# AN EFFICIENT FRAMEWORK FOR CARDIOVASCULAR PROBLEMS USING CLOUD OF THINGS

Thesis Submitted for the Award of the Degree of

**DOCTOR OF PHILOSOPHY**

**in**

**Computer Applications**

**By**

**Shilpa**

**Registration Number: 41900255**

**Supervised By**

**Dr. Tarandeep Kaur (24836)**

**School of Computer Applications (Assistant Professor)**

**Lovely Professional University, Punjab**

Transforming Education Transforming India

**LOVELY PROFESSIONAL UNIVERSITY, PUNJAB**

**2025**

# DECLARATION

I, hereby declared that the presented work in the thesis entitled "**AN EFFICIENT FRAMEWORK FOR CARDIOVASCULAR PROBLEMS USING CLOUD OF THINGS**" in fulfillment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision of Dr. Tarandeep Kaur, working as Assistant Professor, in the School of Computer Applications of Lovely Professional University, Punjab, India. In keeping with the general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigators. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

Shilpa

**(Signature of Scholar)**

Name of the scholar: SHILPA

Registration No.:41900255

Department/school: School of Computer Applications

Lovely Professional University, Punjab, India

# CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled "**AN EFFICIENT FRAMEWORK FOR CARDIOVASCULAR PROBLEMS USING CLOUD OF THINGS**" submitted in fulfillment of the requirement for the award of degree of **Doctor of Philosophy (Ph.D.)** in the School of Applications, is a research work carried out by Shilpa, 41900255, is bonafide record of his/her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

Tarandeep Kaur

**(Signature of Supervisor)**

**Name of supervisor: Dr. Tarandeep Kaur**

**Designation: Assistant Professor**

**Department/school: School of Computer Applications**

**University: Lovely Professional University**

# ABSTRACT

The healthcare industry has become a leading sector with a prompt need for efficient services for the patient such as patient-centric related services delivery, management of the health records, and facilitation of efficient, fastest, and reliable healthcare systems. Recently, the medical industry has been transformed, and there has been an extreme convergence towards digitalized healthcare offerings owing to the advance, engrossment, incorporation, and dependency on the digital advancements that include the Information and Communication Technology (ICT) sector, known as the digital healthcare. The health industry is pitching for more upgradation, an energetic environment, professional enhancements, and sprightly to deliver and support the purveying technologies-enabled healthcare offerings. Digital healthcare innovations facilitate prompt healthcare services regardless of geographical boundaries and time. In recent years, cardiovascular illnesses have become the main cause of mortality and disability across the globe. Cardiovascular diseases comprise an inclusive variety of ailments affecting the heart and blood arteries, including hypertension, Coronary Artery Disease (CAD), heart failure, and strokes.

Many novel and advanced technologies have been integrated to bring the digital revolution in various sectors including healthcare. The healthcare sector is transforming into digital healthcare with the involvement of digital technologies. In the healthcare sector, a drastic change has come in the last few years due to the involvement of Artificial Intelligence(AI), blockchain, Cloud of Things (CoT), cloud computing, robotics, Internet of Things (IoT), and more advanced innovative technologies. Cloud computing and IoT are growing technologies at the global level. Their adoption and use are expected to be more and more pervasive, making them important components of digital services.

Recently, Machine learning (ML) has considered as a transformative technology being utilized in every sector such as healthcare, education, and many other sectors. In the healthcare domain, ML offers tools and digital technologies for enhancements in diagnosis, predictions, patient care, and treatments. Machine learning offers significant innovations for data analysis without human interaction. ML is becoming

more crucial for healthcare professionals to predict meaningful and relevant results from patient health information. In the process of evaluating complicated and diverse medical data, ML algorithms have shown a great deal of promise for finding significant patterns and making correct predictions. The development of such skills has opened the way for the creation of intelligent CVD analysis models, which aid medical practitioners in making informed choices, boosting diagnosis accuracy, and optimizing patient care. Machine learning algorithms are extensively used for decision-making as well as prediction systems. Various ML approaches are used to improve performance such as classification, prediction, and function approximation related to machine learning. In recent times, there is a growing trend for combining machine learning approaches as an ensemble for solving various issues. Considering other machine learning methods, ensemble learning has shown particularly promising outcomes. This approach combines the predictions produced by numerous models to enhance both performance and resilience. It is a part of data science and is used for data analysis based on features to predict the outcomes and make decisions to achieve the objectives. Machine learning has emerged as an advanced approach in healthcare offerings that is being used to detect the cardiovascular diseases (CVD).

The developments of machine learning (ML) and AI approaches in the healthcare sector have transformed the area of medical diagnostics and presented new prospects for early identification of CVD and precise prediction of its progression. AI in healthcare brings the potential of health diagnosis and treatment of patients. AI helps with image analysis of medical as well as health monitoring.

There has been research work carried out for CVD prediction and classification using digital approaches such as ML, and several models have been proposed that focus on CVD prediction and diagnosis However, such models lack subtle recommendations and suggestions for treatments.

Driven by the need for an effective CVD identification and classification, the thesis proposes a PCard model, that is used to predict the prospective occurrence and the probability of the presence and absence of cardiovascular disease using different machine learning algorithms. It uses k-nearest Neighbors, Support Vector Machine, Logistic Regression, Multi-Layer Perceptron, and Deep Forest for the CVD prediction. A recommendation component has been integrated with a prediction system in PCard, where the patients are provided with valuable recommendations for their diagnosis based on the level of CVD predictions. The proposed PCard model has been tested and trained using the CVD data cardio_train taken from the Kaggle repository, and the experimental results indicate that the model achieves an accuracy of 98.5% compared to individual machine-learning approaches.

The PCard model uses ensemble learning that capitalizes on the complementary capabilities of many machine learning algorithms. It tends to improve the diagnostic accuracy and reliability by utilizing an ensemble approach The ensemble approach used in PCard enhances the model's capacity to deal with the complexity and variability that are inherent in CVD datasets and samples. This model allows us to analyze different features and diseases that are related to cardiovascular disease diagnosis.

The proposed model uses ML approaches for the prediction of CVD problems. It is an ensemble machine learning model used to detect the absence and presence of CVD and predict the associated risk levels. The proposed model not only helps to diagnose CVD but also facilitates the level-based segregation for CVD risks, such as high risk, low risk, or medium risk.

Additionally, the proposed model is also centered on the development of an online CVD analysis model for better CVD diagnostics as well as the potential of ML algorithms in real-time. The proposed model is based on ML algorithms such as k-nearest Neighbors (KNN), Logistic regression (LR), random forest (RF), Multi-Layer Perceptron (MLP), and Support vector machine (SVM). The purpose of using these algorithms is for the ensemble of classifiers to utilize the capabilities of each method and produce a diagnostic mechanism that is more reliable and accurate. The model involves data preprocessing, feature selection, and an ensemble of various ML algorithms for the CVD prediction.

The research work has been conducted on the cardio_train dataset available in the Kaggle repository having 70,000 patient records and 11 attributes. In the first stage, the model is trained using an ensemble approach using different ML algorithms, and second stage, the results are compared with the existing ensemble-based proposed model, such as the MLP model proposed in 2022 with an accuracy of 82.47% and 87.5%, the EHDPS model achieved an accuracy. In contrast, the PCard model achieved an accuracy of 98.5%.

# ACKNOWLEDGEMENT

I am deeply grateful to my research advisor Dr. Tarandeep Kaur for her motivating guidance throughout this research work. She provided me with great freedom to choose a research topic of my interest and directed and critiqued me just in time. Without her inspiration, untiring efforts, enthusiastic attitude, and personal care, the accomplishment of this dissertation would have been impossible. I benefited enormously from her valuable feedback, logical analysis, and comments. Her input was vital in shaping this work. She has been a great source of inspiration to me, and I would like to thank her from the core of my heart.

I would like to gratefully acknowledge the School of Computer Applications, Lovely Professional University (Punjab), India, for offering this doctorate program. The University and the department were always very cooperative in extending all the resources and infrastructure facilities during my research span. I am thankful to all the reviewers of my research manuscripts for providing valuable feedback and comments that helped me raise the standard of this research work. I also extend my special thanks to all the cited authors in my thesis, whose research findings provide me with a battleground to choose from in the research directions.

Words are inadequate to express my gratitude to my loving Husband, Mr. Piyush Malhotra, for always being there for me. He never failed to give me great encouragement and confidence to finish my thesis. I am highly grateful to him for all the cooperation, affection, motivation, understanding, patience, and sincere advice. His support and care helped me overcome setbacks and stay focused on my thesis. I greatly value her friendship and sincerely appreciate her belief in me.

I am also grateful to my Mother, Mrs. Manjit, for appreciating my research and support in every hard situation at day and night and my friends for their best wishes and words of encouragement. They cheer me up during the walk through the journey towards attaining the Doctorate Degree. Also, I express my thanks to my other motivators for always being with me at any stage during this work.

Above all, I firmly believe that everything in the world is happening as per the plan and will of the Almighty. So, I thank him for planning this and allowing me to do this thesis. I pray that he continues showering his grace and blessings onto me and all others.

Place: LPU, Phagwara (SHILPA)

Date: (Reg No: 41900255)

# LIST OF ACRONYMS

| Abbreviations | Description |
|---|---|
| CVD | Cardiovascular Disease |
| CAD | Coronary Artery Disease |
| ML | Machine Learning |
| AI | Artificial Intelligence |
| CNN | Neural Network |
| NB | Naive Bayes |
| k-NN | k-Nearest Neighbors |
| MLP | Multilayer Perceptron |
| SVM | Support Vector Machine |
| ECG | Electrocardiogram |
| SL | Supervised Learning |
| USL | Unsupervised Learning |
| RFL | Reinforcement Learning |
| WHO | World Health Organization |
| PAD | Peripheral Artery Disease |
| CHD | Coronary Heart Disease |
| TIA | Transient Ischemic Attack |
| MRI | Magnetic Resonance Imaging |
| IoT | Internet of Things |
| CoT | Cloud of Things |
| HER | Electronic Health Records |
| HIE | Health Information Exchange |
| PHR | Patient Health Record |
| HaaS | Healthcare-as-a-Service |
| SaaS | Storage-as-a-Service |
| ICT | Information and Communication Technology |
| VR | Virtual Reality |
| AR | Augmented Reality |
| GBM | Gradient Boosting Machines |
| LDL | Low-Density Lipoprotein |
| HDL | High-Density Lipoprotein |
| BMI | Body Mass Index |
| BNP | Brain Natriuretic Peptide |
| ASCVD | Atherosclerotic Cardiovascular Disease |
| EMR | Electronic Medical Record |
| RF | Random Forest |
| HD | Heart Disease |

| | |
|---|---|
| ANN | Artificial Neural Network |
| NLP | Natural Language Processing |
| LR | Logistic Regression |
| Pt | Patient |
| RFj | Patient Risk Attribute |
| Levij | Level of CVD associated with each patient |
| RNN | Recurrent Neural Network |
| GNN | Graph Neural Network |
| CK-MB | Creatine Kinase-Myocardial Band |
| AIHW | Australian Institute of Health and Welfare |
| AHA | American Heart Association |
| CDC | Centers for Disease Control and Prevention |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLE

# CHAPTER - 1

# INTRODUCTION

The healthcare industry has become a leading sector with a prompt need for efficient services for the patient such as patient-centric related services delivery, management of the health records, and facilitation of efficient, consistent, and fastest medical sector. Recently, the medical industry has been transformed, and there has been an extreme convergence towards digitalized healthcare offerings owing to the advancements, engrossment, incorporation, and dependency on the advancements such as the Information and Communication Technology (ICT) sector, known as digital healthcare. Consequently, the healthcare industry is pitching for more upgradation, an energetic environment, professional enhancements, and sprightly to deliver and support the purveying technologies-enabled healthcare offerings. Digital healthcare innovations facilitate prompt healthcare facilities regardless of geographical borders and time. This chapter serves as a technological enhancement in the healthcare sector.

Technological enhancements in the medical services and sector help monitor, identify, and treat human diseases. It influences our daily routine and impacts our future in providing health-related services such as telemedicine, mHealth, artificial intelligence, electronic health records, sensors, and other devices that impact digital transformation. Digital transformation has already shaped the healthcare industry with technology enhancement and more transformations will impact the industry with more advancements in virtual reality and augmented reality.

