have been developed. These systems are trained and tested to obtain better disease prediction accuracy [12].

The machine learning field allows for the identification of hidden patterns and the creation of structures analytical, clustering including, classification, regression, and correlation, by integrating and applying various methods, such as models learning machine, information mining, and neural networks. Consequently, methods learning machine have demonstrated potential great to support decision-making clinical, assist clinical development algorithms of management and guidelines, and contribute to of evidence-based clinical practices establishment for the treatment of cardiovascular diseases (CVD). In addition, early detection of cardiovascular diseases using machine learning methods can in expensive clinical and laboratory studies reduce need, which will lead, as rule, to a reduction in the burden financial on both the healthcare system and individuals.

For a vivid example, a gradient boosting model can be used to predict the presence of cardiovascular disease and determine the predictive value most based on the rough set values. After that, a series of deep learning and machine learning methods are used to disease cardiovascular analyze. The main contributions of our study are as follows.

Using split validation and validation cross, discover a algorithm of machine learning type with performance improved to be applied to detect cardiovascular disease.

Applying appropriate feature selection techniques can optimize prediction accuracy. Using a algorithm machine learning robust can improve development early cardiovascular disease (CVD)prediction at facilitating early, stages early, intervention and key features selecting to recovery algorithms support.

Disease cardiovascular type predicting using the most advanced cardiovascular disease dataset.

Providing reliable advice to medical and healthcare professionals on changes significant in the sector healthcare.

Many researchers are studying a number of cardiac disease prediction frameworks using various data analysis methods. They use calculations various and datasets in addition to the test results and future work that will be possible on the structure, and achieve more productive results. Researchers have research numerous completed attempts to achieve effective methods and in recognizing high accuracy heart-related disorders [3].

Tran's research built system intelligent type using the Naive Bayes method modeling data mining. It is application web type in which the assistant answers questions which pre-programmed. It tries to hidden information find in the database and compares the user's values with a set data trained. It can answers provide to complex questions about diagnosing heart disease, allowing professionals healthcare type to make clinical decisions more informed than traditional support systems decision. It also reduces treatment costs by providing efficient care.

Gnaneswar demonstrates the importance of heart rate monitoring while cycling. Cyclists can manage cycling events such as cycling cadence to determine their activity level by monitoring their heart rate during acceleration. By controlling the load on the pedals, cyclists can overtraining avoid and heart failure. A cyclist’s heart rate can be used to intensity determine of exercise. Heart rate can be measured by using a wearable sensor. Unfortunately, a sensor does not all capture the regular intervals information, such as second one, seconds two, etc. So, we need a heart rate model expectation to in the gaps fill.

Gnaneswar's work aims to use Feedforward Brain Organization to build a predictive model for heart rate, taking into account cyclic rhythm. In the second, the data sources are heart rate and rhythm. The predicted result is the heart rate for the next second. Using the direct transmission structure brain, the between relationship heart rate and cycling rhythm is represented statistically. Mutiyarsa extends the administration of medical care on these arguments based. Breakthroughs numerous in communication remote have been made in of heart disease anticipation. The use of data mining (DM) methods to detect and localize coronary disease is very useful. During their evaluation, a comparative analysis of several information mining calculations for single and mixed breeds is performed to determine which calculation disease coronary most accurately predicts [15].

Yeshvendra states that of AI the use computing for predicting is growing various diseases. This concept is so diverse and important because of the AI computing ability of to have a perspective comparable human to improve the coronary disease accuracy prediction. Patil states that the correct diagnosis of heart disease is most of the one important biomedical problem that needs to be solved. Three information mining methods: support vector machine, decision tree and naive Bayes. These methods to create an emotional support network were used for their option preferred. Tripoli states that identification of diseases with high prevalence rates, such as Alzheimer's, Parkinson's, breast cancer, coronary heart disease and diabetes is one of the most important biomedical tests that needs immediate attention. Gonsalves attempted to predict CVD coronary using learning machine and medical data historical. Overview by Oikonomou provides an of the diverse information that is found in the context of chronic diseases. Using numerous machine learning techniques, they elucidated extreme value theory to better chronic measure severity disease and risks.

According to Ibrahim, learning-based machine systems can be used for the diagnosis and prediction of heart diseases. Active learning (AL) methods improve the classification accuracy by integrating feedback between the user and the expert system with sparse data. In addition, Pratiyush investigated the ensemble classifier's role in XAI system in predicting heart diseases based on datasets CVD. The work proposed used a dataset including 14 attributes and 303 instances, with integer, real and categorical types attributes, and the task classification was based on methods classification such as naive Bayes, SVM, KNN, bundling, AdaBoost and LR [9].

In the literature, attempts have been made to create predicting strategies for the heart disease diagnosis. Due to the high dimensionality of text input, many algorithms machine learning traditional cannot incorporate it into the simultaneously process prediction. As a result we can see, this paper develops and investigates a robust set algorithms of machine learning to early prediction improve of disease cardiovascular development, allowing for recovery rapid and intervention.

3.2. Materials and methods

Data always plays a key role in the prediction process in research experiments. To develop this prediction model, we considered heart disease data available in the UCI repository learning machine. The consists data of the following attributes.

S No	Attribute Name	Type of Attribute
1	Age	Real
2	Sex	Binary
3	Chest Pain Type	Nominal
4	Resting Blood Pressure	Real
5	Cholesterol	Real
6	Fasting Blood Sugar	Binary
7	Resting Electrocardiographic Results	Nominal
8	Maximum Heart Rate Achieved	Real
9	Exercise Induced Angina	Binary
10	Old peak	Real
11	Slope Of The Peak Exercise ST Segment	Ordered
12	Number Of Major Vessels	Real
13	Thal	Nominal
14	Num	Real

Figure 3.1 – Steps required to create a heart disease prediction system

The distribution of the data can be seen in the figures below.

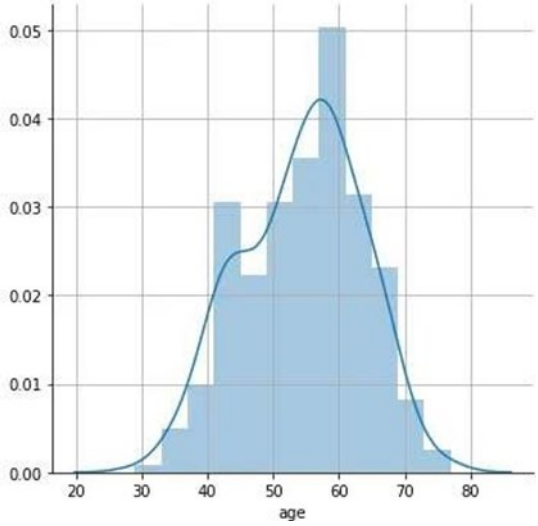

Figure 3.2 – Age distribution

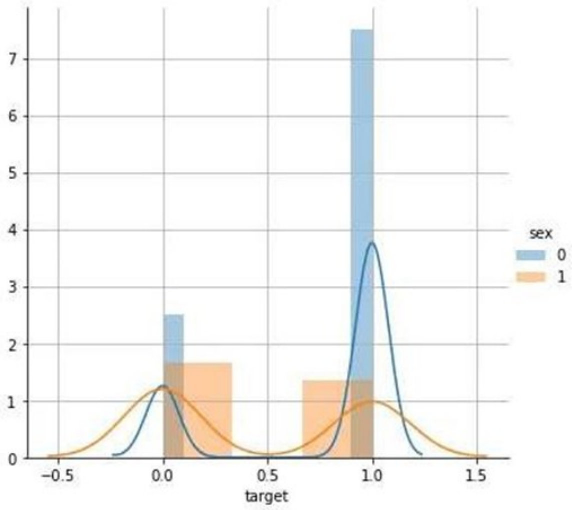

Figure 3.3 – Distribution showing that both men and women are equally susceptible to the influence

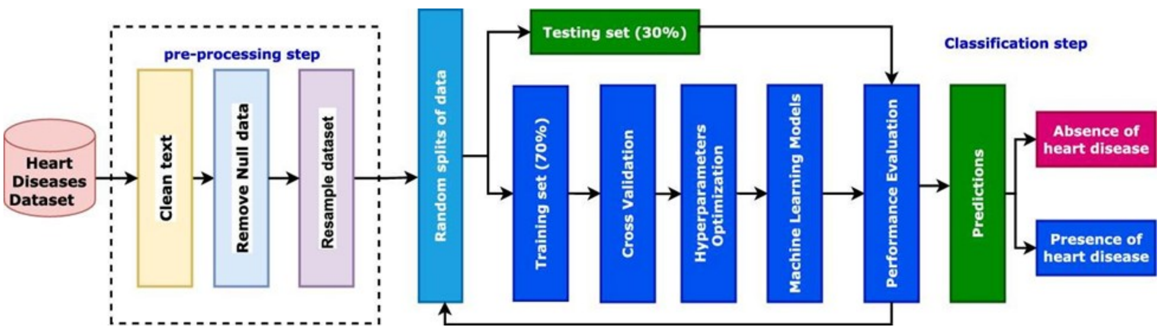

Figure 3.4 – Main stages of the proposed methodology

In the proposed scheme for classifying heart disease cases, an exploratory analysis is first performed. A comprehensive analysis of both the objective and the function is performed, and the categorical variables are converted into numerical values. Different criteria are used to compare the models under consideration. The results analyzed of model each, and the model optimal is selected for the problem under consideration. The model proposed is examined thoroughly, and the Optuna library is used to hyperparameters tune model to see how much improved they are. Proposed Model is three phases divided into: preprocessing, training, and classification (Figure 3.4).

It is important to, before training the selected models, note the missing cholesterol values, which were initially entered as 0. To do this, the data are divided based into groups on the of confirmed heart disease presence, and the mean value of group each is to fill used in the values missing. To assess whether affect variables the prediction of based on their Shapley values heart disease, terms interaction was included in the to capture any possible correlations models ` between the elements data. "SHAP" (Additive SHapley ExPlanations) uses game theory to determine the significance of each feature and to explain both individual model predictions can be used and the aggregate results model. SHAP determines each predictor's contribution magnitude to the output models by marginal contributions averaging the of over all each feature possible combinations of features.

The algorithm machine learning will be trained properly after normalizing and preprocessing the datasets. After modifying the data, they are arbitrarily classified into test set and training set, with rows assigned 70% to the training set and to the test set 30%. A common validation cross method K-fold that involves large number performing a of tests relevant to determine accuracy typical model score. This technique has been around for some time quite. To test the proposed strategy, AI procedures such as SVC, MultinomialNB, K-Neighbor, BernoulliNB, SGD, Random forest, and Decision Tree are deployed for the best outcome conditions.

XGBoost (Extreme Gradient Boosting) is a supervised learning method for improving prediction accuracy by combining multiple decision trees. XGBoost iteratively adds decision trees using gradient boosting, with each subsequent tree attempting to correct errors in previous trees. The final prediction is the weighted sum of all the individual predictions of the tree.

The Naive Bayes algorithm assigns equal weight to all features or qualities. The algorithm becomes more efficient because one feature does not affect another. According to Yasin 2020, the Naive Bayes Classifier (NBC) is a simple, efficient, and well-known text categorization algorithm. NBC uses Bayes' theorem to classify documents since the 1950s, and it is theoretically sound. A posteriori estimation is used to determine the class using the Naive Bayes classifier. Features, for example, are classified based on their highest conditional potential [15].