This chapter has been divided into 9 sections, where Section 1.1 enlightens on cardiovascular disease and its ailments. Section 1.2 gives a detailed description of the Cloud of Things. Section 1.3 highlights the significance of digital healthcare with its technologies, while Section 1.4 defines the basics of Machine Learning, ML-based models, and healthcare analytics. Section 1.5 discusses the problem statement guiding the requirement of the proposed work. Section 1.6 lists the objectives for the proposed work.

Section 1.7 pens down the research questions related to the research work. Sections 1.8 and 1.9 cover the thesis contributions and thesis organization, respectively.

## 1.1 Introduction

In recent years, cardiovascular illnesses, also known as Cardiovascular diseases (CVD), have become the main reason for mortality and disability across the globe. Cardiovascular illnesses comprise a broad variety of ailments that affect the heart and blood arteries, such as hypertension, Coronary Artery Disease (CAD), heart failure, and strokes [1].

As per the report of the World Health Organization (WHO), CVDs have become the cause of around 17.9 million deaths annually, which accounts for 31% of all fatalities that occur throughout the globe [2]. The cardiovascular system, including the heart and its blood vessels, can be affected by various issues that can hinder its proper functioning, which are also known as CVD. These are also called heart diseases, which include different entities such as CAD, Peripheral Artery Disease (PAD), aortic atherosclerosis, and Cerebrovascular Disease [3].

CAD, also known as coronary heart disease (CHD), is a coronary artery blockage where the major vessels struggle to supply proper flow of blood as well as sufficient oxygen level to the heart. The word cerebrovascular is derived from two parts: "Cerebro” means the brain’s large part, and “vascular” is the arteries or veins. Cerebrovascular disease includes stroke and transient ischemic attack (TIA) that affect the flow of blood as well as blood vessels in the brain. The PAD is a situation in which arteries reduce the blood flow to arms and legs [4]. All these CVD problems pose major threats to human life, thereby mandating effective analysis and treatment. Economically, the cost of CVD treatment goes beyond the loss of human life since it also places a significant financial burden on healthcare systems, society, and people [5]. Thereby, an early diagnosis is essential for efficient treatment and management of CVD. This enables medical practitioners to react swiftly, therefore, preventing future problems and improving patient outcomes.

Historically, the diagnosing process of CVD has often used a mix of clinical examinations, diagnostic imaging procedures, and laboratory investigations [5]. However, these methods often have disadvantages, such as subjectivity, high costs, and time-consuming procedures, which may lead to delays in diagnosis. However, the recent diagnosis and treatments indicate that the correct and timely identification of CVD is very necessary for the successful treatment and management of cardiac disorders. A lot of conventional methods have been used for CVD diagnosis including angiography[12], Cardiac Catheterization[10], Computerized Tomography [11], Electrocardiogram[6], Echocardiogram [8], Heart CT scan [11], Holter monitoring[7], Magnetic Resonance Imaging (MRI)[13], and Stress test [9]. However, the conventional methods of diagnosis involve extensive dependencies on tests that are both expensive and inefficient, which may cause a delay in both the diagnosis and therapy [14].

Recently, many novel and advanced technologies have been integrated to bring the digital revolution in various sectors including healthcare. The healthcare sector is transforming into digital healthcare with the involvement of digital technological advancement and innovations. This drastic change has come in the last few years due to the involvement of Artificial Intelligence (AI), blockchain, cloud computing, Internet of Things (IoT), Cloud of Things (CoT), robotics, and more advanced innovative technologies. Technological innovations as listed above extensively utilize the capabilities of cloud computing and IoT. Cloud computing and IoT are growing technologies at the global level. Their adoption and use are expected to be more and more pervasive, making them important components of digital services.

Currently, the developments of Machine Learning (ML) and AI approaches in the healthcare sector have transformed the area of medical diagnostics and presented new prospects for early identification of CVD and also precise prediction of its progression. AI in healthcare brings the potential of health diagnosis and treatment of patients while involving image analysis of medical records as well as active health monitoring [15].

Machine learning offers significant innovations for data analysis without human interaction [16]. ML is becoming more crucial for healthcare professionals to predict meaningful and relevant results from patient health information. In the process of evaluating complicated and diverse medical data, ML algorithms have shown a great deal of promise for finding significant patterns and making correct predictions. The development of such skills has opened the way for the creation of intelligent CVD analysis models, which aid medical practitioners in making informed choices, boosting diagnosis accuracy, and optimizing patient care [16].

Machine learning algorithms are extensively used for decision-making as well as prediction systems. Various ML approaches are used to improve performance such as classification, prediction, and function approximation related to machine learning [17]. In recent times, there has been a growing trend for combining machine learning approaches as an ensemble for solving various issues. As compared with machine learning methodologies, ensemble learning has shown particularly promising outcomes [18]. This approach combines the predictions produced by numerous models to enhance both performance as well as resilience.

## 1.2 Cloud of Things

A novel paradigm where an integration of cloud and IoT is termed as Cloud of Things [19]. CoT provides a way to manage massive volumes of data with its complexity and IoT helps in creating an extended portfolio of services that can be provided with the use of sensors. It is expected that in the current year, 50 billion things will be interconnected via the Internet which will generate a massive amount of data thereby demanding larger storage space [20]. To handle such things, CoT will predict the things as a service to handle such circumstances via APIs.

CoT helps in providing and offering the various services associated with the healthcare sector including data monitoring, patient data storage, patient health data processing, data analysis, and visualization capabilities with the integration of IoT devices, and also enables the feature of communication and collaboration among patients and healthcare professionals [21]. CoT promotes smart healthcare applications and health-related services that promote the extensive use of the cloud through things such as sensors, which opens various issues as well as opportunities [22, 23]. It develops advanced communication between health analysis devices and also handles ever-increasing data demands [24, 25].

The incorporation of digital technologies such as cloud computing and IoT has brought various benefits to the healthcare sector [19]. Due to the technological integration, there is a drastic change in the medical field that has brought by improving the quality of healthcare services and reducing the time for healthcare offerings. Cloud computing helps healthcare industries to increase efficiency by automating operations, online record storage, offering health apps, and conclusively reducing the acquisition and operation costs.

The fast-growing IoT and cloud-enabled data processing help to provide better opportunities to develop smart and interconnected healthcare systems [26].

## 1.3 Digital Healthcare

Digital healthcare, often referred to as eHealth or digital health, is a broad and evolving field that involves the use of technologies such as CoT, ML, AI, and other digital tools for providing digital healthcare services. Digital healthcare enhances the delivery of healthcare services, improves patient outcomes, and streamlines healthcare processes [27]. Digital healthcare comprehends of variety of technologies and applications [19, 28], as shown in Figure 1.1. The various technologies related to digital healthcare are CoT, AI and ML, Blockchain, and health analytics. Figure 1.1 also depicts the various associated applications related to technologies that include EHR, Telemedicine, Telehealth, robotics, Digital Therapeutics, etc.

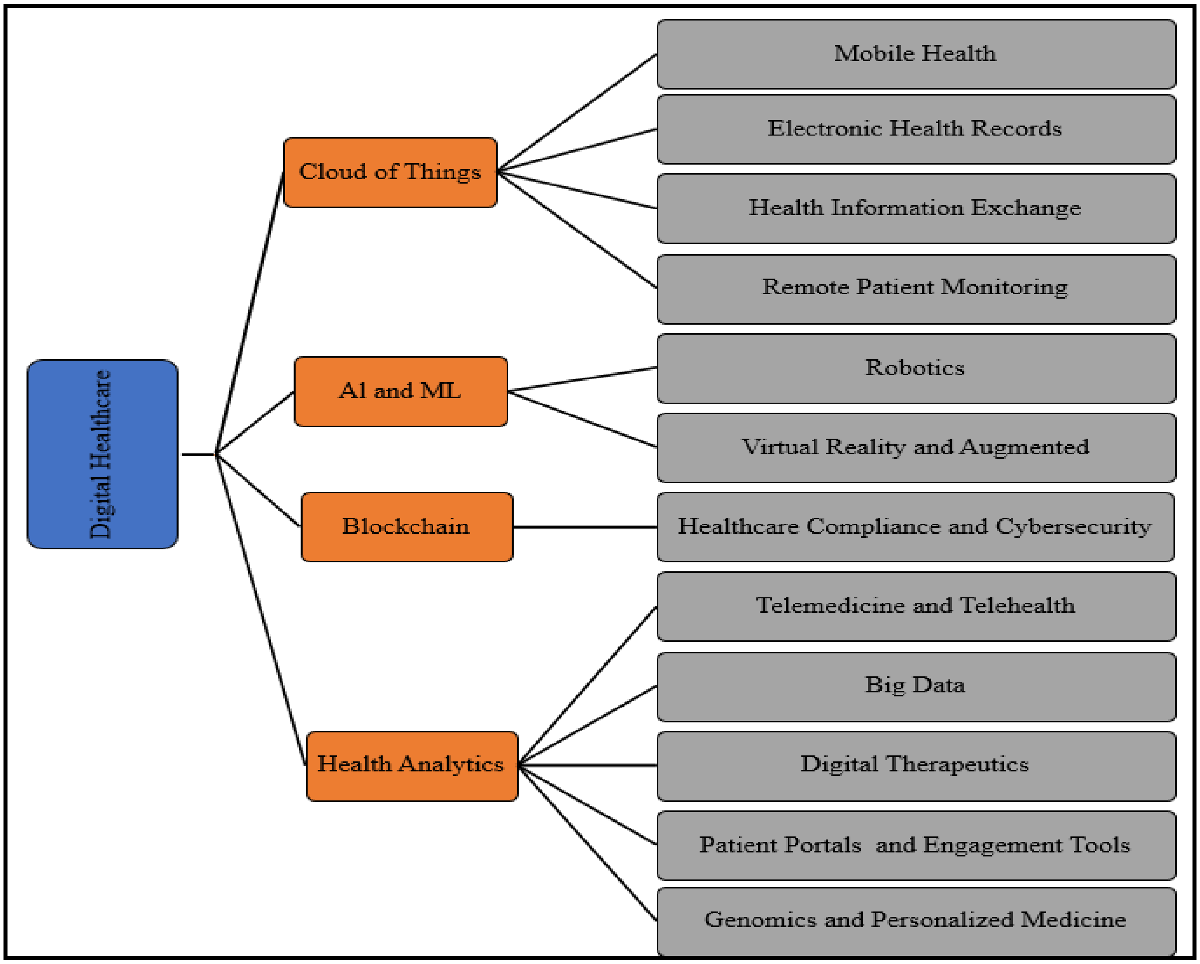



**Figure 1.1 Digital Healthcare Technologies**

### I. Cloud of Things (CoT)

CoT provides patient care facilities in terms of real-time health data analysis [19] and monitoring that helps to improve the healthcare services at local as well as global level healthcare industries.

- mHealth (Mobile Health): mHealth is related to the medical and public health services that refer to the mobile devices, such as smartphones and wearables, to monitor health, provide health information, and support medical care. Mobile apps can track vital signs, medication adherence, and fitness activities [29].
- Electronic Health Records (EHRs): EHRs replace paper-based medical records with digital versions. They allow healthcare providers to access and update patient

information easily, leading to better coordination of care and reducing errors at the global level as data is stored and analysis is performed in the cloud [30]. Cloud computing helps with data storage, analysis, and visualization of the data with remote accessibility as well as control.

- Health Information Exchange (HIE): HIE is defined as the digital transformation of medical reports. It enables the security for patient health information among healthcare organizations and providers during transformation. It facilitates continuity of patient care and reduction in duplication of tests and procedures [31].
- Remote Patient Monitoring: An IoT devices are used for real-time data collection of patient's health. It is particularly useful for monitoring patients with chronic conditions [32].

**II. Artificial Intelligence and Machine Learning**

The advanced technologies such as AI and ML approaches are used to support with diagnostics, treatment recommendations, drug discovery, and predictive analytics. They can help healthcare providers and professionals make more accurate and timely decisions [16]. AI can include the following: -

- Robotics in Healthcare: Robots are being extensively employed in the healthcare sector. They can assist with surgeries, patient care, and medication dispensing the healthcare procedures [16].
- Virtual Reality (VR) and Augmented Reality (AR): These technologies are used for pain management, medical training, and therapy. They can create immersive healthcare experiences [16,19].

**III. Blockchain in Healthcare**

Blockchain technology helps for the enhancement in the security and health data integrity of healthcare data, ensuring that patient records are tamper-proof and easily traceable.

- Healthcare Compliance and Cybersecurity: With the digitalization of healthcare data, ensuring compliance with regulations like HIPAA and robust cybersecurity measures are essential to protect patient information.
- Digital healthcare holds the promise of improving healthcare accessibility, quality, and efficiency while reducing costs [36].

## IV. Health Analytics

Health analytics provide the way for diagnosis and analysis of patient data for providing better health services using telemedicine and telehealth services [33]. Such services are described below:

- Telemedicine and Telehealth: Telemedicine use the concept of ICT technologies that involves the use of telecommunications technology to provide remote medical services such as medical care. It includes virtual doctor visits, remote monitoring, and teleconsultations, making healthcare more accessible to patients, especially in rural or underserved areas [33].
- Health Analytics and Big Data: Data analytics methods are used by healthcare organizations for healthcare analytics, and big data tools to large datasets, identify trends, and make data-driven decisions for improving patient care, population health management, and cost control [34].
- Digital Therapeutics: These are software-based interventions that treat or manage medical conditions. They can be used independently or in conjunction with traditional treatments and medications [35].
- Patient Portals and Engagement Tools: Patient portals give patients access to their health records, appointment scheduling, and communication with healthcare providers. Engaging patients in their care can lead to better outcomes [19].
- Genomics and Personalized Medicine: Advances in genomics and DNA sequencing enable custom-made medicine, where interventions as well as treatments are tailored to genomic makeup [19].

Overall, the Healthcare sector is boosting up with the integration of the above technologies to provide the services and treatment in digital ways such as EHR, PHR, and MRI, etc. With the amalgamation of digital technologies such as CoT, ML and AI in the healthcare sector , it is known as digital healthcare or Healthcare as a Service (HaaS) [28]. HaaS involves a wide number of services that include as shown in the Figure 1.2 [37]:

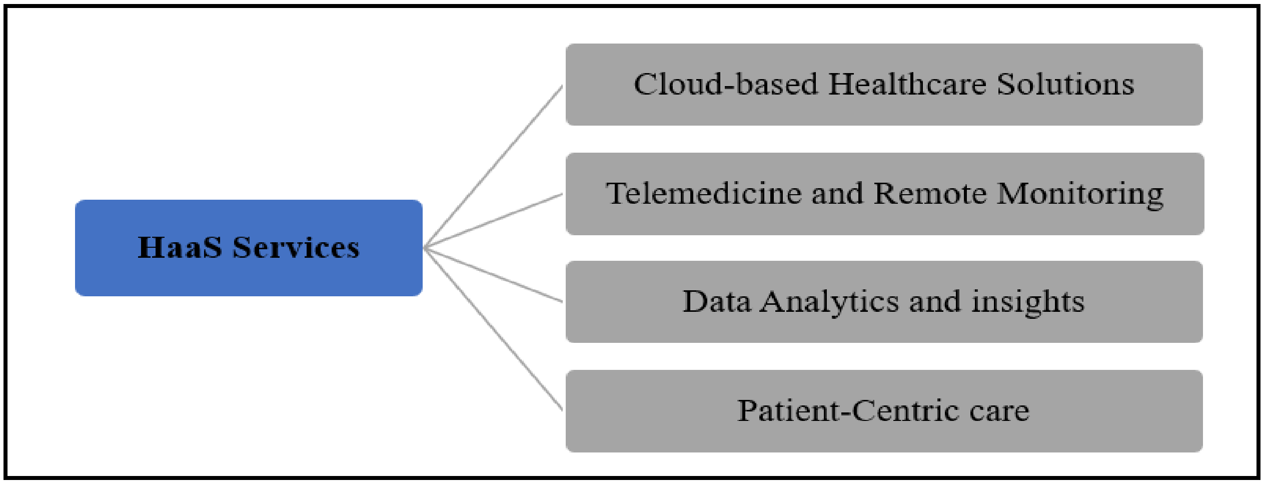



**Figure 1.2 HaaS Services**

- **Patient-centric care**: HaaS typically involves advanced technologies such as cloud computing for patient data storage and management in the form of electronic health records( EHRs), mhealth and clinical images. Cloud-based storage helps healthcare professionals and doctors to access patient information at the global level. Additionally, HaaS emphasizes the patient-centered data that allows the patients to control their health information as well as engaged them in decisions related to their health.
- **Telemedicine and remote monitoring**: HaaS provides a way to consult healthcare professionals with patients through telemedicine at the remote level. It provides the facility of the patient health monitoring through connected devices such as wearable watch and IoT devices.
- **Data analytics and insights**: By collecting and analyses of a large amount of the patient's health data, HaaS provide the valuable insights for healthcare professionals for better health treatment and lead for the better decision making and health care.

## 1.4 Machine Learning

Machine Learning (ML) is considered as a branch of Artificial intelligence (AI) that helps the machine to analyse the past data and also helps for model building as well as prediction using historical data. The ML algorithms are trained using the large volume of the data. It usually explores the dataset, constructs the data model, and helps for the prediction as well as generates the rules for recommendations [38]. Figure 1.3 shows the general machine-learning process [39].

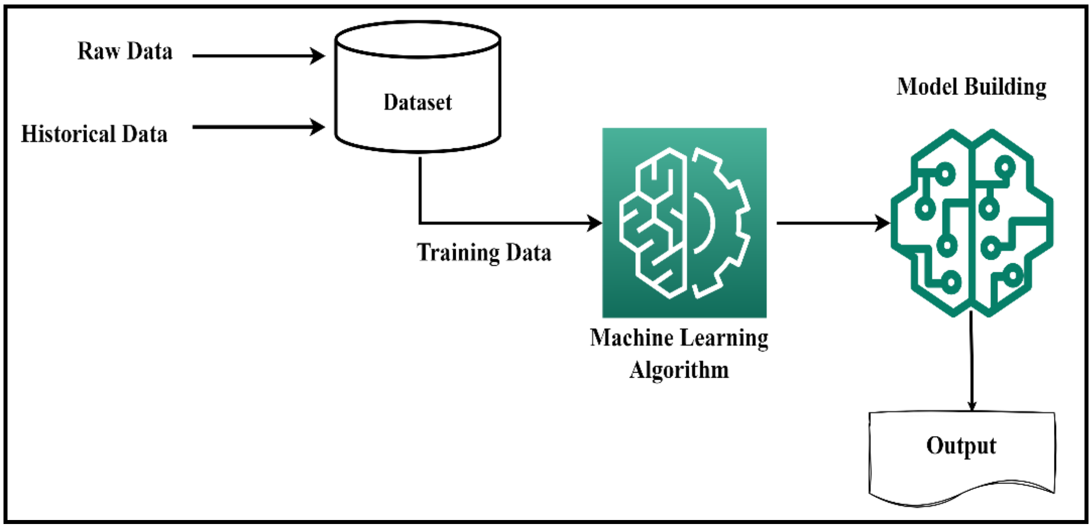



**Figure 1.3 General Machine Learning Process**

Machine learning approaches are classified as supervised learning (SL), unsupervised learning (USL), and reinforcement Learning (RFL). Supervised learning is like a supervisor in which the model is being trained using well-labeled data, like the presence of a teacher to teach the students in the class. In supervised learning, labeled data is used for model training using classification and regression methodologies [40]. Whereas Unsupervised learning is a way in which a machine works with the data without supervision, the model is being trained without labeled data using clustering and Association. Reinforcement learning acts as feedback-based learning using agents [41]. Table 1.1 shows the comparison between types of Machine Learning.

**Table 1.1 Comparison Between Different Types of Machine Learning**

| Sr No | Criteria | Supervised Learning | Unsupervised Learning | Reinforcement Learning |
|---|---|---|---|---|
| **1** | Definition | A model trained by labeled data. | A model trained without labeled data. | Interactively works with the environment. |
| **2** | Type of Data | Data having labeled data. | Unlabeled data is used | Undefined data |
| **3** | Associated Problem | Used for classification and regression | Used for Clustering and Association | Used for Exploitation as well as exploration |
| **4** | Supervision | Supervision required | No supervision required | No supervision is required |
| **5** | Algorithms | Linear regression, SVM, KNN, LR, Naïve Bayes | K-means, C-means, Apriori | Q-Learning, SARSA |
| **6** | Prospective Applications | Evaluation of risk | Recommender System | Healthcare |

### 1.4.1 Machine Learning Models

ML is the branch of AI that specifically emphasizes the advancements of ML algorithms and statistical-based models to enable the computer to perform specific tasks and activities. There are various ML models [42] that are being used for prediction and analysis as shown in Figure 1.4. ML models are the core of any prediction system that allows one to learn from the data and helps in decision-making and prediction.

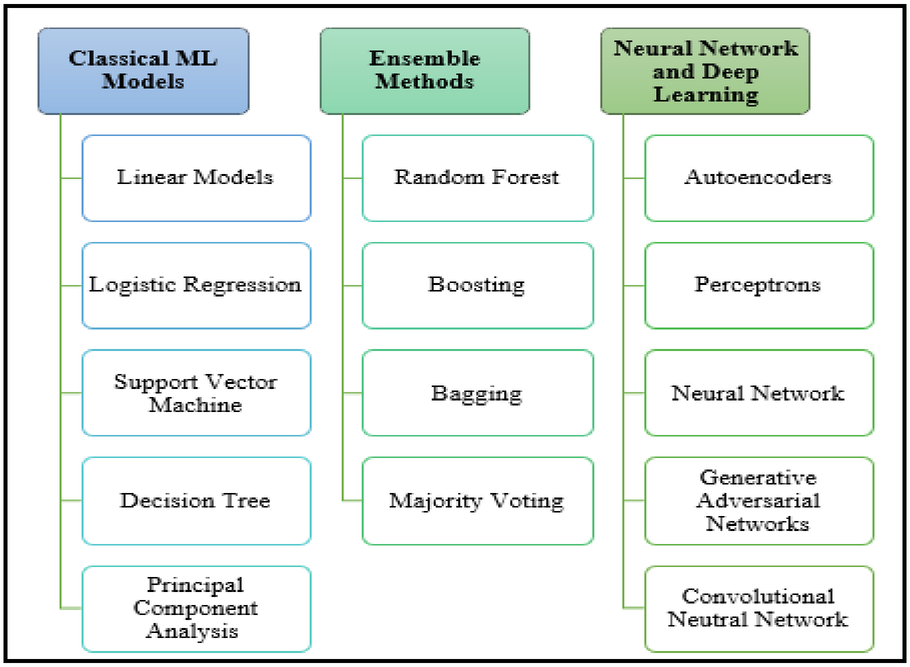



**Figure 1.4 Machine Learning Models**

1. **Classical ML Models:** Classical ML models are the fundamentals for the data science used for the prediction. Such models build the foundations for many advanced techniques and are used across many industries to solve a myriad of problems. A comprehensive detailed description of classical models is discussed in Table 1.2 below:

**Table 1.2 Classical ML models**

| Model | Types | Description | Use cases | Advantages |
|---|---|---|---|---|
| Linear Models [55] | Regressions | The model is proposed to initiate the linear association between the input and output variables as a linear equation where the input variables are considered as | Continuous Values Predictions | • Simple to implement<br>• Interpretable<br>• Fast to train |

| Model | Types | Description | Use cases | Advantages |
|---|---|---|---|---|
| | | features and the output variables are predicted as targets. Such models are used for the predictions. | | |
| Logistic Regression [56] | Classification | The proposed model is for the prediction of the probability in binary outcomes such as 1/0, and yes/no using logistic functions. Such models are mostly used for detection. | Binary classification | • Output probabilities<br>• Easy to interpret |
| Support Vector Machines [57] | Classification | SVM separates the different classes and creates complex decision boundaries. | Text classification and image recognition | • Effective in high-dimension spaces.<br>• Robust in overfitting |
| Decision Tree [58] | Classification and regression | The model splits the data into branches or subsets to make decisions based on the input values such as features. | Risk assessment and loan approval | • Handles no-linear data<br>• Easy to interpret |

| Model | Types | Description | Use cases | Advantages |
| --- | --- | --- | --- | --- |
| Principal Component Analysis[59, 60] | Dimensionality Reduction | The model transforms the data into new coordinate systems and reduces the dimensions. | Data compression and noise reduction | • Reduces computational costs<br>• Noise removal |
| Random forest [61] | Classification and regression | Random forest is an ensemble of multiple decision trees to improve the performance and overfitting reduction. | Predictions | • Reduce overfitting<br>• Handling of high-dimensional data |
| K-Nearest Neighbors (KNN) [62] | Classification and regression | KNN makes the classification of the data point based about the classes of the neighbors. | Recommendation system and pattern recognition | • Simple<br>• Non-parametric |
| Naive Bayes [63] | Classification | The naive Bayes classification is created using the Bayes theorem, which assumes the features' independence. | Text classification and spam detection | • Fast<br>• Suitable with high-dimensional data |
| K-Means clustering[64] | Clustering | K-Means clustering is considered for the partitioning of the | Customer segmentation and image | • Simple<br>• Scalable with large |

| Model | Types | Description | Use cases | Advantages |
|---|---|---|---|---|
| | | data into different clusters that is represented by the mean of its point. | compression | datasets |
| Gradient Boosting Machines (GBM) [65] | Classification and regression | GBM builds an ensemble of weak models to improve the performance. | Ranking, Classification, and Regression | • High prediction accuracy<br>• Handles mix data types |