Naive Bayes Bernoulli is a statistical method that produces logical based results on the absence or of presence the text desired. A discrete Bernoulli distribution is introduced this classifier into. When unwanted keyword identifying or designating a specific type of word in a text fragment, this simple Bayesian classifier type is useful. It also differs from the approach multinomial in that it generates a output binary such as Yes–No, True–False, 1–0. A procedure or is system stochastic one that has a match random as part of its solution. Stochastic descent gradient (SGD) a few data samples randomizes rather than the data entire set in iteration each. As a result, instead of sum calculating of the gradients for instances all, each iteration calculates the gradient of the cost function for a single instance. SGD is an iterative method for optimal smoothness properties determining of a subdifferentiable or differentiable function objective.

A decision tree is a well-known technique learning machine in which data is partitioned repeatedly based on parameters certain. A tree has two traversal entities: leaves and nodes. Leaves decisions or outcomes represent, while decision nodes partition data. To solve problems (learning ensemble) decision trees can be used in combination. The forest random algorithm solves the overfitting problems associated with algorithms tree decision. The algorithm is able for solving classification and regression problems, as well as large number evaluating of attributes type to which ones are most important determine. Random data, as rule, can be learned without well-planned data changes.

The Neighbor K-Nearest (K-NN) algorithm classifies new based observations on their known examples distance. On the majority based of its neighbors vote and the distance function as a measurement tool, an instance is assigned to the class with the frequency among highest its neighbors k-nearest. In tasks classification, returns k-NN the membership class. Whereas in tasks regression, it returns the value of the feature property. k-NN is used for regression or classification affects the result. Since this method relies on classification distance, normalization can improve significantly the data training. If the correspond features to physical different scales or units, standardization can improve significantly the accuracy of the data training.

The research experiments used Google Colab as a platform deployment for learning machine models. The platform have virtual machine a running on servers Google and providing users Python environment access that includes data science popular libraries such as Scikit-Learn, TensorFlow, and PyTorch. Colab Google is a based cloud environment for notebooks Jupyter that offers access free to resources computing such as a machine virtual with up to 100 GB of disk hard space and 12 GB of RAM. The virtual machine memory allocated to 25 GB, and you can also large RAM enable options up to 52 GB for models large or data sheets [11]. The virtual machine runs on Google servers and is equipped with an NVIDIA Tesla K80 GPU, allowing us to efficiently train deep learning models. In addition, Google Colab provides a wide range of pre-installed tools and libraries, making easy it to set up and the use of dependencies necessary. The machine virtual runs on the Linux operating system, which ensures the stability and reliability of the implementation environment. In addition, the system operating used by the machine virtual is Ubuntu Linux, which comes with various system pre-installed tools and libraries.

The heart disease data used is a synthesis of datasets in this study from the Repository Machine Learning UCI and eleven features contains that can, as rule, be used to predict the presence of heart failure, a common cardiovascular disease that significantly increases the likelihood of cardiovascular disease. The variable target is a attribute binary that diagnosis indicates of failure heart if HeartDisease = 1 (Figure 3.5). In addition, a list of variables and descriptions of features in dataset disease heart are also presented (Figure 3.6).

Age	Sex	Type chest pain	BP resting	Cholesterol	BS fasting	ECG resting	HR max	Angina exercise	Old peak	ST slope	Disease of heart
41	M	ATA	142	287	0	Nor1	173	N	0.0	Upper	0
48	F	NAP	162	182	0	Nor1	157	N	1.0	Flat1	1
38	M	ATA	132	273	0	ST	98	N	0.0	Upper	0
49	F	ASY	136	224	0	Nor1	109	Y	1.5	Flat1	1
53	M	NAP	152	185	0	Nor1	123	N	0.0	Upper	0

Figure 3.5 – Sample Heart Failure Dataset

Variable	Interpretation
Age	Patient's Age/year
Gender	Patient's Gender, Male/Female
Type of chest pain	Type of chest pain: i. TA: Typical Angina ii. ATA: Atypical Angina iii. NAP: Non-Anginal Pain iv. ASY: Asymptomatic
Resting blood pressure	Patient's Blood Pressure/mmHg.
Total Cholesterol	Patient's Cholesterol (mg/dl).
Blood Glucose level (Fasting)	Patient's fasting blood glucose level. i. glucose >120 mg/dL =1 ii. glucose below 120 mg/dL =0
ECG at rest	Electrocardiography (at rest): i. Normal ii. ST: ST segment and/or T wave abnormality iii. LVH: Probable or Definite Left Ventricular Hypertrophy
Heart Rate at Maximum	Maximum Heart Rate, heart beats per minute.
Angina on Exercising	Exercise-associated Angina, present /absent.
Old peak	Measure of ST Depression.
ST_Slope	Slope of Peak Exercise. i. Up: up sloping ii. Flat iii. Down: down sloping

Figure 3.6 – Symptoms, signs, and laboratory tests of a set of heart diseases

Datasets	#Observations
Cleveland	303
Hungarian	294
Stalog (Heart) Data Set	270
Long Beach VA	200
Switzerland	123
Total	1190
Duplicated	272
Final dataset	918

Figure 3.7 – Different used datasets to heart disease create dataset

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease
Count	918	918	918	918	918	918	918
Max	77	200	603	1	202	6.20	1
Min	28	0	0	0	60	-2.6	0
Mean	53.51	132.39	198.79	0.23	136.81	0.89	0.55
Std	9.43	18.51	109.38	0.42	25.46	1.06	0.49
25%	47	120	173.25	0	120	0	0
50%	54	130	223	0	138	0.60	1
75%	60	140	267	0	156	1.50	1

Figure 3.8 – Summary numeric variables statistics

	Sex	TypeChestPain	ECGResting	AnginaExercise	ST_Slope
Count	920	920	920	920	920
Unique	2	3	4	2	4
Top	M	ASY	Normals	N	Flat1
Freq	735	486	562	557	470

Figure 3.9 – Summary variables categorical statistics

Variable	Value	Total patients	Proportion of heart disease
Sex	M	725	90.2%
	F	193	9.8%
ChestPainType	ASY	**496**	**77.2%**
	NAP	203	14.2%
	ATA	173	4.7%
	TA	46	3.9%
RestingECG	Normal	**552**	**56.1%**
	ST	178	23.0%
	LVH	188	20.9%
ExerciseAngina	Y	371	**62.2%**
	N	**547**	37.8%
ST_Slope	Flat	**460**	**75.0%**
	Up	395	15.4%
	Down	63	9.6%

Figure 3.10 – Share of heart disease

The dataset by merging various was created that were available independently previously and had not been combined before. In this dataset, five cardiac datasets are combined under 11 common characteristics, making it the largest cardiac disease dataset available for purposes research. The datasets specific used to curate this composite dataset (Figure 3.7).

The Heart Disease dataset contains 12 columns and 918 observations. You can see the summary of the basic statistics for the numeric characteristics (Figure 3.8). It is that clear the average age is 53 years and the 77 years is maximum (Figure 3.8). The statistics for the categorical attributes are presented in a similar way (Figure 3.9), where the unique values of the ChestPainType attribute are shown as 4 and the top value is “ASY”.

You can also see the main details of the numerical characteristics summarized (Figure 3.10). It is that clear the Gender variable two values main has: female (F) and male (M), so the disease heart proportion for F is 9.8% and for M is 90.2%. Similarly, the statistics for the attribute ChestPainType are presented (Figure 3.10), there are four values (TA, ATA, NAP, and ASY), and the frequent most is ASY 77.2%.

3.3. Results of work

It is noteworthy that the classifications in the meaning of the attribute of heart disease are quite balanced well. From 918 patients the 508 who in the study participated were diagnosed with heart failure, and 410 were not. Patients have a mean age of 57 years with disease heart, while patients without disease heart have age of 51 years a typical. As shown (Figure 3.11), about 63% of men have heart disease, while about 25% of women have heart disease [7]. A woman has a 25.91% probability of heart disease. A man has a 63.17% probability of heart disease.

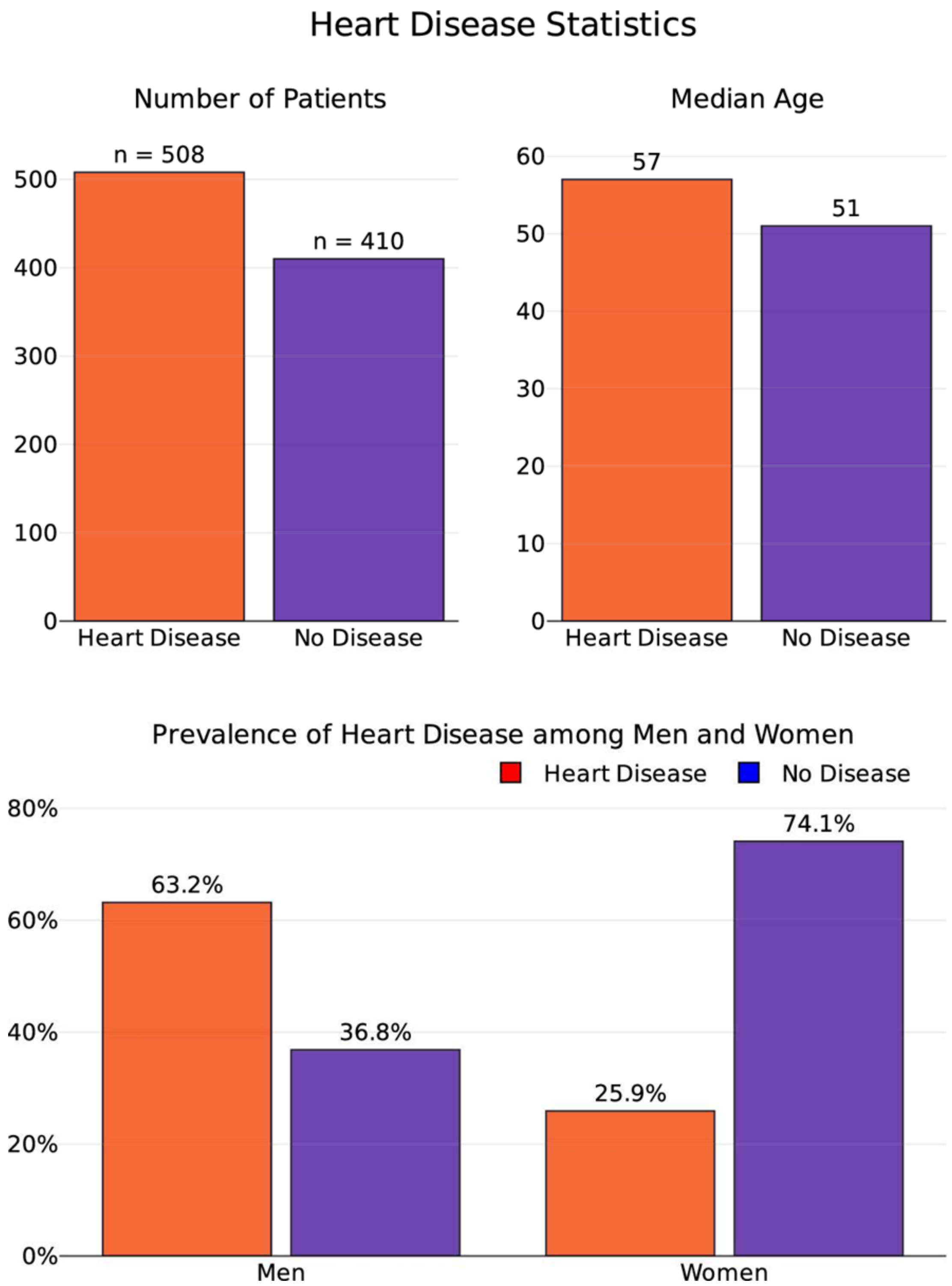

Figure 3.11 – Heart disease statistics

The ranges for disease heart for age, cholesterol, systolic pressure blood, rate heart, and depression ST-segment will be shown below (Figure 3.12). Age chart of patients with heart disease vascular disease falls between the ages of 62 and 51, as depicted in the age chart. There also are a younger few people in this category that fall outside the lower limit. People who do not have cardiovascular disease have range age that is somewhat more variable, but more even, and outliers there are no. The majority vast of patients who fall are quite young into this category, between the ages of 57 and 43.