2. **Ensemble Methods:** Ensemble learning is a machine learning model that combines two or more learners to produce a better prediction model. The main purpose is to integrate the predictions of multiple models to improve performance. A detailed description of ensemble methods [66] is discussed in Table 1.3 given below:

**Table 1.3 Ensemble Learning methods**

| Ensemble method | Description | Algorithms used | Use cases | Advantages |
|---|---|---|---|---|
| Boosting [67] | In boosting, various ensemble models are trained , each model address the specific errors related to the models output of the | AdaBoost,<br>Gradient Boosting, XGBoost, LightGBM, CatBoost | Binary classification, spam detection, image recognition, House price prediction, customer behavior prediction,<br>Anomaly detection | • Reduces both bias and variance<br>• Lead to high accuracy |

| Ensemble method | Description | Algorithms used | Use cases | Advantages |
|---|---|---|---|---|
| | preceding one. The final model is weighted. | | | |
| Bagging [68] | Bagging trains the multiple models in parallel on different training data subsets. The final prediction model is made using regression and classification of the predictions of all the models. | Random forest | Fraud Detection, Spam detection, sales forecasting, Customer segmentation | • Reduce variance<br>• Prevent overfitting |
| Voting [69] | In voting, the predictions of multiple models are combined together by regression and classification. | VotingClassifier | Customer Churn Prediction | • Simple to implement<br>• Performance improvement |
| Stacking [70] | In stacking, base learners (multiple models) are trained and their predictions are | StackingClassifier | Prediction of Real Estate | • Pattern capturing<br>• Improve the performance |

| Ensemble method | Description | Algorithms used | Use cases | Advantages |
| --- | --- | --- | --- | --- |
| | used for higher-level models as input features to make the final predictions. | | | |

3. **Neural Network and Deep Learning:** Neural networks and deep learning are the subfields of machine learning that enable expressive advances in digital areas such as speech recognition, natural language processing, and computer vision. A neural network consists of layers of connected neurons, where each neuron represents mathematical operations and functions. There are various methodologies related to neural networks and deep learning, which are discussed in Table 1.4 below:

**Table 1.4: Neural Network and Deep learning methods**

| Methods | Description | Algorithm used | Use cases | Advantages |
| --- | --- | --- | --- | --- |
| Autoencoders [71] | Autoencoders are artificial neural networks used as unsupervised learning for efficient coding. | Backpropagation | • Fraud and anomaly detection<br>• Noise removal<br>• Data compression | • Labeled data not required |
| Perceptrons [72] | Perceptrons are artificial neural networks used as building | Perceptron based weighted, Linear regression | • Pattern recognition<br>• Classification | • Computational Efficiency<br>• Adaptability |

| Methods | Description | Algorithm used | Use cases | Advantages |
|---|---|---|---|---|
| | blocks for larger neural networks. | | | |
| Neural Network [73] | Neural Network is a machine learning algorithm class consisting of layers of interconnected neurons to learn the patterns. | Backpropagation | • image recognition,<br>• autonomous systems<br>• natural language processing<br>. | • Flexibility<br>• Scalability<br>• Automatic feature extraction<br>• |
| Generative Adversarial Networks [74] | Generative Adversarial Networks is a machine learning class consisting of two neurons, one generator, and another discriminator. | Stochastic Gradient Descent | • Image Generation<br>• Video Generation<br>• Text-to-image synthesis | • Super Resolution<br>• Improve stability<br>• Efficiency |
| Convolutional Neural Network [75] | A deep neural network is used for grid processing. | | • Image classification<br>• Object detection<br>• Segmentation | • Enhance capability<br>• Easy adaptability |

### 1.4.2 Machine Learning and Digital Healthcare Provisioning

Machine learning is considered as a subset of AI that plays an important and crucial role in the medical sector. It helps healthcare professionals monitor real-time in less time, provides them with recommendations for better treatment facilities, and helps in clinical data management [43, 44]. The ML and digital healthcare provisions are facilitated with the support of CoT. ML based healthcare models using cloud capability.

ML is the way of providing the data set to the computers as well as algorithms to process it and perform the analysis [43]. ML plays a critical and vital role in the disease prediction which is not easy manually. Machine learning has different applications associated with the healthcare sector [44] as shown in Figure 1.5. These applications significantly renovate the healthcare field and industry by improving the patient outcomes, cost reduction, and enhancement[45,46] in healthcare delivery in efficiency.

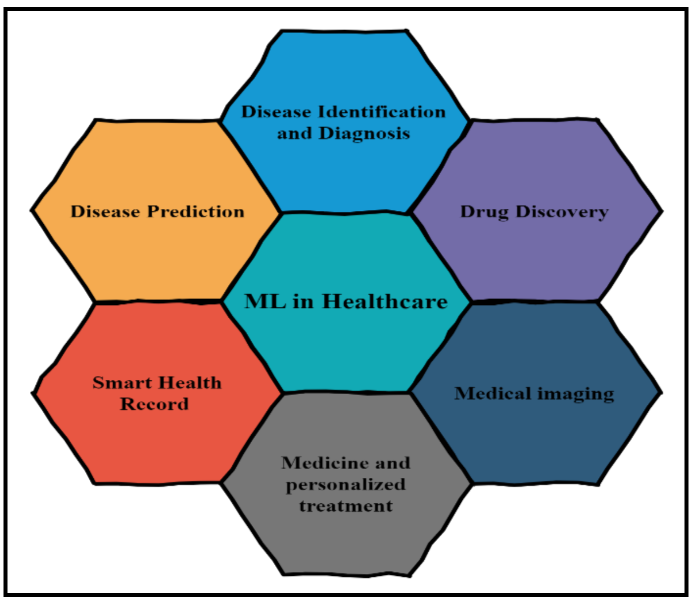



**Figure 1.5 Machine Learning in Healthcare**

ML is significantly transforming medical imaging with the improvement of accuracy, efficiency, and capability of image analysis and diagnosis. ML algorithms are used to analyze medical images that including MRI and X-rays for anomaly detection. Additionally, ML also diagnoses and automates the disease and analyzes the disease patterns. The role of ML in medicine and personalized treatment is transforming healthcare by enabling more accurate diagnosis services, patient outcomes prediction, and tailoring treatments to individual patients. Various ML algorithms can analyze the clinical data to recommend personalized treatment plans most likely most effective for each and every patient by considering their unique behavioral characteristics and medical history.

ML in smart health records enhances patient data management, analysis, and utilization. This leads to improvements in healthcare delivery. Integrating Machine learning and NLP methodologies can extract and drive meaningful insights from unstructured data in electronic health records (EHRs), such as clinical notes and pathology reports, enhancing the comprehensiveness of patient records. ML plays a crucial role in disease prediction by leveraging vast amounts of healthcare data for pattern identification and making accurate predictions about disease onset, progression, and outcomes. The integration of genetic information with other related health data, ML can predict which patients are likely to benefit from specific treatments, enabling personalized medical approaches.

ML helps in the identification of diseases based on hidden patterns and image analysis. ML can also identify the early signs of diseases that lead to disease diagnosis and treatment. ML can predict the likelihood of certain diagnoses, aiding clinicians in considering possible conditions based on presenting symptoms and medical history. ML models can accelerate the process of drug discovery by predicting how different compounds will interact with targets in the body, potentially reducing the time and cost of developing new medications.

**1.4.3 Healthcare Analytics**

Healthcare analytics is the process of collecting, analyzing, and interpreting healthcare data to improve healthcare service delivery, patient outcomes, and overall healthcare system efficiency using data analysis and ML capabilities. It encompasses the application of multiple methods and instruments to extract insights from extensive amounts of healthcare-related information, including Electronic Health Records (EHRs), claims information, patient questionnaires, and others.

In modern healthcare [47] management, Healthcare analytics is considered as a crucial process, as it helps healthcare organizations to make data-driven decisions to perform various operations. Table 1.5 depicts how healthcare analytics helps in healthcare sectors.

**Table 1.5 Categorization of Healthcare Analytics data-driven decisions**

| Sr no | Categories | Description |
|---|---|---|
| 1 | Improve patient care [48] | • Categorize patterns insights and trends from patient data to provide personalized treatment plans.<br>• Predict and prevent negative patient’s outcomes such as hospital readmissions or complications.<br>• Monitor and manage chronic diseases more effectively. |
| 2 | Enhance operational efficiency [49] | • Optimization of resources allocation, including staff scheduling and inventory management.<br>• Reorganize administrative processes, such as billing and claims processing.<br>• Reduce healthcare costs by identifying areas for cost containment and waste reduction. |
| 3 | Quality improvement [50] | • Measure the healthcare quality metrics, such as patient satisfaction and clinical outcomes.<br>• Identify best practices and areas for improvement in the delivery of healthcare services.<br>• Support evidence-based decision-making in clinical practice. |
| 4 | Population health management [51] | • Identify the risk populations and the development of targeted interventions for better health-related outcomes.<br>• Track health trends in communities and respond to public health issues more effectively. |

| | | |
|---|---|---|
| 5 | Research and Development [52] | • Accelerate medical research by analyzing large datasets to discover new treatments or therapies.<br>• Support clinical trials and drug discovery by identifying suitable candidates and monitoring outcomes. |

### 1.4.4 Diverse Healthcare Analytics Techniques in Healthcare Sector

Healthcare analytics have a substantial impact on patient care and the overall healthcare system. It helps the healthcare providers, payers, and policymakers to make better decisions that improve the patient outcomes, reduce costs, and enhance the quality of healthcare services. However, it also raises important ethical and privacy considerations regarding the sensitivity of the patient data, which must be managed carefully to ensure compliance with healthcare regulations and protect patient privacy [53, 54]. There are some common techniques and technologies used in healthcare analytics [53] as shown in Figure 1.6 :

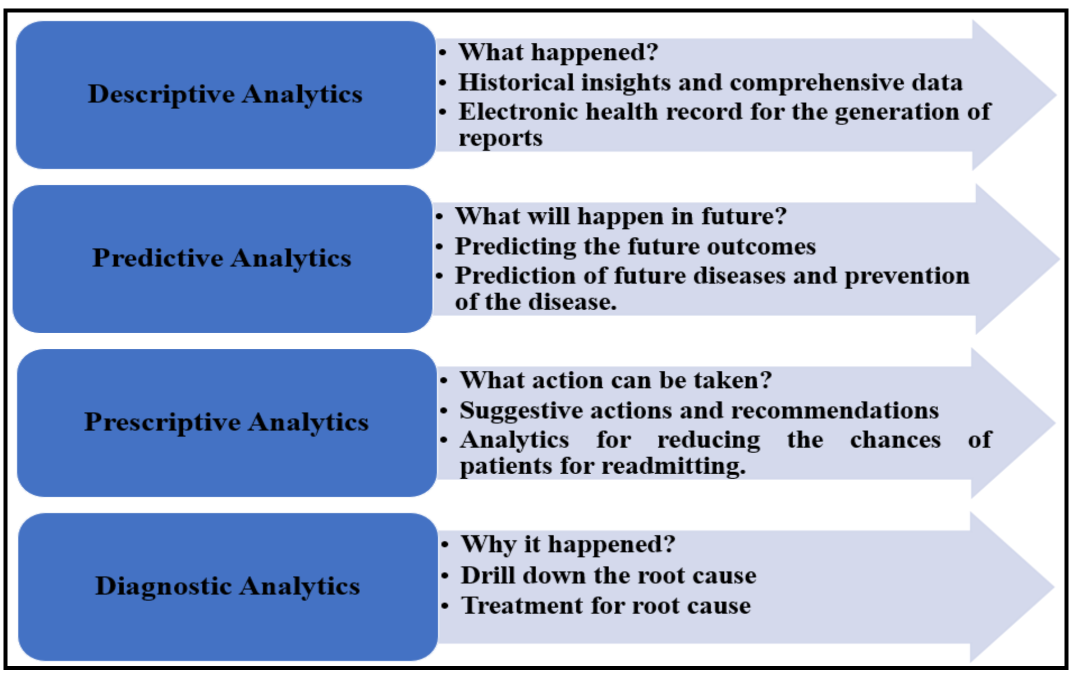



**Figure 1.6 Healthcare Analytics Categories**

a) **Descriptive Analytics:** In the healthcare sector, Descriptive analytics helps in the patient's historical data to gain insights from past issues, patterns, trends, and outcomes. It becomes easy to provide better treatment based on existing symptoms [54]. Descriptive analytics provides insights into what has happened within the healthcare system and also helps organizations to make decisions and improve patient care.