The between boxplots the groups for the Pressure Pulse variable are similar remarkably. Both have lower and upper outliers, with the majority vast of having blood pressures patients falling between 145 and 120 mm Hg. As shown (Figure 3.12), the mean pressure blood in both groups is approximately 130 mm Hg. Furthermore, for the variable cholesterol, the cholesterol appears distribution to be particularly among skewed to the right, those with disease heart, where a number significant of observations were recorded with values cholesterol of 0. As shown (Figure 3.12), the mean heart rate in those with disease is 150 beats per minute, while in those without disease heart it is 126 per minute beats.

In the case of variable ST-segment depression (OldPeak), there is a between variance the distributions of groups depression ST-segment. ST depression is variable more in patients with disease heart with larger outliers multiple. Most of these patients have between ST depressions 2 and 0 mm with 1.2 mm mean. In without disease heart patients, is narrower range, between 0 and 0.6 mm, with a ST depression median 0 mm, but the this group distribution is generally more skewed (Figure 3.12).

Consider the correlation matrix (Figure 3.13) associated with the data set on heart disease. Heart disease has the strongest positive relationship between OldPeak (0.4 correlation) and the relationship negative strongest with MaxHR (0.4correlation), according to the matrix correlation. MaxHR and Age have also a fairly relationship high with a 0.38correlation.

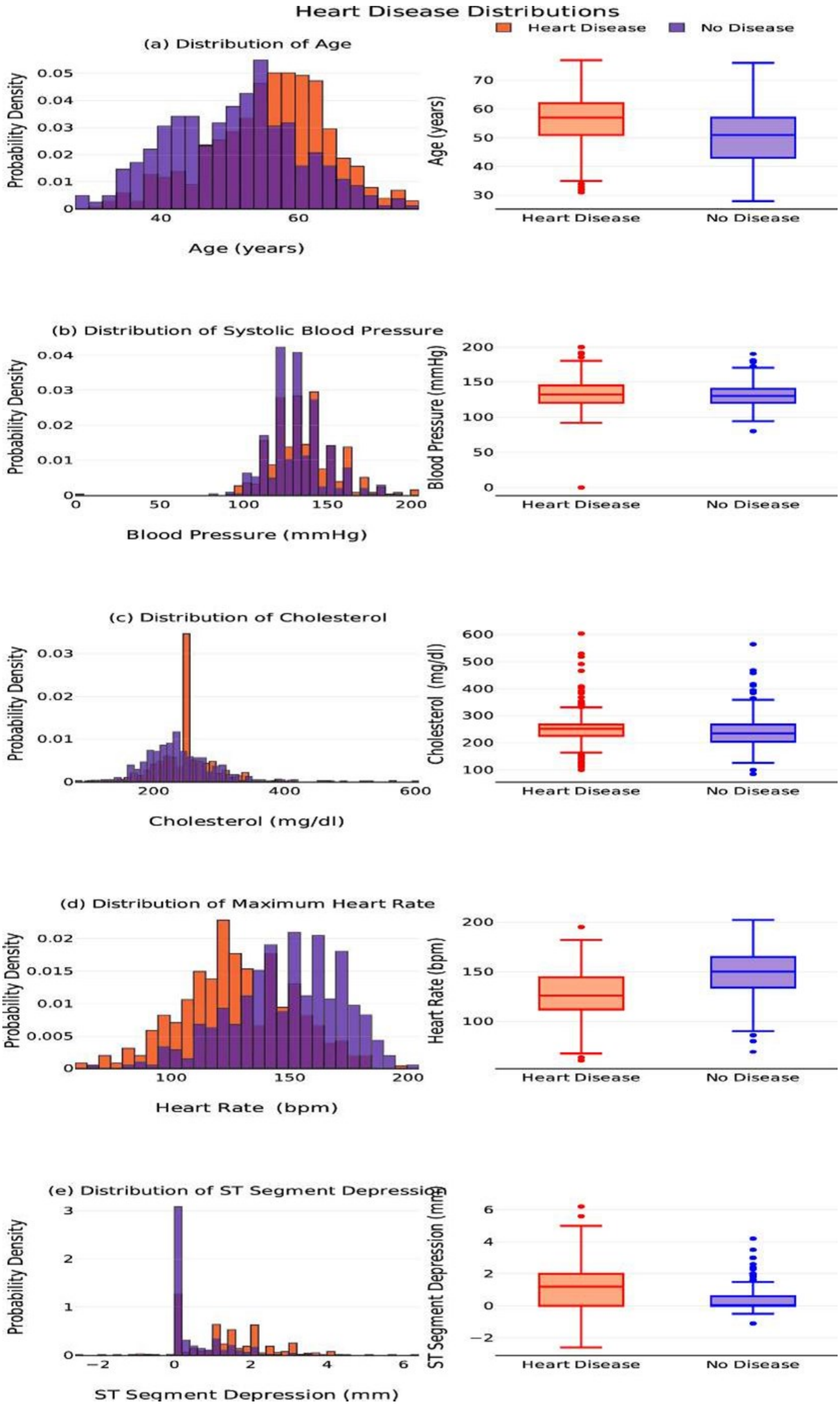

Figure 3.12 – Prevalence of heart disease

As can be seen (Figure 3.13), rate heart tends to with age decrease. The observe results a weak between correlation the numerical features and the variable target matrix based. OldPeak (a number associated with depression) is positively correlated with disease heart. Disease heart is correlated negatively with maximum heart rate. Cholesterol has an interesting negative relationship with heart disease.

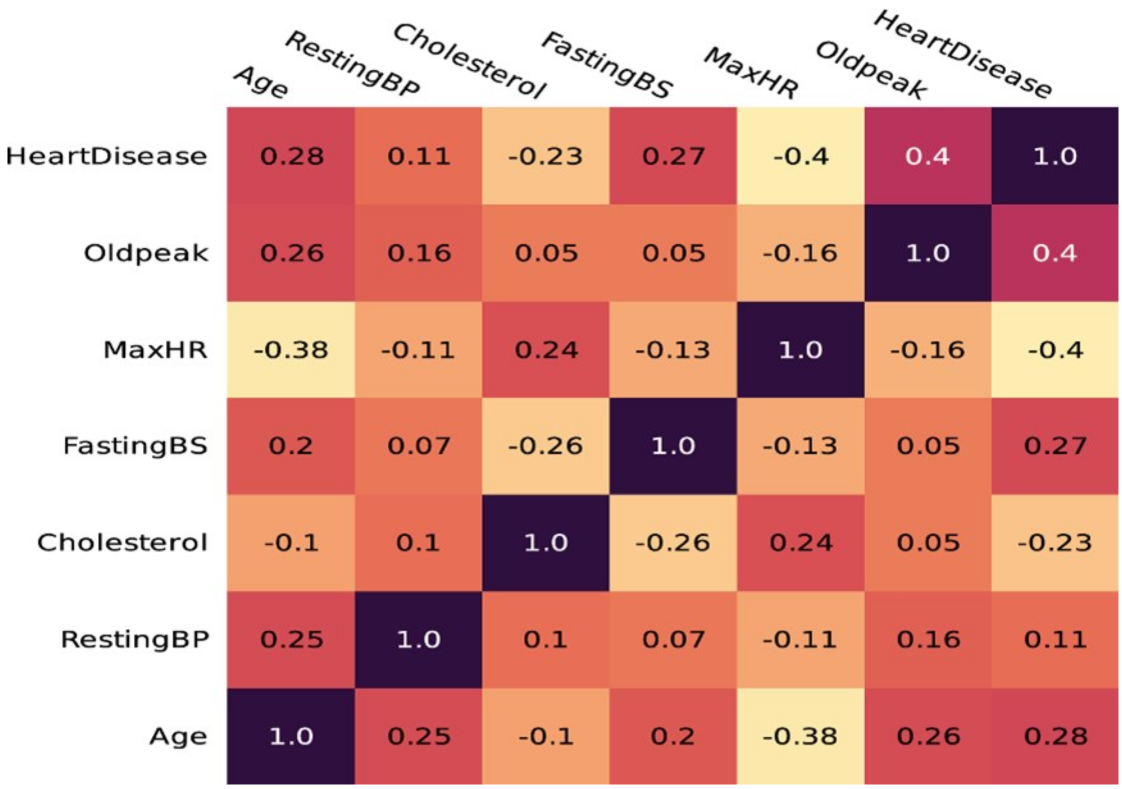

Figure 3.13 – Matrix correlation for the dataset Disease Heart

The between correlation disease heart and the variables is illustrated below categories (Figure 3.14). About 80% of diabetics have problems heart. Patients with induced exercise angina have greater an even disease cardiovascular incidence, 85%over. Over patients 65% with diagnosed heart disease had abnormalities wave ST-T on ECG resting, the highest percentage among the categories. Patients with flat or depressed ST slopes during exercise had the highest incidence of cardiovascular disease, 77.8%and 82.8%.

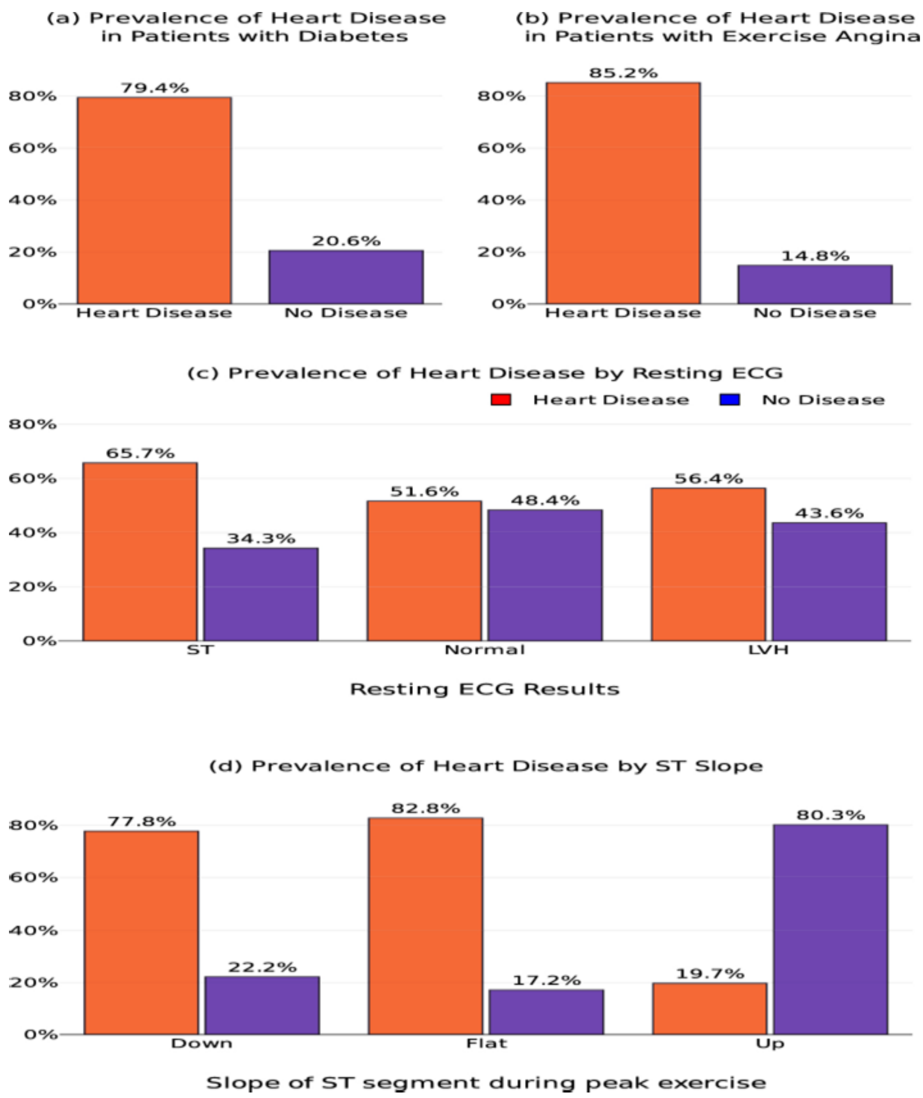

Figure 3.14 – Prevalence of heart disease based on ECG results.

Asymptomatic pain chest in disease heart in almost 77%, absence of pain chest is the most common symptom in patients with disease heart (Figure 3.15). In addition, disease heart is approximately times nine more common in men than women among patients with a diagnosis cardiovascular. A patient with chest asymptomatic pain (ASY) is approximately times six more likely to have disease heart than a patient with angina atypical (ATA).

Overall information obtained from the data analysis exploratory, the data for the variable target is balanced almost. The relationship between the numerical characteristics and the variable target is weak. Oldpeak (number associated with Depression) is positively correlated with heart disease.

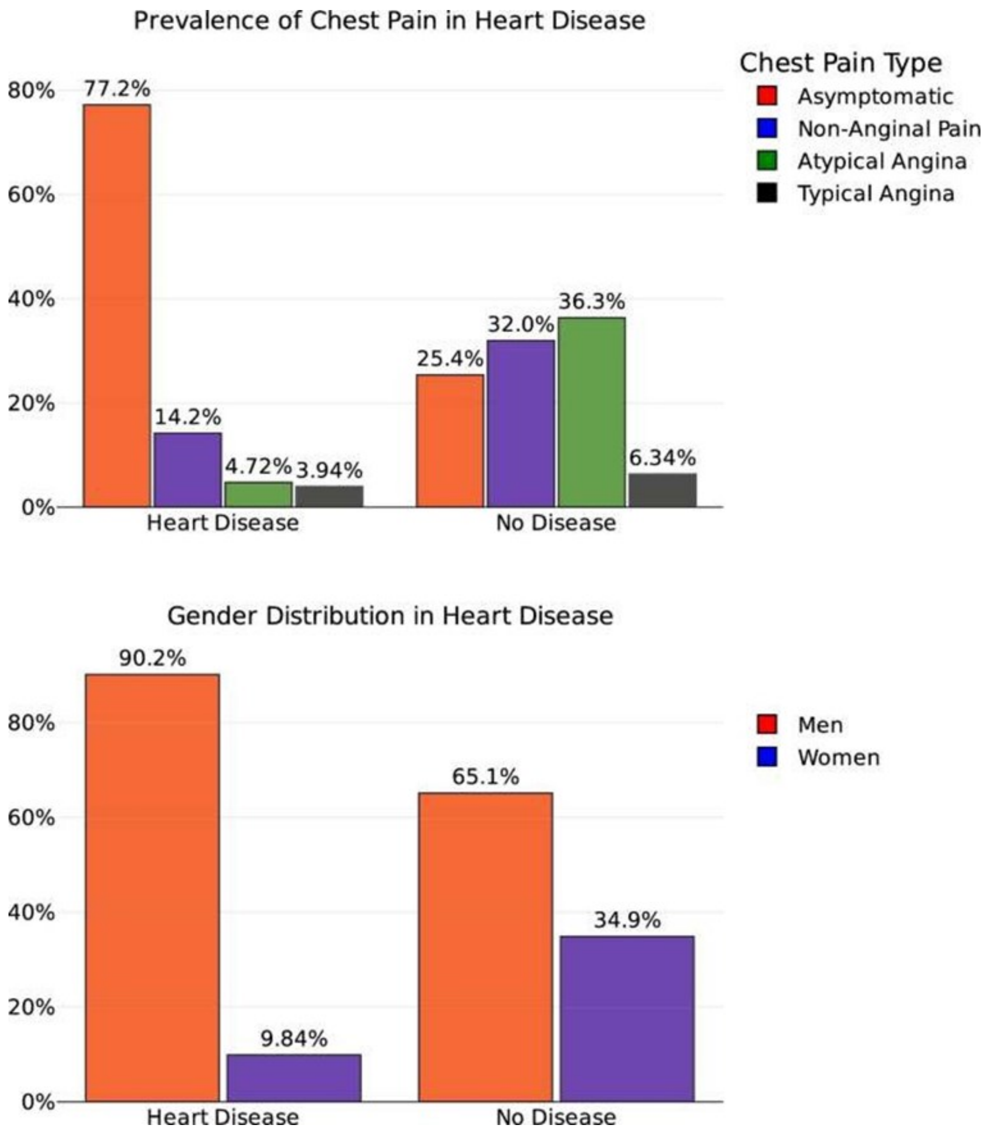

Figure 3.15 – Prevalence of chest pain among heart disease data

Heart disease is correlated negatively with maximum rate heart. Interestingly, there is a association negative between cholesterol and disease heart. Men are 2.44 times more approximately likely to have disease heart than women. There are clear differences between types of pain chest. Patients with pain chest asymptomatic (ASY) are times six approximately more likely to have disease heart than those with angina atypical (ATA). Resting ECG: Resting ECG comparable values are. Patients with abnormalities wave ST-T have a risk higher of developing disease heart than without those [16]. Angina exercise-induced: People who angina induced exercise have are almost 2.4 more times likely to have disease heart than people who do not. The slope of the segment ST at exertion maximum is altered. Slope Up ST has a lower risk significantly of disease cardiovascular than the two other segments. Angina exercise-induced with a score of "Yes" is almost more likely 2.4 times to lead to disease heart than angina exercise-induced with a score of "No."

3.4. Software

Machine learning PyTorch is a based platform on the library Torch, used for applications such as vision computer and language natural processing, developed originally by AI Meta and now part of umbrella Foundation Linux. It is source open and free software released under a license modified BSD. While the interface Python is polished more and the focus main of development, PyTorch include a interface C++.

A deep learning applications number have been built on PyTorch, including Tesla Autopilot, Uber's Pyro, Hugging Face's Transformers, PyTorch Lightning, and Catalyst.

PyTorch features two high-level provides: computing tensor (e. g. NumPy) with acceleration powerful using processing graphics units (GPUs) and neural deep networks built on a automatic tape-based system differentiation.

Meta (as Facebook known formerly) works with both PyTorch and the convolutional architecture for rapid embedding feature (Caffe2), but the defined models by the frameworks two were incompatible mutually. The “Open Neural Network Exchange (ONNX)” project was created by Microsoft and Meta in 2017 to convert models frameworks between. Caffe2 was merged with PyTorch in late March 2018. [10] In Autumn 2022, was announced that PyTorch would be managed by the Foundation PyTorch, a formed newly organization independent and a subsidiary of the Foundation Linux.

PyTorch defines a called class of Tensor (ect. torch.Tensor) for storing and manipulating uniform rectangular multidimensional numbers arrays.

Tensors PyTorch are similar to arrays NumPy, but can run also on NVIDIA GPUs that support CUDA. PyTorch is also support developing for platforms other GPU, such as AMD's ROCm and Apple's Metal Framework.

A tensor in physics is similar to a PyTorch tensor in that it is mostly a multidimensional array. The only additional feature of a physics tensor that is not present in a PyTorch tensor is that when indexing its entries, some indices are written below or above the index.

The tensor in physics supports four basic operations:

- addition;
- tensor product;
- abbreviation;
- change of basis.

Of these, only the basis change operation is affected by the co/contravariance distinction. The other three operations are not affected by co/contravariance and are therefore easy to implement for PyTorch tensors.

3.5. Auxiliary devices

Coronary heart disease, or ischemic heart disease (IHD), develops when the main vessels blood that your heart supply become diseased or damaged. Arteries coronary are the vessels blood that oxygen supply and to heart blood. Often, IHD can cause a attack heart. It is the most common form of disease heart in the USA, where it causes than more 370,000 per year deaths.

Over the past decade, continuous wireless monitoring health has been a application significant of new technology health. Overall, the idea of health continuous system monitoring and medical smart devices has huge business potential and human being well. This technology provides a device time real that early can alert people about their status health. Before main points explaining, it would be better to discuss sensors physiological some that can be used to systems build continuous monitoring health. These sensors can be used to people track in time real. In addition, they can be used for monitoring and diagnosis both. Continuous remote health monitoring uses various networks and protocols wireless, Bluetooth including, Zigbee, Wi-Fi and others (WAN, LAN). Physiological sensors, gateway (access connection) and cloud are the basic building blocks of health monitoring continuous (data storage). Rate heart shows the health heart our and helps to condition determine of system cardiovascular. In settings clinical, rate heart is monitored under conditions controlled, such as measurements blood, heart measurements tone and electrocardiogram (ECG). However, it also monitored can be at home. Our hearts beat to pump oxygen-rich blood to our tissues and remove cellular waste products from them [8]. The harder our hearts work to achieve these goals, the more we use our muscles, the faster our hearts will beat to blood pump more. A monitor rate heart is essentially a system wearable that takes our heart snapshot rate and measures the number of per minute beats (such bpm) so that we can monitor accurately our rate heart in detail. There are types two of heart methods monitoring: optical and electrical. The electrical approach has an average error rate of 1 percent and an average cost of $150.00. The optical method has an accuracy rating of 15 percent and an average cost of $20. The average resting heart rate for an adult male is approximately 70 beats per minute, and for an adult female, it is approximately 75 beats per minute. Rate heart greatly varies from one person to another on fitness depending, genetics and age. There are methods several for measuring rate heart, such as pressure blood waveform, ECG, phonocardiogram measurement, and measurement pulse, but methods such are expensive and clinical.

Over the past two decades, machine learning algorithms have become increasingly popular in various fields. In recent years, machine learning algorithms have been used in the healthcare industry to analyze large amounts of data to improve understanding of diseases. As a result, machine learning algorithms can be useful for discovering hidden patterns in medical datasets. Deep learning and machine learning have made great progress. In Masethe & Masethe (2014), an experiment was conducted to predict heart attacks and a comparison was made to find the best prediction method. This can be a useful tool for doctors to predict critical cases in their practice and provide appropriate advice. The prediction accuracy determined by the J48, REPTREE, and SIMPLE CART algorithms indicates that the parameters used are reliable indicators for predicting the presence of heart disease, but this study has the disadvantage that it cannot predict heart disease in real time. Chaurasia & Pal (2013) uses three classifiers: ID3, CART and DT to build the model, with CART being the most accurate with 83.49% and 0.23 s. The important features of heart disease are cp (chest pain), slope (slope of peak exercise segment), Exang (exercise-induced angina) and Restecg (resting electrocardiographic score). However, this study has the disadvantage that it cannot predict heart disease in real time. Sudhakar & Manimekalai (2014) describe various methods that have been used in recent years to calculate the prediction coefficient of heart disease. ANN, BN, decision trees and classification algorithms are some of the methods used, but this study has the disadvantage that it cannot predict heart disease in real time. In the study of diagnosis heart disease using arbitrage neural networks Olaniyi, Oyedotun, and Adnan (2015) used two methods: SVM and ANN. Support vector machine is the best algorithm for diagnosing heart disease with an accuracy rate of 87.5 percent and high sensitivity and specificity values, but the prediction of heart disease is static and not real-time. In Otoom et al. (2015), a pulse sensor is used for monitoring.