b) **Predictive Analytics:** Predictive analytics helps in forecasting future outcomes or trends based on historical health data and statistical models. It allows the healthcare providers to anticipate and prevent the problems and made easy to predict and diagnose the diseases [54]. Predictive analytics helps for the improvement in patient care outcomes and quality of patient care. Predictive analytics can reduce healthcare costs through resource optimization such as staff, equipment, and facilities, and improvement in operational efficiency.

c) **Prescriptive Analytics:** Predictive analytics comprises of past patient data, and advanced ML algorithms to recommend the actions that will achieve the desired outcomes such as optimizing decisions using predictive models to reduce the chances of disease to be occurred again [53, 54]. Predictive analytics help in the improvement of patient outcomes and quality of care by providing personalized and evidence-based recommendations.

d) **Diagnostic Analytics:** Diagnostic analytics is a data-driven technique that helps to find the root causes of diseases and identify why this disease happening, also it tends to analyze the reasons for the medical history based on symptoms and different features [53, 54]. Diagnostic Analytics specifically investigates past performance to analyze the reasons behind any outcomes.

## 1.5 Problem Statement

The need for an accurate and efficient model for the study of CVD is a prominent issue that needs a lot of research. CVD prediction has emerged as a prominent issue due to high mortality rates owing to CVD and associated problems. There has been extensive growth in the adoption and implementation of digital healthcare paradigm in the last decade. Thus, a digitally supported CVD prediction strategies and tools are of current demand in the healthcare sector. Driven by this, a model that would assist medical practitioners in the rapid diagnosis, risk assessment, and tailored treatment of cardiovascular diseases must be developed.

The need also originates up from the drawbacks and challenges in the traditional healthcare methods for diagnosing cardiovascular diseases. Such challenges are subject to high costs, and lengthy procedures. As a result, diagnosis and the therapy process are often delayed. Moreover, the complexity and variety of CVD data pose difficulties in making reliable predictions and classifications. Thus, an intelligent analysis model for the identification of CVD is required.

An intelligent digital analysis model can predict the CVD and associated levels, considering multiple factors that contribute to the CVD and associated risks. Such models intend to not only collect, process, and analyze the patient's EHR data to predict the CVD but also offer recommendations to reduce the CVD risks.

In the thesis work, an ensemble model is proposed that combines various classifiers, such as Convolutional Neural Network (CNN), Naive Bayes (NB), k-nearest Neighbors (kNN), Multilayer Perceptron (MLP), and Support Vector Machine (SVM), to improve the accuracy and robustness of CVD analysis. The model is used for feature extraction from patient data, while the different views and decision-making skills of the other classifiers are included into the ensemble model. This allows the ensemble model to harness the complementary capabilities of these classifiers.

Challenges Driving the Need for the CVD analysis and prediction model:

- **Diagnostic approach**: Existing diagnostic approaches for cardiovascular diseases may lack the requisite accuracy and reliability owing to the subjective and theoretical character of these methods or their limited capacity to manage the complexity and variety of cardiovascular disease datasets and samples. To facilitate more effective clinical decision-making, a more precise and reliable analytical model that can provide correct predictions and classifications is required.

- **Delayed diagnosis and the subsequent beginning of therapy:** Traditional diagnostic procedures sometimes require long processes, which results in delayed diagnosis and, as a consequence, the initiation of treatment. This may have an unfavorable impact on the patient's overall result. To facilitate early identification and rapid intervention, which will ultimately result in an improved patient prognosis, a CVD analytic model that is both fast and efficient is needed.
- **Difficulty in Accessing Specialized Healthcare Facilities and Experts**: People who live in rural or underdeveloped regions sometimes have difficulty in gaining access to specialized healthcare facilities and professionals. This gap may be bridged with the use of an online CVD analysis model that offers remote access to accurate risk assessment and individualized recommendations, hence enhancing accessibility to high-quality healthcare services.
- **Decision-making in real time**: Cardiovascular diseases require the use of decision-making to evaluate risks, monitor patient status, and give suitable treatment methods. An online model that is capable of processing data in real time and providing instant feedback may be of tremendous assistance to medical practitioners in terms of making educated choices and beginning treatments at the appropriate moment.

By addressing these challenges, the proposed model intends to develop a smart and intelligent platform for the analysis of CVD. This model will leverage ensemble learning techniques, integrate multiple classifiers, and make use of web-based technologies in order to provide accurate, efficient, and easily accessible diagnosis, risk assessment, and personalized management of cardiovascular diseases.

Driven by this, this thesis presents a digital CVD prediction and recommendation model, named PCard. This model intends to analyze the EHRs stored in the cloud and sensor data of patients obtained from IoT sensors and predict the CVD risk levels.

## 1.6 Research Objectives

Every research work has some key stages that are integrated with research objectives that define the research. To frame the research objectives, need to do surveys and review many existing studies are reviewed that are similar to the proposed work. Below are some objectives that are framed during the literature survey as shown in Table 1.6.

**Table 1.6 List of Research Objectives**

| ID | Description |
|---|---|
| RO-01 | To study and analyze the existing Cloud of Things (CoT) based frameworks for healthcare. |
| RO-02 | To design and develop a framework for offering Healthcare as a Service using the CoT. |
| RO-03 | To implement the proposed framework on cardiovascular problems. |
| RO-04 | To compare the proposed framework with the existing frameworks |

To accomplish these objectives such as RO-01 to RO-04, various steps have been followed that are have been discussed below:

1. **Dataset Development:** Develop a complete and representative dataset for cardiovascular disease analysis that comprises a variety of clinical measures, ECG signals, and medical samples. This is to be done so that an analysis of cardiovascular disease can be performed. Integrating data from a variety of demographics, CVD subtypes, and severity levels ensures the quality of the dataset, as well as its dependability and diversity.

2. **Development of an Ensemble Model:** Conceive of and put into action an ensemble based model that incorporates the various classifiers such as CNN, kNN, MLP, NB, and SVM. It helps to create an integrated decision-making system by developing suitable algorithms for ensemble learning, which will utilize the capabilities of individual classifiers. The ensemble approach maximizes the model's performance by optimizing its architecture and hyperparameters to achieve the highest possible level of accuracy, sensitivity, specificity, and computing efficiency.

3. **Evaluation of Performance**: The ensemble model evaluates the performance of the proposed model by conducting exhaustive experiments and evaluations. In this phase, Performance evaluation of the ensemble model has been conducted in comparison to the performance of individual classifiers techniques to CVD analysis. When determining how well the model is doing, it is important to make use of relevant assessment measures such as performance matrices

4. **The design and development of a Web-Based Interface**: A web-based interface is designed and constructed that is both user-friendly and compatible with an online CVD analysis model. The following are the features of the online CVD analysis model:

- Integrate user-friendly, interactive elements that enable patients and medical professionalists to enter pertinent details, see the outcomes of those analyses, and get individualized suggestions.
- Ensure that the interface can be used easily, that it responds appropriately, and that it is compatible with a variety of devices and web browsers.

5. **Validation and Case Studies**: To evaluate the usefulness, efficacy, and practicability of the suggested model, you will need to carry out a number of comprehensive validation experiments and case studies. The validation process is performed in the following steps:

- Evaluate the performance of the model using real-world datasets on CVD, taking into account a variety of demographic groupings and clinical circumstances. Demographic grouping includes factors such as age, gender, ethnicity, and socioeconomic status, along with clinical circumstances such as a patient’s history of chronic conditions like diabetes and hypertension. It also takes into account lifestyle factors, including diet, alcohol consumption, smoking habits, and physical activity levels.
- Solicit the comments and suggestions of medical practitioners and patients so that the functionality of the model and the quality of the user experience may be further developed and enhanced.

The purpose of this thesis is to contribute to the area of CVD analysis by offering an accurate, robust, and easily available online model for the early diagnosis, risk assessment, and individualized treatment of cardiovascular illnesses. This will be accomplished by achieving the research goals listed above. These findings have the possible to significantly advance the diagnosis of CVD, improve the outcomes for patients, and make substantial strides in expanding the use of machine learning and artificial intelligence in medical settings.

## 1.7 Research Questions Addressed

The research work has appropriate questions, which are listed below in Table 1.7:

**Table 1.7 Research Questions and Their Probable Answers**

| Sr No | Research Questions | Answers to the Questions |
|---|---|---|
| RQ1 | In terms of performance and specificity in CVD analysis, how does the suggested ensemble model, which incorporates CNN, NB, kNN, MLP, and SVM classifiers, compare to individual classifiers? | The performance of the ensemble model will be assessed in comparison to the performance of individual classifiers using a variety of assessment measures. The findings will shed light on the efficiency of the ensemble method and demonstrate its superiority over that of individual classifiers. |
| RQ2 | How does the ensemble model deal with the variability and complexity of CVD data in comparison to standard diagnostic methods? | The capability of the ensemble model to deal with diverse and complicated CVD data will be evaluated by comparing its performance with that of established diagnostic methods. The accuracy of the model in predicting cardiovascular disease subtypes and severity levels will be the primary emphasis of the assessment. This will be shown by the model's capacity to evaluate a variety of clinical measures and patient samples. |
| RQ3 | What kind of effect would the incorporation into the ensemble | The contribution of CNN to the processing of ECG signals and visual representations will be examined |

| Sr No | Research Questions | Answers to the Questions |
| --- | --- | --- |
| | model when it comes to the processing of ECG data and visual representations in CVD analysis? | by contrasting the performance of the ensemble model when it is used with and when it is used without the CNN classifier. The purpose of this investigation is to shed light on the efficiency of CNN in extracting significant characteristics from CVD-related data and its contribution to overall diagnostic accuracy. |
| RQ4 | How does the fact that the CVD analysis model is available online increase accessibility and make it possible to make decisions in real time? | We will evaluate the online CVD analysis model's usability, responsiveness, and effectiveness in terms of giving rapid risk assessment and individualized suggestions in order to determine the influence it has on accessibility and real-time decision-making. Case studies and feedback from users will provide light on how successful the approach is in increasing accessibility and paving the way for timely responses. |
| RQ5 | How well does the ensemble model perform in contrast to other state-of-the-art CVD analysis methods? | In order to determine how successful it is, the performance of the ensemble model will be compared against the performance of other techniques to CVD analysis that are considered to be state of the art. In order to evaluate the performance of the model and figure out the extent of its competitive advantage over other methods already in use, the comparison will take into account measures. |
| RQ6 | How does the ensemble model respond to new data and shifting trends in CVD diagnosis? | The capacity of the ensemble model to adapt to new data and shifting patterns in CVD diagnosis will be investigated by continually updating the |

| Sr No | Research Questions | Answers to the Questions |
|---|---|---|
| | | model with fresh data and measuring its performance over time. This will be done in order to test the adaptability of the ensemble model. The flexibility of the model, as well as its continued relevance in today's rapidly changing healthcare systems, will be shown by this examination. |
| RQ7 | Regarding the suggested online CVD analysis model, from the point of view of healthcare professionals, what are the strengths and limitations of this model, and how may these be improved? | It is planned to conduct surveys or conduct interviews with healthcare experts in order to obtain their feedback on the strengths and limitations of the online CVD analysis model. The model's practical implications, possible advantages, and opportunities for development may all be better understood with the help of the material provided here. |
| RQ8 | Regarding the use of the online CVD analysis model for risk assessment and treatment, what are the viewpoints and experiences of patients? | Through questionnaires or in-person interviews, responses from patients about their experiences using the online CVD analysis model will be collected. These data will help throw light on the user experience, usability, and efficacy of the model from the viewpoint of the patient, which will assist in the refining of the model and its individualized suggestions. |
| RQ9 | In terms of accuracy and sensitivity, how does the ensemble model perform when applied to diverse demographic groups? | To determine the ensemble model's degree of accuracy and sensitivity, we will examine how well it performs across a variety of demographic categories, including age, gender, and ethnicity, among others. This study will assist evaluate whether or not the model exhibits any potential |

| Sr No | Research Questions | Answers to the Questions |
|---|---|---|
| | | biases or differences in its performance, and it will make it possible to design CVD analytic tools that are more inclusive and equitable. |
| RQ10 | Regarding the deployment of an online CVD analysis model in real-world clinical settings, what are some of the possible obstacles and ethical issues to take into account? | We will investigate the various difficulties and ethical issues that may arise during the deployment of the online CVD analysis model in real-world clinical settings. In this study, topics like as data privacy and security, as well as the ethical ramifications of depending on AI-based systems for important medical choices, will be discussed. It will shed light on the essential protections and criteria that must be adhered to in order to guarantee the appropriate and ethical use of the model in clinical settings. |

## 1.8 Thesis Contribution

This thesis intends to contribute to the area of cardiovascular disease detection by constructing an ensemble model. The proposed model will exploit online capabilities for higher accuracy and real-time analysis, and it will integrate the strengths of different classifiers to create a more robust diagnosis. The findings of this thesis have the potential to bring about a revolution in the diagnosis of CVD, making it possible to diagnose these life-threatening disorders earlier, to provide medication on an individual basis, and to effectively control them. The proposed research work offers significant benefits for the healthcare sectors that are highlighted below:

1. The proposed model is available at the remote level for healthcare professionals as well as patients from a distant location remotely, hence making CVD risk assessment levels easier. This accessibility is especially helpful for those who live in rural or underdeveloped regions and have restricted access to specialist medical institutions because of their location.

2. The research has the latent to improve the early diagnosis of cardiovascular disease and provide the appropriate outcomes and recommendations for patients.
3. The suggested approach can improve the accuracy and efficiency of CVD analysis by leveraging the power of ensemble learning and making it accessible online. This may lead to early diagnosis of these life-threatening illnesses, as well as individualized treatment methods and successful management of these disorders.
4. The study contributes to the expanding area of ML-based medical diagnostics, which validates the practicality and efficiency of AI technologies in the medical industry.
5. With the help of the proposed model, the drawbacks of existing diagnostic methods can be solved, by investigating the possibilities offered by ensemble learning and devising an original CVD analysis model, known as PCard model.
6. The findings of this study have the potential to completely revolutionize the area of cardiovascular disease diagnosis, making available to both medical professionals and patients an instrument that is dependable, easily accessible, and highly effective for the early identification, risk assessment, and individualized treatment of cardiovascular illness.
7. The integration of numerous classifiers with the process of ensemble learning may further contribute in the accuracy improvement and resilience of the proposed model.
8. The improvement of CVD diagnosis and the overall performance of the analysis model may both be improved by exploring the potential of ensemble learning methods, which can facilitate to the ensemble learning techniques.

## 1.9 Thesis Organization

This thesis investigates the use of machine learning in the study of CVD, with a particular emphasis on the development of the framework using ensemble approach. The thesis is divided into 6 chapters which are organized as follows:

Chapter 1: It draws attention to the Cardiovascular diseases with its ailments. This provides a clear grasp of the Cloud of Things and digital healthcare. It showcases the use of machine learning in healthcare, various ML-based models and healthcare analytics. It highlights the objectives of the thesis and research based questions. In the end, it provides an overview of the structure of the thesis. Chapter 1 derives from:

- Shilpa, Kaur, T. (2022). Digital Healthcare: Current Trends, Challenges and Future Perspectives. In: Arai, K. (eds) Proceedings of the Future Technologies Conference (FTC) 2021, Volume 2. FTC 2021. Lecture Notes in Networks and Systems, vol 359. Springer, Cham. https://doi.org/10.1007/978-3-030-89880-9_48.
- Shilpa, Kaur, T. (2019). Blockchain as Reverse Approach From Centralized To Decentralized Database. Think India Journal, 22(17), 2769-2773.
- "Data Analysis using Association Rule in Data Mining", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.5, Issue 11, page no.316-319, November 2018.
- Cardiovascular Disease and Covid-19: Unveiling Shared Risk Factors and the Role of Technology in Advancing Healthcare Innovation for Sustainable Development----- (Communicated: Journal: Discover Sustainability)

Chapter 2, which is devoted to the study of the relevant literature, presents an overview of CVDs, and the conventional methods for conducting CVD analysis are presented. This chapter investigates the uses of advanced technologies, especially in the context of CVD analysis, and outlines the current obstacles that are present in the area. The chapter comes to a close with a summary, which focuses on the most important conclusions from the literature review. The chapter comes to a close with a summary, which reviews the salient issues that were covered. Chapter 2 partially derives from:

- Shilpa, Kaur, T. (2022). Blockchain and Cloud Technology: Leading the ICT Innovations. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. Lecture Notes in Networks and Systems, vol 321. Springer, Singapore. https://doi.org/10.1007/978-981-16-5987-4_41.

- Shilpa,, Kaur, Tarandeep and Garg, Rachit. "Digital healthcare: A topical and futuristic review of technological and robotic revolution" Paladyn, Journal of Behavioral Robotics, vol. 14, no. 1, 2023, pp. 20220108. https://doi.org/10.1515/pjbr-2022-0108.
- Healthcare-as-a-Service Provisioning using cloud-of-things: A Contemporary Review of Existing Frameworks Based on Tools, Services and Diseases ( Scalable computing Indexed: Scopus ) Accepted

The purpose of this chapter 3 is to provide an introduction to CVD analysis. This offers a general introduction to machine learning, with a particular emphasis on the ensemble model approach. The architecture and layers of the proposed model are broken out here, as is the method of training and evaluating the model. In addition to that, the chapter describes the performance measures that are used for CVD analysis and covers the selection and preparation of datasets. The chapter comes to a close with a summary, which reviews the salient issues that were covered.

The experimental design and data collection are the primary focus of Chapter 4. It gives a comprehensive explanation of the experimental setup that was used for the research, including the procedures for data collection and preparation. This chapter investigates several methods for augmenting data and discusses the preparation stages that are carried out on the data that was obtained. The outcomes of the comparative study with previously established methods are then evaluated and examined. A discussion on the performance of the model is also included in this chapter, in which both the model's strengths and shortcomings are highlighted.

The conclusion of the thesis may be found in Chapter 5, which provides the conclusion and future directions. The significant contributions made by the thesis are emphasized, along with the consequences such contributions have for analyses of CVD and healthcare. The thesis comes to a close with a complete conclusion that focuses on the most important findings from the study and recommendations for more research that should be conducted in the future.

## Conclusion

This chapter discusses the healthcare and advanced technologies that are integrated with healthcare to promote healthcare as digital healthcare. We have also discussed the basic ML algorithms with their categories and applications. The chapter also highlights healthcare analytics. In this chapter, about research objectives and scope are highlighted.

# CHAPTER-2

# LITERATURE REVIEW

Globally, Cardiovascular Disease (CVD) is the main cause of mortality and disability. Owing to the high mortality rates due to CVD, cutting-edge medical technologies must be developed that can be useful in aiding CVD diagnosis. Not only diagnosis, but technological innovations in healthcare can enhance risk assessment, treatment planning, and monitoring of CVD problems. Overall, the identification and treatment of CVD have been made easier with the help of several technical tools, methods, and platforms used in healthcare. This chapter serves as an introduction to the many healthcare technologies that have been and are currently being used in the study of CVD and associated risks. It highlights the use of machine learning approaches in offering digital healthcare services. It discusses the traditional and contemporary approaches used for CVD prediction and analysis. Additionally, it presents the various risk factors that augment the CVD problems. A comparative analysis of the existing CVD prediction models based on Cloud of Things (CoT) has been given.

People with CVD have a better chance of surviving and thriving because of enhanced technologies and digital methods of early identification and diagnosis. Individual differences are taken into account in personalized treatment regimens developed using data analytics and algorithms for artificial intelligence, with the goals of improving patient outcomes while reducing unwanted side effects.

The chapter comprehensively and comparatively analyzes and reviews the CVD predictions and diagnosis methods (Sections 2.2 and 2.3). It has been divided into several sections. Section 2.1 expansively gives an in-depth study of CVD with its various types and also unfolds risk factors associated with CVD. It mentions different studies related to the categories of risk factors' impact according to the geographic domains as well as a person’s demographic features, with the risk rates in various states. Section 2.2 provides a detailed description of the existing traditional approaches for CVD prediction and analysis, whereas Section 2.3 defines the current approaches such as digital and machine learning-based CVD prediction and diagnosis frameworks. Section 2.4 enlightens the overlapping-based comparative analysis of CVD prediction models.

## 2.1Introduction

CVD refers to the disordered vessels of the heart that include various associated diseases such as cerebrovascular disease, congenital heart disease, coronary heart disease, deep vein thrombosis, peripheral arterial disease, pulmonary embolism, and rheumatic heart disease. It has been estimated that more than 17.9 million patients have died from CVD at the global level [76]. This statistic includes death rates due to CVD in low-income as well as middle-income countries. The major reasons for high CVD deaths are attributed to unhealthy food habits and lack of a healthy lifestyle. The excessive consumption of unhealthy foods with a lack of fruits and vegetables and more ingestion of sodium-rich foods, fatty foods, and sugars adds to the risks of CVD. Another reason is less physical activity in daily routines. Additionally, alcohol and tobacco consumption can increase damage to the body and heart [77].

Cardiovascular illnesses, often known as CVDs, are conditions that may impede a person's ability to pump blood effectively through their cardiovascular system since they impact both the heart and the blood vessels and arteries. Certain illnesses include coronary artery disease, heart failure, stroke, hypertension, and arrhythmias are all included in the category of CVDs. They are responsible for a sizeable portion of the fatalities and impairments occurring globally, thus making them a huge threat to the public health [78].

### 2.1.1 Issues related to CVD

CVD involves a large composition of related problems that are harmful for the overall health and well-being. The broad spectrum of CVD issues include:

- Coronary Arteries Disease (CAD): The development of CAD happens when the arteries that provide blood to the heart become constricted or obstructed owing to the accumulation of cholesterol plaque. It is the most frequent kind of cardiovascular disease and the primary reason behind people suffering from heart attacks and angina attacks [79].

- Heart Failure: In heart failure, the heart can pump less blood than what the body needs, according to different investigations. It may be the consequence of several different reasons, including CAD, high blood pressure (hypertension), cardiac valve diseases, and cardiomyopathies. Heart failure may cause a variety of symptoms, including difficulty in breathing, extreme weariness, and accumulation of fluid in the body[79].
- Stroke: A stroke occurs when the brain not getting the enough and sufficient blood and there is an interruption of the blood flow to the brain, either because of a blockage (which causes an ischemic stroke) or a burst blood vessel (which causes a hemorrhagic stroke). Strokes are a common cause of brain injury, which may result in difficulties with mobility, speech, and cognition[79,80].
- Hypertension: Hypertension is a chronic disorder that is defined by higher blood pressure levels. Hypertension is also known as high blood pressure. The presence of hypertension for an extended period causes damage to the blood artery and an increase in the heart disease risk levels , stroke, and other consequences [79,81].
- Arrhythmias: Arrhythmias are abnormal heart rhythms that may emerge as irregular heartbeats, rapid or slow heart rates, or other disruptions in the heart's electrical activity. They can be either harmless or life-threatening and can influence heart functionality with disrupted blood circulation [82].
- Valvular heart disease: Valvular heart disease is characterized by anomalies or damage to the heart valves, which compromises their capacity to control the flow of blood through the heart. The common valvular abnormalities include aortic stenosis, mitral regurgitation, and mitral valve prolapse [82].
- Cardiomyopathies: Cardiomyopathies are illnesses that affect the heart muscles and may cause abnormalities in both the heart structure and functionality of the heart. They may be categorized as dilated, hypertrophic, or restricted cardiomyopathies, and can be caused by hereditary factors, infections, or other disorders that lie behind the surface [83].

Majorly, the CVD problems are attributed to various factors. Choosing to live an unhealthy lifestyle (such as smoking, eating unhealthy food, or engaging in sedentary behavior); being overweight or obese; having diabetes; having a family history of CVD; growing age; and having some specific medical problems are all risk-posing factors for CVD. Therefore, proper and regular diagnosis; risk assessment; and therapy of CVD must be performed promptly to diminish the risk of CVD complications and improvement in the health outcomes.