To build the model are used three classifiers: FT, BayesNet and SVM, with SVM being the accurate most with an 88.3% accuracy. Joshi confirm that three disease cardiovascular methods prediction such as Tree Decision, Nearest Neighbor and Naive Bayes were used to model develop using based classification methods mining data, with Tree Decision being the accurate most with 92.3% accuracy [16]. This work did not perform real-time prediction. Sultana, Haider & Uddin (2016) state that in the analysis of data mining methods for heart disease prediction, five classifiers are used: KStar, J48, SMO, Bayes Net and Multilayer Perceptron, with SMO showing the highest performance with an accuracy of 89%. This work did not perform real-time predictions. For patients with heart problems, sensors are used to provide continuous monitoring. For example, consider the work of Ali (2017). Using a heart rate sensor and a threshold algorithm to detect abnormalities, an Arduino-based heart rate monitoring and heart attack detection system was proposed. Due to the imperfect nature of the components, the system has limitations related to not providing accurate results of the monitoring system. Majumder, Mondal & Deen (2017) presents a state-of-the-art study on physiological parameters and activity monitoring systems developed on a wearable platform. Deep learning methods have attracted significant interest among medical researchers. A recent survey shows that about half of healthcare organizations plan to use deep learning (Kamilaris & Prenafeta-Boldú, 2018). Deep learning algorithms offer a great opportunity for early detection of heart disease. Thus, the use of deep neural networks in the analysis of a dataset allows for much faster and more reliable diagnosis of heart disease than with traditional methods. Methods called deep learning learn to represent objects from a low level (raw input) to a high and more abstract one (Makantasis et al., 2015). It consists of several layers that represent different levels of representation; hence the name deep learning (Gulli & Pal, 2017). Deep learning is defined as an automatic structured algorithm or hierarchical structure that emulates human learning in order to acquire certain knowledge. It stands out in that it does not require pre-programmed rules, but the system itself is able to independently learn to perform tasks during the pre-training phase. It is also characterized by the fact that it consists of interwoven artificial neural networks to process information. It is mainly used to automate predictive analysis. The algorithms that make up a deep learning system are located in different neural layers, which are made up of weights (numbers). The system is basically divided into three layers: the input layer (it consists of neurons that assimilate input data such as images or a database), the output layer (it is a network that performs information processing and performs intermediate calculations. The more neurons in this layer, the more complex the calculations are),

the output level (this is the last link in the chain, and it is the network that makes the decision or draws a conclusion by providing the output data).

How does deep learning work? Do computer programs that use deep learning go through the same process as a young child learning to identify an object? Each algorithm in the hierarchy applies a nonlinear transformation to its input data and uses the learning rate to build a statistical model as output. The iterations continue until the result reaches an acceptable level of accuracy. This is called deep learning because all data must pass through all hidden layers for the processing step. A typical and significant example of a deep learning-type method is multilayer perceptrons, which are basically a mathematical function that maps the input data to the output data. They are a composition (or network) of several simpler nonlinear functions (Karlik & Olgac, 2011), called perceptrons neurons or (Figure 4.1), where the associated weights with each composition are trained (usually by backpropagation) to learn a function that relates the input data to the outputs of a supervised problem learning [12].

IN framework SHDML used many methodologies, for example: wearable embedded system, Android, desktop, DL, ML.

In recent years, smartwatches have been launched by major companies electronics as well as startups new. One of the first offerings was the Galaxy Samsung Gear in Autumn 2013.

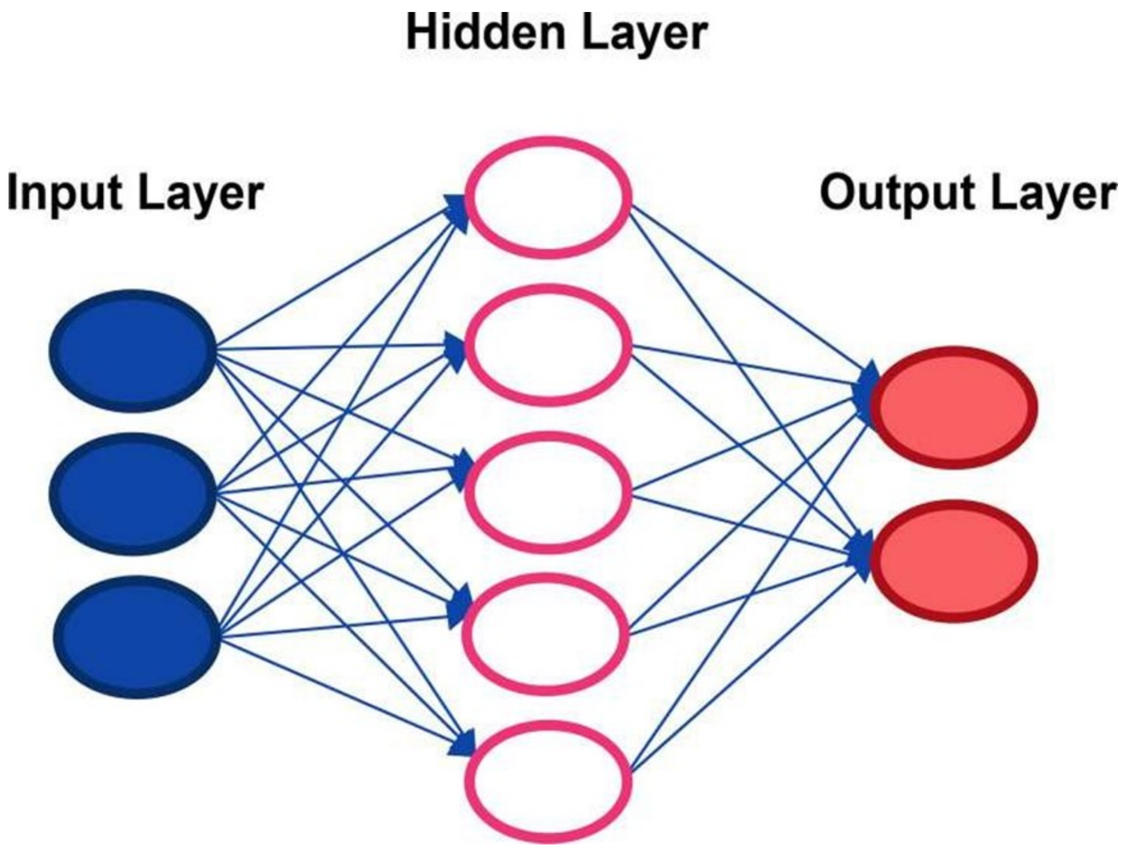

Figure 3.16 – Representation of a network neural classical

About a year later, in April 2015, the Apple Watch was released. Most proposed health monitoring interfaces use three architectures: “Wireless Body Area Network” (WBAN), which consists of devices wearable, as the data acquisition unit sensors, networking and communication, and a layer service. An embedded system has been proposed that includes a microcontroller connected to a heart rate sensor that measures the person's heart rate. The microcontroller transmits the collected health data to an Android application via Bluetooth wireless communication. The Android application stores the heart rate on a cloud server, and the user can access the measured heart rate; steps are counted in the Android application. The user confirms whether he or she suffers from heart coronary disease or not using a deep learning module via an Android/desktop app.

Built-in function used for creation portablecomputer. An system embedded is a "computer system" that consists of combinations microprocessors, memory and peripheral devicesspecial purpose I/O with embedded modern systems based primarily on an integrated microcontroller.

The wearable system is implemented based on an atmega32 microcontroller with a heart rate sensor and other modules, the heart rate sensor is based on technology PPG, sensors photoplethysmography use technology light-based to measure flow blood velocity, which is controlled by the pumping action of the heart. A Bluetooth module was added to interact with an Android application to send BPM beats per minute and an LCD monitor to view BPM to the wearer. The proposed structure is implemented by developing

"pulse sensor library", which handles the interfaces of the analog-to-digital converter ADC and TIMER with microcontrollers (Figure 3.17). The LED emits light that falls directly on the vein. In the veins there will be more blood flow in them only when the is pumping heart (systolic), the veins will have blood little flow in during them (diastolic), so if we monitor blood flow, we can also monitor heart contractions. The light sensor (Figure 3.18), will capture more light as it is reflected by the flow blood. The value threshold “T”, which is defined as the point starting for beat each (25% or 50%) of the signal, the converter analog-to-digital (ADC) is the start of reading the analog signal at point T, which represents the received light blood flow. The signal is converted to a digital value, with the application of noise suppression (Figure 3.19) (a dicrotic notch that represents the closing of the aortic valve). Every 2 ms the ADC stops the microprocessor to read from the sensor, this operation using was mechanism interrupt.

The SHDML structure consists of stages two. The stage first of the SHDML is able to patient's monitor rate heart. The heart rate is detected by a photoplethysmography (PPG) sensor [9]. The signal is processed by microcontroller ATmega32 to determine the rate heart per minute. It then sends the heart rate, represented as beats per minute, to the application via a Bluetooth connection.

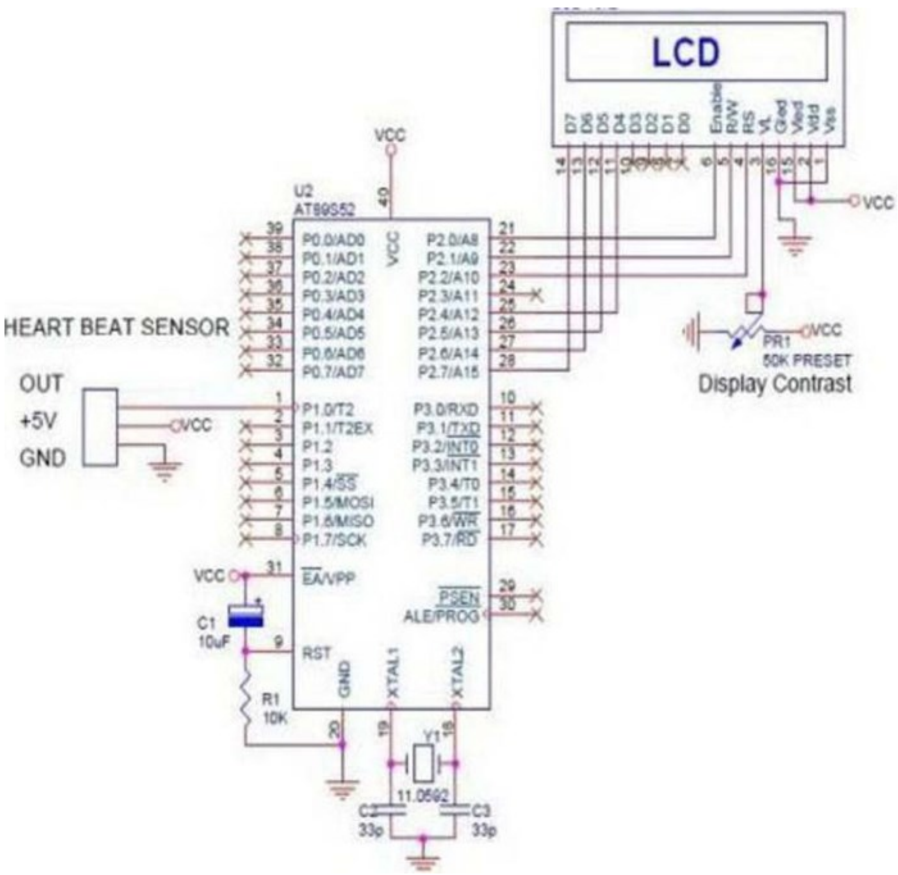

Figure 3.17 – Interfaces of analog-to-digital converter ADC and timer with microcontrollers