For the assessment of the heart function and to diagnose CVD, a variety of diagnostic instruments and examinations are being used. These include electrocardiograms, echocardiograms, stress tests, and cardiac biomarkers. For CVD prevention, several efforts can be made that include public health initiatives, education, and early identification [84]. For CVD treatment and precautions, modifications to one's lifestyle (such as a balanced diet and regular exercise), use of medicines (such as antiplatelet drugs, statins, and antihypertensives), interventional procedures (such as angioplasty and stenting), and surgical treatments (such as bypass surgery and valve repair/replacement) can be carried out.

Conclusively, CVD covers a wide variety of heart and vascular problems that have a substantial influence on the health of people all over the world. It is crucial for healthcare practitioners, policymakers, and researchers who are trying to battle these illnesses and improve cardiovascular health outcomes to have a comprehensive understanding of the many forms of CVDs, their associated risk factors, diagnostic techniques, and treatment choices.

### 2.1.2 Risk Factors Associated with CVD

Several risk factors are associated with raising CVD and thus growing the chances of heart attack. Risk factors such as habits, behaviors, and circumstances or conditions increase the chances of a heart attack. Such risk factors are classified as modifiable and non-modifiable as shown in Figure 2.1. Modifiable risk factors are considered controlled risk factors that can be modified and controlled with lifestyle changes and control on smoking, alcohol, diet, and increased physical activities. Whereas, non-modifiable risk factors are not controllable and cannot be modified such as age, family history, genetic behaviors, and ethnicity cannot be controlled.

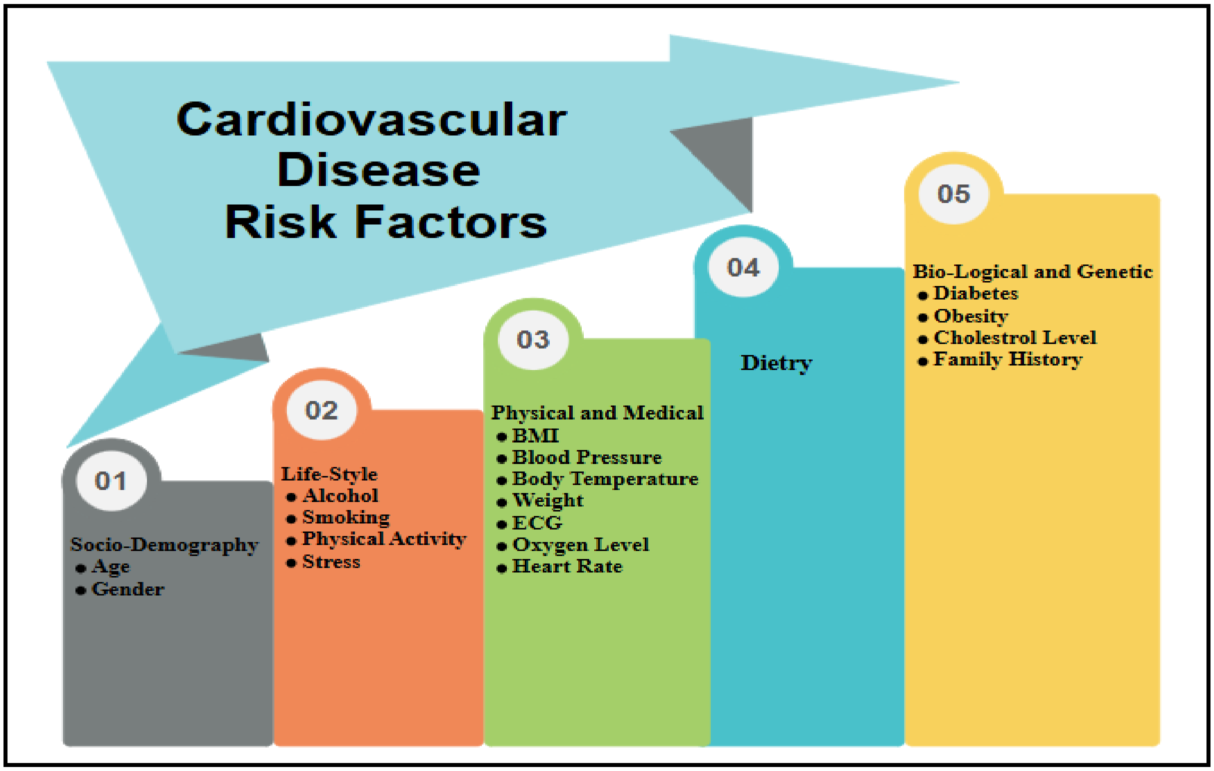



**Figure 2.1 : Associated Risk Factors**

- Socio-demographic factors are related to social and demographic aspects. They include age and gender-related to the patient. Growing age naturally affects the heart and changes the heart's blood vessels. Additionally, arteries become narrow and stiffer which raises the blood pressure level. Men are generally at high risk of CVD as compared to women. Most heart attack chances develop after age 40 and above 65 are at high risk of CVD.

- Lifestyle: Lifestyle is related to the basic habits of the patients, such as alcohol intake, smoking, physical activities, and stress, which affect the heart. Excessive alcohol intake may increase blood pressure, affect the heart, and also increase triglycerides, contributing to heart disease. Smoking damages the arteries and reduces the oxygen level of the heart. Stress and Physical inactivity increase blood pressure and weight gain.

- Physical and Medical Factors: Various physical and medical factors such as BMI, increased blood pressure, body temperature, overweight, imbalance in Oxygen level, irregular heart rate, and inaccurate ECG contribute the heart failure.
- Dietary: A proper diet is crucial for the good health. Saturated and trans fats in the diet increase the LDL cholesterol level which reduces the blood flow to the heart veins and brain which leads to heart attack. Whereas, High sodium intake increases blood pressure, and processed food high inflammation in the body
- Bio-Logical and Genetic: There are various risk factors associated with CVD that are non-modifiable such as diabetes, obesity, cholesterol level, and family history. Such Biological and genetic factors identify people with a high risk of CVD.

The studies carried out in [85-91] state that more saturated salt intake can also increase the risk level of CVD, less physical activity will increase weight and more development of hypertension. One of the studies elaborates that CVD-associated risk factors are increasing global trends such as diabetes, dyslipidemia, family history, hypertension, obesity, and smoking. These risk factors rates survey has been done in a few of the existing studies such as smoking with rates of 45.9% [92], 90% [93], 6.5%[94], and 4.7% [95].

Another risk factor such as smoking-related studies indicates that it is also one of the higher significant factors of CVD such as 43.1% [92] below age 45, 44.2% [96] in the US, and 63.02% [97] in Romania. The study[92, 94, 97, 98] states, that the diabetes rate in India (15.7%), Pakistan (18%), Romania (15.13%), and the UK (12%) is the reason for CVD. The associated factor hypertension rate is 20.2%, 32.5%, and 25% in Pakistani, Indian, and Nigerian populations respectively [92,94,95].

Table 2.1 lists the various risk factors and their range respectively. The risk factors and their range have been obtained from internationally approved medical associations and standards such as AIHW[149], AHA [150], and CDC[151].

**Table 2.1**: **Risk factors range with Risk Levels**

| Sr No | Risk Factor | Range |
|---|---|---|
| 1 | Age | • <45: **Low-Risk**<br>• Between 45 and 64: **Risk Level increased** with hypertension, dyslipidemia, insulin resistance, and obesity. Poor diet, Less Physical activity as well as stress.<br>• >65: High Risk |
| 2 | Gender | Women have risk than Men. |
| 3 | Alcohol | • High<br>• Low |
| 4 | Smoke | • Yes<br>• No |
| 5 | Physical Activity | • Yes<br>• No |
| 6 | Stress | Normal and Acute stress |
| 7 | BMI or Obesity<br>BMI = weight (kg) / height $(m)^2$ | • Underweight<br>• Normal Weight:<br>• Overweight |
| 8 | Body Temperature | • 36-38 : No risk<br>• 38.1-39: moderate risk<br>• >39 : high Risk |

| Sr No | Risk Factor | Range |
|---|---|---|
| 9 | Blood Pressure | • Normal<br>• Elevated<br>• Hypertension Stage 1 and<br>• Hypertension Stage 2 |
| 10 | Oxygen Level | • >90 : Normal, No risk<br>• 85-90: Low Risk<br>• 80-85: Moderate Risk<br>• <80: High Risk |
| 11 | Diabetes | • Less than 100 mg/dL: Normal<br>• 100-125 mg/dL: Pre-diabetes<br>• 126 mg/dL and above: Diabetes |
| 12 | Cholesterol Level | • Total Cholesterol:<br>(Total Cholesterol = LDL cholesterol + HDL cholesterol + (Triglycerides ÷ 5) )<br>• LDL Cholesterol (Low-Density Lipoprotein):<br>• HDL Cholesterol (High-Density Lipoprotein): |
| 13 | Family History | Yes: 1<br>No: 0 |

## 2.2Traditional Approaches for CVD Analysis

The conventional analyses of CVD have been performed using various methods that include the finding and management of a variety of cardiovascular disorders. These methods are based on tried-and-true approaches, procedures, and processes that have a long history of use in therapeutic settings. Even though many contemporary digital healthcare methodologies like ML and AI have garnered more attention in recent years, but the classic methods are still providing significant insights and are contributing to the CVD study.

The analysis of cardiovascular disease using more conventional methods provides helpful diagnostic, prognostic, and therapeutic insights. They serve as a basis for clinical decision-making and are the foundation upon which treatment plans are constructed. Even though more modern methods that include the various machine learning algorithms have been developed to improve CVD analysis, classical methods are still an essential component of the normal clinical examination of patients with suspected or confirmed cardiovascular illnesses. The classical approaches are dependable over long periods for monitoring heart function, identifying problems, and directing suitable therapies to enhance patient outcomes. Table 2.2 lists the various conventional methods for CVD analysis along with their advantages and limitations.

**Table 2.2*: Review of Existing Traditional CVD Analysis Methods***

| Sr No | Method | Description | Advantages | Limitations |
|---|---|---|---|---|
| 1 | Electrocardiography (ECG) [99] | Measures the electrical activity of the heart to detect abnormalities | Non-invasive, widely available | Limited sensitivity for certain conditions |
| 2 | Echocardiography [100] | Uses ultrasound waves and create visual image of the heart's structure | Non-invasive, provides detailed information | Operator-dependent, limited acoustic windows |
| 3 | Cardiac MRI [101] | Produces detailed images of the heart using magnetic fields and radio waves | Provides comprehensive assessment | Expensive, time-consuming, limited availability |

| Sr No | Method | Description | Advantages | Limitations |
|---|---|---|---|---|
| 4 | Nuclear Imaging [102] | Involves injection of radioactive tracers to evaluate heart function | Quantitative assessment of blood flow | Radiation exposure, limited availability |
| 5 | Stress Testing [103] | Evaluate heart function during physical exercise or pharmacological stress | Provides functional assessment | False positives/negatives, limited sensitivity |
| 6 | Coronary Angiography [104] | Uses the X-rays and contrast to visualize the coronary arteries | Gold standard for coronary artery assessment | Invasive, potential complications |
| 7 | Biomarker Analysis [105] | Measures specific molecules in blood to assess heart disease risk. | Non-invasive, early detection | Limited specificity, variations among patients |
| 8 | Machine Learning [106] | Utilizes algorithms to analyze large datasets and predict heart disease | High accuracy, automated analysis | Requires extensive training, data dependency |
| 9 | CT Angiography [107] | Utilizes X-rays and contrast dye to visualize the coronary arteries | Non-invasive, high-resolution imaging | Radiation exposure, limited availability |
| 10 | Holter Monitoring [108] | Records the heart's electrical activity over a 24-hour period | Continuous monitoring, detects arrhythmias | Limited duration of monitoring |

| Sr No | Method | Description | Advantages | Limitations |
|---|---|---|---|---|
| 11 | Blood Pressure Monitoring [109] | Measures blood pressure to assess cardiovascular health | Non-invasive, widely available | Single-point measurements, variability |
| 12 | Cardiac Biomarkers [110] | Analyzes specific molecules in the blood for heart disease assessment | Early detection, objective markers | Variability among individuals, non-specific |
| 13 | Heart Rate Variability Analysis [111] | Analyzes variations in the time interval between heartbeats | Non-invasive, assesses autonomic function | Interpretation challenges, data preprocessing |
| 14 | Genetic Testing [112] | Examines genetic mutations associated with inherited heart conditions | Identifies hereditary risks, personalized medicine | Limited availability, high cost |
| 15 | Cardiac Catheterization [113] | Involves threading a catheter into the heart to assess its function | Detailed assessment, therapeutic interventions | Invasive, potential complications |
| 16 | Cardiopulmonary Exercise Testing [114] | Measures the cardiovascular response to exercise | Assess functional capacity, disease severity | Requires patient effort, limited availability |
| 17 | Pulse Wave Analysis [115] | Analyzes arterial waveforms to assess arterial stiffness and function | Non-invasive, provides vascular insights | Interpretation challenges, device dependency |

| Sr No | Method | Description | Advantages | Limitations |
|---|---|---|---|---|
| 18 | Heart Sound Analysis [116] | Analyzes heart sounds to detect abnormalities and murmurs | Non-invasive, portable devices are available | Interpretation challenges, noise interference |