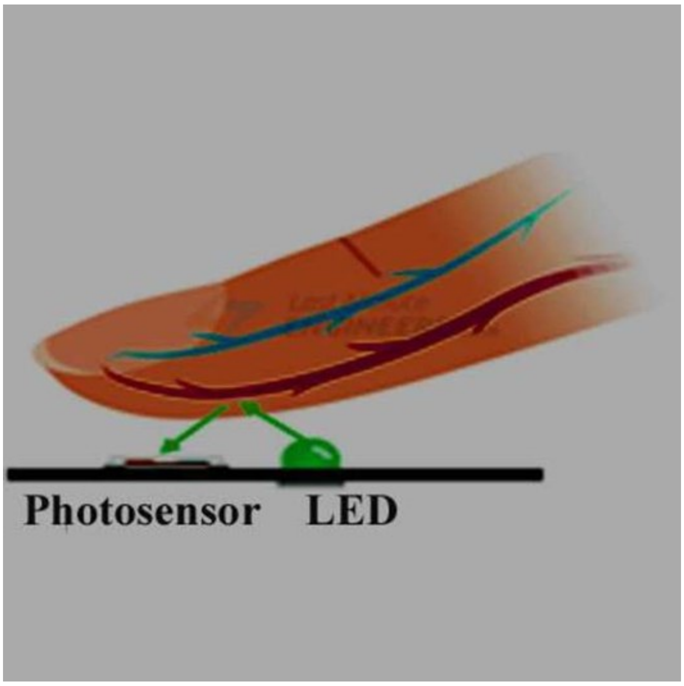

Figure 3.18 – Light sensor

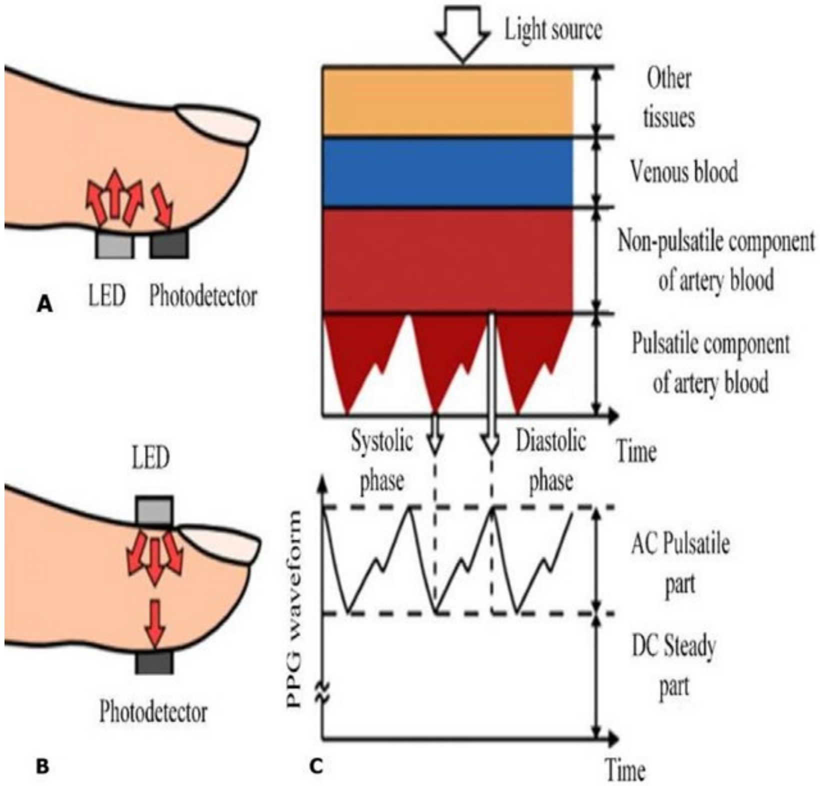

Figure 3.19 – Representation of a patient's electrocardiogram from training data

The stage second, SHDML, has been used in decision medical systems support for the diagnosis and of prediction disease heart. By deep training and techniques machine learning that analyze user data to detect disease heart (Figure 3.20).

Since our data on which the machine will be trained includes the desired outcome, the types of DL and ML used are supervised machine learning.

Input data: The data on which will be trained the model is collected and must be accurate, reliable, and balanced. Used Firebase Cloud real-time database for collecting and storing authentic data as opposed to a traditional relational database (Figure 4.7). Firebase: This is a real-time server-side database as a service (BaaS) that allows you to store a list of objects in a tree format [14]. Data can be synchronized across devices (Figure 3.21).

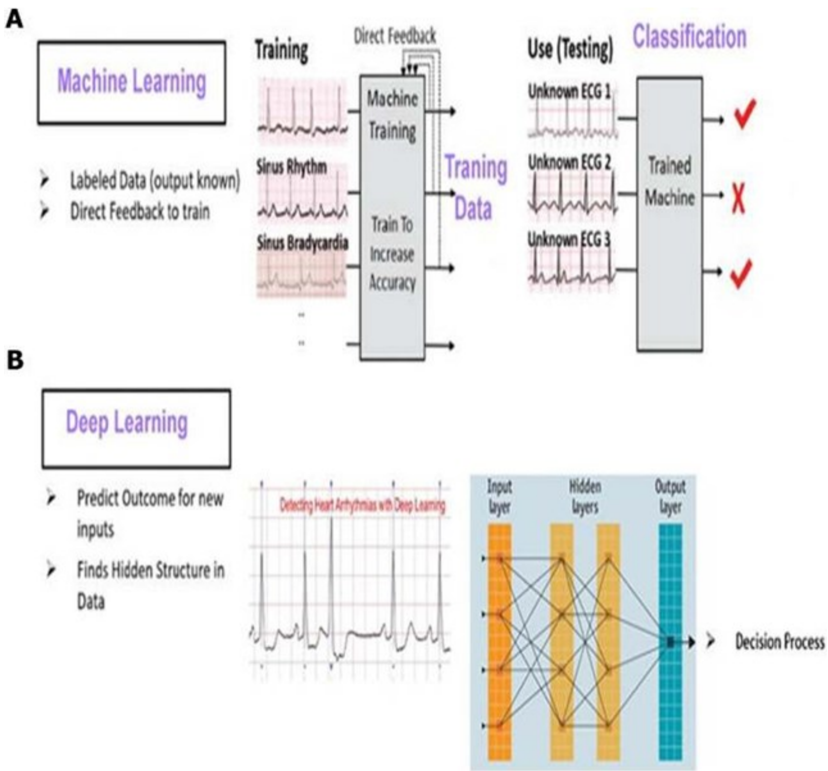

Figure 3.20 – The SHDML structure consists of ML and DL

3.6. Work scheme

Learning deep is a subfield of learning machine that is important in our project because it helps an embedded system think and an Android app predict that a user will have a attack heart in the next years ten. We used the Pytorch library because it is an easy-to-learn and easy-to-build library, and it can be used to deal with Android apps in a connected way, making it the best choice among libraries learning deep. We used CNN to help us improve our performance by extracting relationships and features, which increases the characteristics of our data by finding new relationships between all of them to get the pattern best and our data combination to results predict [13].

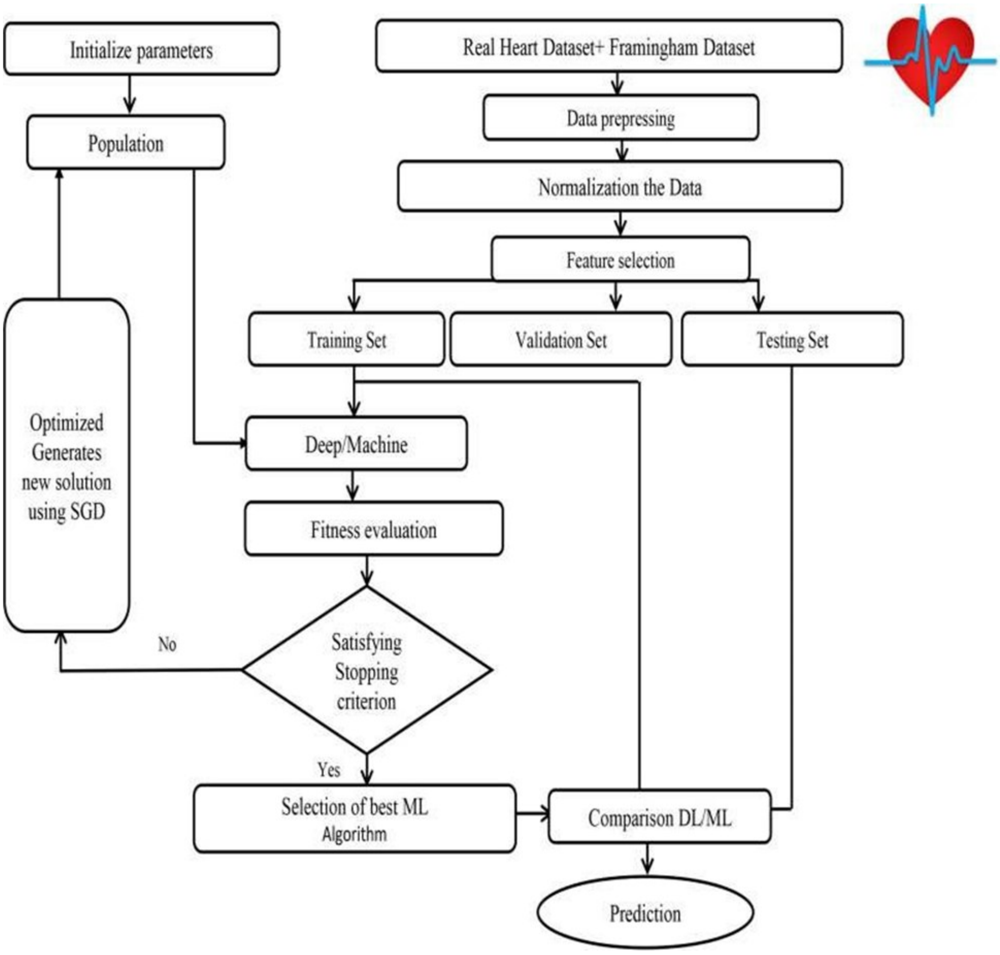

Figure 3.22 – SHDML infrastructure flowchart

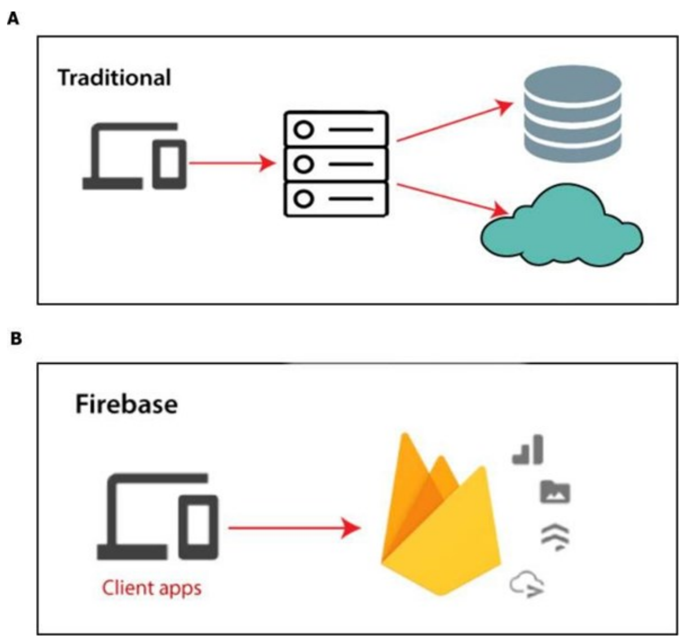

Figure 3.23 – Firebase real-time database

Biologically derived networks neural artificial, also known as perceptrons multilayer (MLPs), are capable of modeling very complex nonlinear functions. One of the important most machine learning techniques is artificial neural networks.

Neural networks (ANNs). These are brain-oriented systems designed to mimic human thinking, as the word implies “neural”. The output, hidden, and input layers of networks neural consist of layers three. The layer hidden typically blocks consists that transform data input into a pattern that the layer output can manipulate. ANNs are excellent tools for extracting and training a computer to identify patterns that are too complex or vague for a human programmer to recognize [16]. Neural networks have been used since the 1940s, and in recent decades they have become an important part of artificial intelligence thanks to the emergence of a new technique known as “backpropagation,” which allows networks to know how to modify their layers hidden of neurons in cases where the results do meet not the expectations creator’s.

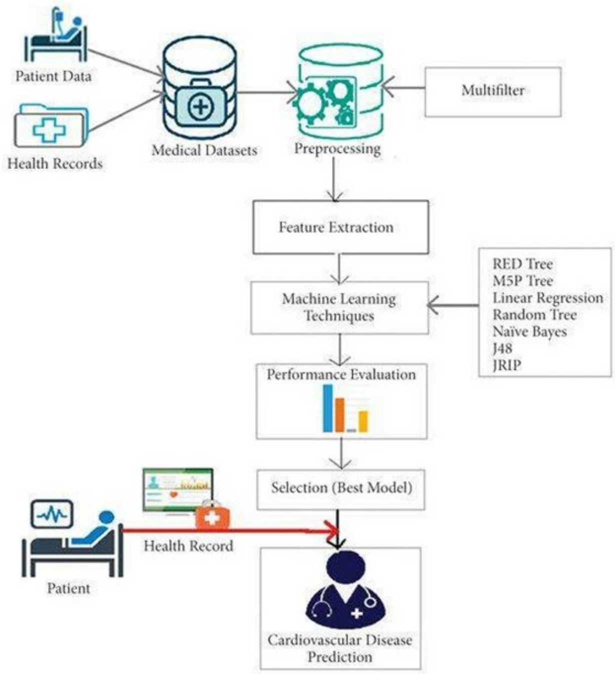

Figure 3.24 – Framework of the proposed cardiovascular disease prediction system

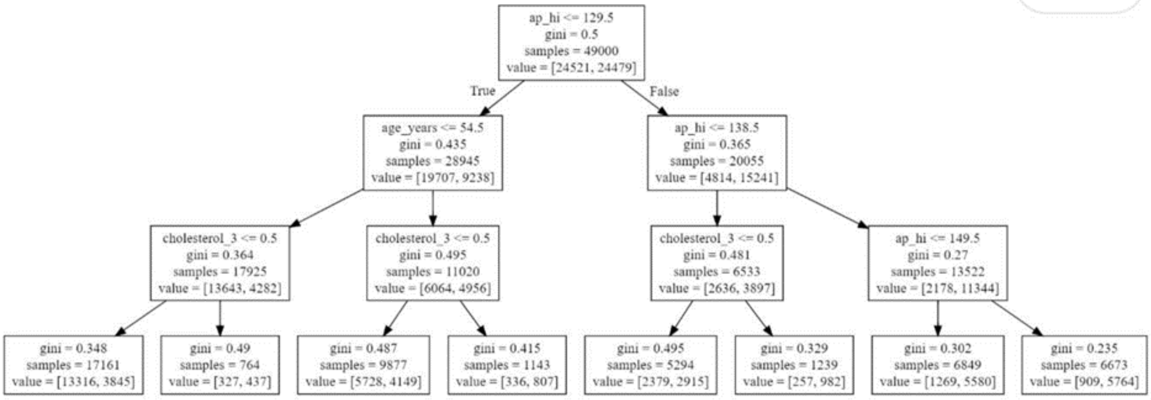

Figure 3.25 – Results of patient data collection and their visualization

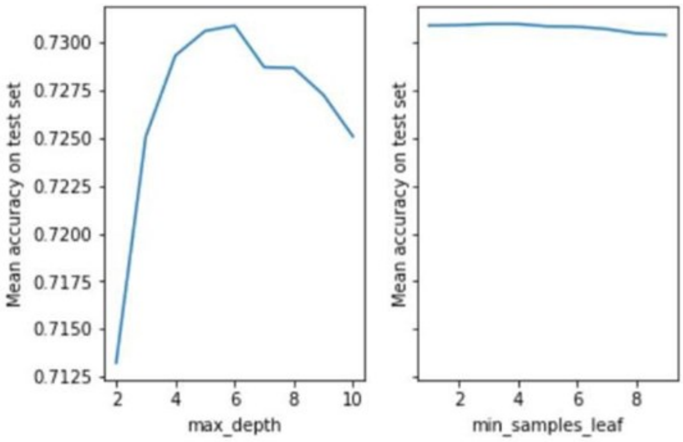

Figure 3.26 – data processing results in the form of a graph of the dependence of the parameters max_depth and min_samples_leaf

The outputs from previous states are fed as inputs to the current state in recurrent neural networks (RNNs). The hidden layers of RNNs have the ability to remember details. The outputs generated in the previous state are used to update the hidden state. RNNs can be used for time series forecasting because they have long-short-term memory (Ali et al., 2021), which allows them to recover previous inputs.

4 SAFETY OF LIFE, BASIC LABOR PROTECTION

4.1. Effects of electromagnetic radiation on the human body

A large body of literature exists on the response of tissues to electromagnetic fields, primarily in the extremely-low-frequency (ELF) and microwave-frequency ranges. In general, the reported effects of radiofrequency (RF) radiation on tissue and organ systems have been attributed to thermal interactions, although the existence of nonthermal effects at low field intensities is still a subject of active investigation. This chapter summarizes reported RF effects on major physiological systems and provides estimates of the threshold specific absorption rates (SARs) required to produce such effects. Organ and tissue responses to ELF fields and attempts to characterize field thresholds are also summarized. The relevance of these findings to the possible association of health effects with exposure to RF fields from GWEN antennas is assessed.

Nervous System

The effects of radiation on nervous tissues have been a subject of active investigation since changes in animal behavior and nerve electrical properties were first reported in the Soviet Union during the 1950s and 1960s.1 RF radiation is reported to affect isolated nerve preparations, the central nervous system, brain chemistry and histology, and the blood-brain barrier.

In studies with in vitro nerve preparations, changes have been observed in the firing rates of Aplysia neurons and in the refractory period of isolated frog

sciatic nerves exposed to 2.45-GHz microwaves at SAR values exceeding 5 W/kg.2,3,4 Those effects were very likely associated with heating of the nerve

preparations, in that much higher SAR values have not been found to produce changes in the electrical properties of isolated nerves when the temperature was controlled.[5, 6] Studies on isolated heart preparations have provided evidence of bradycardia as a result of exposure to RF radiation at nonthermal power densities,[7] although some of the reported effects might have been artifacts caused by currents induced in the recording electrodes or by nonphysiological conditions in the bathing medium.[8,9,10] Several groups of investigators have reported that nonthermal levels of RF fields can alter Ca^{2+} binding to the surfaces of nerve cells in isolated brain hemispheres and neuroblastoma cells cultured in vitro (reviewed by the World Health Organization[11] and in Chapters 3 and 7 of this report). That phenomenon, however, is observed only when the RF field is amplitude-modulated at extremely low frequencies, the maximum effect occurs at a modulation frequency of 16 Hz. A similar effect has recently been reported in isolated frog hearts.[12] The importance of changes in Ca^{2+} binding on the functional properties of nerve cells has not been established, and there is no clear evidence that the reported effect of low-intensity, amplitude-modulated RF fields poses a substantial health risk.

Results of in vivo studies of both pulsed and continuous-wave (CW) RF fields on brain electrical activity have indicated that transient effects can occur at SAR values exceeding 1 W/kg.[13,14] Evidence has been presented that cholinergic activity of brain tissue is influenced by RF fields at SAR values as low as 0.45 W/kg.[15] Exposure to nonthermal RF radiation has been reported to influence the electroencephalograms (EEGs) of cats when the field was amplitude-modulated at frequencies less than 25 Hz, which is the range of naturally occurring EEG frequencies.[16] The rate of Ca^{2+} exchange from cat brain tissue in vivo was observed to change in response to similar irradiation conditions.[17] Comparable effects on Ca^{2+} binding were not observed in rat cerebral tissue exposed to RF radiation,[18] although the fields used were pulsed at EEG frequencies, rather than amplitude-modulated. As noted above, the physiological significance of small shifts in Ca^{2+} binding at nerve cell surfaces is unclear.

A wide variety of changes in brain chemistry and structure have been reported after exposure of animals to high-intensity RF fields.19 The changes include decreased concentrations of epinephrine, norepinephrine, dopamine, and 5-hydroxytryptamine; changes in axonal structure; a decreased number of Purkinje cells; and structural alterations in the hypothalamic region. Those effects have generally been associated with RF intensities that produced substantial local heating in the brain.

Extensive studies have been carried out to detect possible effects of RF radiation on the integrity of the blood-brain barrier.20,21 Although several reports have suggested that nonthermal RF radiation can influence the permeability of the blood-brain barrier, most of the experimental findings indicate that such effects result from local heating in the head in response to SAR values in excess of 2 W/kg. Changes in cerebral blood flow rate, rather than direct changes in permeability to tracer molecules, might also be incorrectly interpreted as changes in the properties of the blood-brain barrier.

Effects of pulsed and sinusoidal ELF fields on the electrical activity of the nervous system have also been studied extensively.22,23 In general, only high-intensity sinusoidal electric fields or rapidly pulsed magnetic fields induce sufficient current density in tissue (around 0.1-1.0 A/m2 or higher) to alter neuronal excitability and synaptic transmission or to produce neuromuscular stimulation. Somewhat lower thresholds have been observed for the induction of visual phosphenes (discussed in the next section) and for influencing the electrical activity of Aplysia pacemaker neurons when the frequency of the applied field matched the endogenous neuronal firing rate.24 Those effects, however, have been observed only with ELF frequencies and would not be expected to occur at the higher frequencies associated with GWEN transmitters. Recent studies with human volunteers exposed to 60-Hz electric and magn.

Electromagnetic radiation can be classified into two types: ionizing radiation and non-ionizing radiation, based on the capability of a single photon with more than 10 eV energy to ionize oxygen or break chemical bonds. Ultraviolet and higher frequencies, such as X-rays or gamma rays are ionizing, and these pose their own special hazards: see radiation and radiation poisoning. By far the most common health hazard of radiation is sunburn, which causes over one million new skin cancers annually.

4.2 Types of hazards

Electrical hazards

Very strong radiation can induce current capable of delivering an electric shock to persons or animals.[citation needed] It can also overload and destroy electrical equipment. The induction of currents by oscillating magnetic fields is also the way in which solar storms disrupt the operation of electrical and electronic systems, causing damage to and even the explosion of power distribution transformers, blackouts (as occurred in 1989), and interference with electromagnetic signals (e.g. radio, TV, and telephone signals).

Fire hazards

Extremely high power electromagnetic radiation can cause electric currents strong enough to create sparks (electrical arcs) when an induced voltage exceeds the breakdown voltage of the surrounding medium (e.g. air at 3.0 MV/m). These sparks can then ignite flammable materials or gases, possibly leading to an explosion.

This can be a particular hazard in the vicinity of explosives or pyrotechnics, since an electrical overload might ignite them. This risk is commonly referred to as Hazards of Electromagnetic Radiation to Ordnance (HERO) by the United States Navy (USN). United States Military Standard 464A (MIL-STD-464A) mandates assessment of HERO in a system, but USN document OD 30393 provides design principles and practices for controlling electromagnetic hazards to ordnance.

On the other hand, the risk related to fueling is known as Hazards of Electromagnetic Radiation to Fuel (HERF). NAVSEA OP 3565 Vol. 1 could be used to evaluate HERF, which states a maximum power density of 0.09 W/m^2 for frequencies under 225 MHz (i.e. 4.2 meters for a 40 W emitter)/

Biological hazards

The best understood biological effect of electromagnetic fields is to cause dielectric heating. For example, touching or standing around an antenna while a high-power transmitter is in operation can cause severe burns. These are exactly the kind of burns that would be caused inside a microwave oven.[citation needed]

This heating effect varies with the power and the frequency of the electromagnetic energy, as well as the distance to the source. A measure of the heating effect is the specific absorption rate or SAR, which has units of watts per kilogram (W/kg). The IEEE and many national governments have established safety limits for exposure to various frequencies of electromagnetic energy based on SAR, mainly based on ICNIRP Guidelines, which guard against thermal damage.

There are publications which support the existence of complex biological and neurological effects of weaker non-thermal electromagnetic fields , including weak ELF magnetic fields and modulated RF and microwave fields. Fundamental mechanisms of the interaction between biological material and electromagnetic fields at non-thermal levels are not fully understood.

Lighting.

Fluorescent lights.

Fluorescent light bulbs and tubes internally produce ultraviolet light. Normally this is converted to visible light by the phosphor film inside a protective coating. When the film is cracked by mishandling or faulty manufacturing then UV may escape at levels that could cause sunburn or even skin cancer.

LED lights.

High CRI LED lighting.

Blue light, emitting at wavelengths of 400–500 nanometers, suppresses the production of melatonin produced by the pineal gland. The effect is disruption of a human being's biological clock resulting in poor sleeping and rest periods.

EMR effects on the human body by frequency

Warning sign next to a transmitter with high field strengths

While the most acute exposures to harmful levels of electromagnetic radiation are immediately realized as burns, the health effects due to chronic or occupational exposure may not manifest effects for months or years.[citation needed]

Extremely-low frequency

High-power extremely-low-frequency RF with electric field levels in the low kV/m range are known to induce perceivable currents within the human body that create an annoying tingling sensation. These currents will typically flow to ground through a body contact surface such as the feet, or arc to ground where the body is well insulated.

Shortwave

Shortwave (1.6 to 30 MHz) diathermy heating of human tissue only heats tissues that are good electrical conductors, such as blood vessels and muscle.

Adipose tissue (fat) receives little heating by induction fields because an electrical current is not actually going through the tissues.

4.3 Road Transport Safety

The basic strategy of a Safe System approach is to ensure that in the event of a crash, the impact energies remain below the threshold likely to produce either death or serious injury. This threshold will vary from crash scenario to crash scenario, depending upon the level of protection offered to the road users involved. For example, the chances of survival for an unprotected pedestrian hit by a vehicle diminish rapidly at speeds greater than 30 km/h, whereas for a properly restrained motor vehicle occupant the critical impact speed is 50 km/h (for side impact crashes) and 70 km/h (for head-on crashes).

As sustainable solutions for all classes of road have not been identified, particularly low-traffic rural and remote roads, a hierarchy of control should be applied, similar to classifications used to improve occupational safety and health. At the highest level is sustainable prevention of serious injury and death crashes, with sustainable requiring all key result areas to be considered. At the second level is real time risk reduction, which involves providing users at severe risk with a specific warning to enable them to take mitigating action. The third level is about reducing the crash risk which involves applying the road design standards and guidelines (such as from AASHTO), improving driver behavior and enforcement.

4.4 Conclusions

A serious workplace injury or death changes lives forever for families, friends, communities, and coworkers too. Human loss and suffering is immeasurable. Occupational injuries and illnesses can provoke major crises for the families in which they occur. In addition to major financial burdens, they can impose substantial time demands on uninjured family members. Today, when many families are operating with very little free time, family resources may be stretched to the breaking point. Every person who leaves for work in the morning should expect to return home at night in good health. Can you imagine the knock on the door to tell you your loved one will never be returning home? Or the phone call to say he's in the hospital and may never walk again? Ensuring that husbands return to their wives, wives to their husbands, parents to their children, and friends to their friends that is the most important reason to create a safe and healthy work environment. But it isn't the only reason.

CONCLUSIONS

Heart disease is serious and causes many deaths every year. If a patient ignores the warning signs of heart disease, dire consequences can occur quickly. The main objective was to outline several data analysis methods that can be effectively applied to predict heart disease. That is, to make predictions that are accurate and efficient while using fewer features and tests. This study uses various preprocessing methods and machine learning algorithms to conduct in-depth analysis and obtain results. A variety of supervised learning classifiers were used to classify the diagnosis of heart problems, while Success in heart disease diagnosis classifications should be evident if the features are treated properly. If the features are stored properly, the categorization of cardiovascular diseases can work properly.

This work can be extended to diagnose heart disease using a machine learning technique that is understandable to establish the accuracy, fairness, transparency, and performance of the model. Additionally, using a self-generated dataset will allow this research to be expanded.

REFERNCES

1. Mancia, S. Oparil, PK Whelton, M. McKee, A. Dominiczak, FC Lu ft, K. AlHabib, F. Lanas, A. Damasceno, D. Prabhakaran, G. La Torre The technical report on sodium intake and cardiovascular disease in low-and middle-income countries by the joint working group of the World Heart Federation, the European Society of Hypertension and the European Public Health Association Eur Heart J, 38 [Text], (10) (2017). 712-719 p.

2. AND. Coronary Heart disease [Text], (2020), 468 p.

3. AND. D'Souza To predict cardiac disease, data mining techniques are applied 74-77 Int. J.Res.Eng.Sci., 3 (3) [Text], (2015) 75 p.

4. Sanjiv, JS (2015), [Phenomapping for Novel Classification of Heart Failure With Preserved Ejection Fraction], Circulation, #131, [Text], pp. 269–279.

5. Choi E. (2017) [Using recurrent neural network models for early detection of heart failure onset], Journal of the American Medical Informatics Association, No. 24, [Text], pp. 361–370.

6. Koelio LP, Richard V. (2016) [Construction of machine learning systems in Python], DMK Publishing House, [Text], 303 p.

7. WJ Loesche Periodontal disease as a risk factor for heart disease Compendium, 15 (8) [Text], (1994), 976-978 p.

8. SK Sen Using machine learning methods in heart disease detection and prediction Int. J.Eng.Comput. Sci., 6 (6) [Text], (2017), pp. 21623-21631.

9. R. Hertel, R. Benlamri A deep learning segmentation-classification pipeline for x-ray-based covid-19 diagnosis” Biomedical Engineering Advances (2022), Article 100041.

10. Chattopadhyay, M. Maitra MRI-Based brain tumor image detection using CNN based deep learning method Neurosci.Inf. (2022), Article 100060.

11. SK Mamatha, HK Krishnappa, N. Shalini Graph theory based segmentation of magnetic resonance images for brain tumor detection Pattern Recogn Image Anal, 32 (1) [Text], (2022), pp. 153-161.

12. GN Ahmad, H. Fatima, AS Saidi Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV IEEE Access [Text], (2022), 576-596 p.

13. MM Rahman, MR Rana, Nur-A-Alam, MSI Khan, KMM Uddin A web-based heart disease prediction system using machine learning algorithms Netw. Biol., 12 (2) [Text], (2022), pp. 64-81.

14. SK Dey, MM Rahman, A. Howlader, UR Siddiqi, KMM Uddin, R. Borhan, EU Rahman Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: a machine learning approach PLoS One, 17 (7) (2022), Article e0270933.

15. DVB Oliveira, JF da Silva, TA de Sousa Araújo, UP Albuquerque Influence of religiosity and spirituality on the adoption of behaviors of epidemiological relevance in emerging and re-emerging diseases: the case of dengue fever J Relig Health, 61 (1) [Text], (2022), pp. 564-585.

16. Vrakina K. P. Proceedings of XII International Scientific and Practical Conference. 2022. C. 161-164. URL: EURASIAN-SCIENTIFIC-DISCUSSIONS-18-20.12.2022.pdf (sci-conf.com.ua).

17. Y. Leshchyshyn, L. Scherbak, O. Nazarevych, V. Gotovych, P. Tymkiv and G. Shymchuk, «Multicomponent Model of the Heart Rate Variability Change-point,» 2019 IEEE XVth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH), Polyana, Ukraine, 2019, pp. 110-113, doi: 10.1109/MEMSTECH.2019.8817379.

18. Lytvynenko, S. Lupenko, O. Nazarevych, G. Shymchuk and V. Hotovych, «Mathematical model of gas consumption process in the form of cyclic random process,» 2021 IEEE 16th International Conference on Computer Sciences

and Information Technologies (CSIT), LVIV, Ukraine, 2021, pp. 232-235, doi: 10.1109/CSIT52700.2021.9648621.

19. V. Kozlovskyi, Y. Balanyuk, H. Martyniuk, O. Nazarevych, L. Scherbak and G. Shymchuk, «Information Technology for Estimating City Gas Consumption During the Year,» 2022 International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 2022, pp. 1-4, doi: 10.1109/SIST54437.2022.9945786.

20. I. Lytvynenko, S. Lupenko, N. Kunanets, O. Nazarevych, G. Shymchuk and V. Hotovych, "Simulation of gas consumption process based on the mathematical model in the form of cyclic random process considering the scale factors", 1st International Workshop on Information Technologies: Theoretical and Applied Problems ITTAP 2021, 16–18 November 2021.

21. Approach to gas consumption process forecasting on the basis of a mathematical model in the form of a random cyclic process / Serhii Lupenko, Iaroslav Lytvynenko, Oleg Nazarevych, Grigorii Shymchuk, Volodymyr Hotovych // ICAAEIT 2021, 15-17 December 2021. – Tern. : TNTU, Zhytomyr «Publishing house „Book-Druk“» LLC, 2021. – P. 213–219. – (Mathematical modeling in power engineering and information technologies).

22. Конспект лекцій з дисципліни «Грід-системи та технології хмарних обчислень» для студентів освітніх рівнів «бакалавр», «магістр» / Укладачі : Шимчук Г.В., Маєвський О.В., Назаревич О.Б., Стадник М.А. – Тернопіль : Тернопільський національний технічний університет імені Івана Пулюя, 2016 – 340 с.