The other important aspects related to the traditional CVD analysis methods include:

- **Electrocardiography (ECG):** Electrocardiography, often known as ECG, is a non-invasive diagnostic procedure that records the electrical activity of the heart. It is a frequently used diagnostic method. Evaluation of heart rhythm, detection of arrhythmias, identification of anomalies in the conduction system, and evaluation of myocardial ischemia or infarction are all possible with the use of an ECG. It offers very helpful insight into the electrical processes that take place inside the heart [99].
- **Echocardiography**: This technique employs ultrasonic waves to generate pictures of the anatomy and function of the heart that are shown in real time. It is helpful in determining the size of the heart chambers, irregularities in wall motion, the operation of the valves, and the ejection percentage. In the process of identifying problems such as congenital heart abnormalities, heart failure, and valvular disorders, echocardiography is a very helpful diagnostic tool [100].
- **Stress testing:** Stress testing, which might include exercise stress tests or pharmacological stress tests, examines the reaction of the heart to artificially created or naturally occurring stress as well as to the effects of physical activity. It helps to diagnose coronary artery disease, evaluate exercise capacity, and establish whether or not inducible ischemia or arrhythmias are present in the patient's heart [103].
- **Cardiac Catheterization and Angiography** During cardiac catheterization, a catheter is inserted into the heart in order to monitor pressures, evaluate blood flow, and conduct angiography. Angiograms provide in-depth pictures of the coronary arteries, which enable the diagnosis of any obstructions or narrowing in those vessels.

When it comes to detecting coronary artery disease, this method is held in the highest regard.

- **Determining Blood Pressure:** Determining your blood pressure is an essential part of the study of cardiovascular disease [109]. An important contributor to CVD such as hypertension, stroke, and heart disease, is having a blood pressure reading that is too high. Monitoring blood pressure regularly is helpful for the early diagnosis, treatment, and prevention of cardiovascular diseases.
- **Cardiac Biomarkers:** Cardiac biomarkers[110] are used to assess cardiac damage, myocardial infarction, and heart failure. Some examples of cardiac biomarkers are brain natriuretic peptide (BNP), creatine kinase-MB (CK-MB), and troponin. These biomarkers contribute to the diagnosis, the categorization of risk, and the monitoring of CVD.
- **Lipid Profile Assessment:** This test determines the amounts of cholesterol and triglycerides present in the blood. Atherosclerosis and coronary artery disease may both be caused by abnormal lipid levels, notably high levels of LDL cholesterol and low levels of HDL cholesterol. Abnormal lipid levels contribute to the development of both of these conditions. Monitoring lipid profiles helps with risk assessment and provides direction for treatment approaches.
- **Bruce Protocol Test:** Bruce protocol or the modified Bruce protocol test examines the heart rate when it is subjected to physical activity. It does this by evaluating the patient's ability to exercise, their symptoms, and the changes in their electrocardiogram, which helps uncover any possible underlying heart issues.
- **Medical record analysis:** Taking a medical past history related to the disease and performing a physical examination are essential parts of the investigation of CVD. The patient's symptoms, risk factors, family history, and previous medical issues are documented by medical experts as they collect information on the patient. Signs of CVD, such as aberrant heart sounds, murmurs, and peripheral vascular abnormalities, may be identified via the process of doing a full physical examination.
- **Risk assessment tools:** An individual's risk of experiencing cardiovascular events

during a certain period may be estimated with the assistance of several risk assessment tools. Some examples of these tools are the Framingham Risk Score, ASCVD Risk Estimator, the Reynolds Risk Score. These programs take into account a variety of variables, including age, blood pressure, cholesterol levels, gender, patient history, and smoking habits among others.

## 2.3 Current Approaches for CVD Prediction and Analysis

In the 21st century, the healthcare industry has transformed into a more innovative industry in terms of services, techniques, technologies, and applications. The healthcare sector has now become a more digitally transformed healthcare model. Due to the advancements in the technologies in the healthcare market, the healthcare sector is becoming more patient-centric. Patients are being offered on-demand healthcare services. Wherein, healthcare facilities are available easily. Nowadays, about 52% of healthcare-related browsing is feasible through mobile phones, 2.7 billion people have access to mobile phones and 77% of patients use smartphones to book appointments to get better healthcare offerings [117].

### 2.3.1 Digital Approaches for CVD Prediction and Analysis

The field of medicine is now in the midst of a technological upheaval, which is being propelled by recent developments in robotics and other cutting-edge technologies. The delivery of healthcare is being revolutionized by these technologies, which are also enhancing the results for patients and bringing about changes in medical processes. The technology is transforming the landscape of healthcare in a variety of ways, including robotic surgery, telemedicine, and diagnostics powered by Artificial Intelligence (AI) [118]. Figure 2.2 shows the latest technologies that are being used extensively in the healthcare sector for providing digital healthcare services and provisioning.

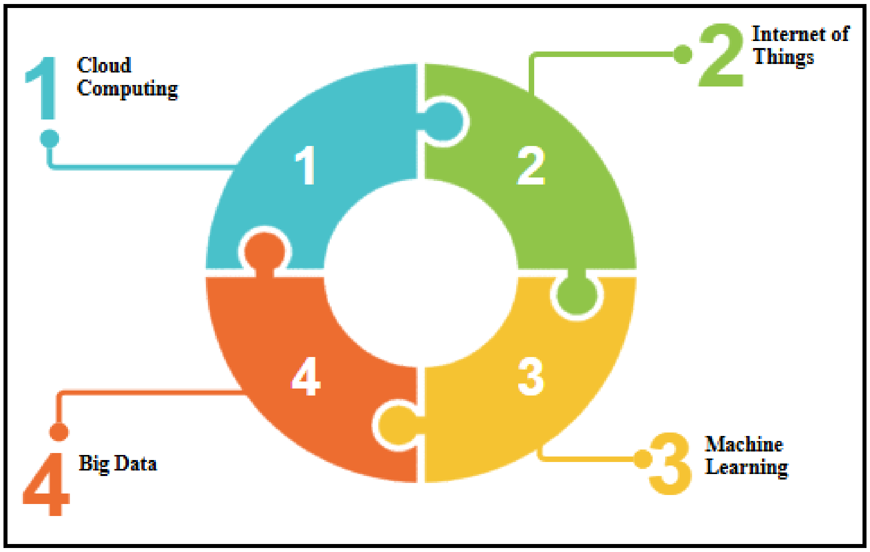



**Figure 2.2: Latest Digital Technologies**

Cloud computing is one of the vast technologies being used in every industry for better services. In the healthcare sector, with the help of cloud computing, it has become easier and safer to share patient records, automate backend operations, and also the development of mobile applications facilitating digital healthcare. The cloud facilitates everything as a service model. Thus, digital healthcare is also delivered as a service, resulting in the Healthcare-as-a-Service era. It is estimated that by the end of 2030, almost 95% of applications will be embedded with cloud computing [119]. In the healthcare sector, data is almost transmitted, stored, and retrieved electronically with the help of digital technologies. The various healthcare components are integrated, such as [120,121] EMR, computerized physician detail entry, telemedicine, smart card, EHR, digital images, and routine data backups [122]. The handling of large amounts of data is more challenging in the healthcare sector. To deal with such issues, cloud computing offers Storage-as-a-Service (SaaS). Cloud computing also offers better opportunities for the doctors as well as hospitals for patient engagement and accessibility of records at the global level.

IoT defines the interconnection of billions of devices throughout the internet that are collecting and sharing patient datasets & samples. Specifically, IoT provides the provision of implantation of various devices ranging from sensors to smartphones and wearable devices with the patient's body. When these devices are connected with an automated system it becomes easy to collect the required information, analyze, and take action to provide better treatment to the patient in real-time. When IoT is integrated with healthcare, there are more potential to develop healthcare standards that are beneficial for investigating the work to be done, remote service offers, efficient time management, and financial gains[123]. IoT provides real-time patient monitoring, telemedicine, and diagnosis services with the help of Biosensors, robotic sensors, and IoT devices. As per the latest survey, it has been estimated that 99% of organizations will adopt IoT-based technologies by 2025. IoT devices were USD 28.42 Billion in 2015 at market level whereas it is expected to increase in 2025 by USD 337.41 billion as per their betterment. [124].

Recently, CoT has emerged as a significant paradigm amalgamating the capabilities of both cloud and IoT. CoT is one of the high-performance application areas for managing, monitoring, and controlling IoT-enabled devices.

Big data supports the healthcare sector with the use of descriptive, prescriptive, and predictive analytics to get insight into the data [125]. Big data in the healthcare sector supports threefold [125] augmented digital healthcare sector such as firstly, the improvement of the clinical outcome with patient datasets and samples, secondly, workforce productivity is boosted by leverage datasets and samples and lastly, Healthcare financial data is used to improve the revenue stream of hospitals, organizations, etc. As we know, complex and massive data is time-consuming and expensive to analyze. Big data helps health professionals to drive and provide the decisions and solutions for better treatment.

In the healthcare sector, ML is a growing field of research with many potential applications such as pattern analysis, image processing, and medical record garnering for probing and analysis. ML-based algorithms enable machines to learn from insights to make predictions and decisions without human interventions, which has immense potential to revolutionize the various aspects of healthcare. The most significant application of ML in the healthcare sector is medical diagnostics. ML analyzes huge amounts of patient-related medical data that includes medical images such as MRI, X-ray, EHR, and genetic information and helps to identify patterns and detect diseases with a high degree of accuracy.

Several proposed ML models have been used for the prediction and diagnosis of the early signs of medical conditions such as cancer, heart disease, and diabetic retinopathy. The major benefit of such models is that they outperform human clinicians in terms of speed and accuracy. Furthermore, ML also helps to the improvement of operational efficiency in healthcare settings.

### 2.3.2 Existing ML Models for CVD Identification and Prediction

Machine learning is considered as one of the recent and significant digital paradigms used in CVD prediction and analysis [148]. ML has momentous advancement for the identification and prediction of CVD. ML helps to leverage patient data, such as EHR, medical imaging (MRI, X-ray), and genetic information (family history), to predict the likelihood of developing CVD and to identify the presence and absence of these conditions with high accuracy.

The researchers have proposed various models and frameworks for CVD prediction based on various associated risk factors. For the validation of these ML-based proposed models, researchers have performed the analysis on a variety of existing datasets, and different approaches have been devised. Table 2.3 states the comparison of the existing ML models for CVD prediction in terms of objective, novel approach, and accuracy.

**Table 2.3 Comparative Analysis of Existing ML-based CVD Prediction Models**

| Sr No | Ref no | Existing Models | Objective | Novel Approach | Accuracy Level | Dataset |
|---|---|---|---|---|---|---|
| 1 | [126] | HealthCloud | A proposed model is used for monitoring of heart status using a machine-learning approach. | SVM, KNN, Neural Networks, LR and Gradient Boosting Trees | 96% | UCI Machine Learning Heart Disease Dataset |
| 2 | [127] | Machine Learning based Cardiovascular Disease Diagnosis framework | A model is proposed for heart disease levels using an ensemble approach as well as boosting methods. | LR and KNN classifiers | 95.5% | Framingham, Heart Disease and Cleveland Heart Dataset. |
| 3 | [128] | Fog-Based Smart Cardiovascular Disease Prediction System | A proposed smart healthcare system diagnoses heart diseases such as CVD using deep learning. | Data Not available | | |
| 4 | [129] | Stacking model with | The prediction model predicts the | Random Forests, Linear | 96% | Cleveland, Hungarian, |

| Sr No | Ref no | Existing Models | Objective | Novel Approach | Accuracy Level | Dataset |
|---|---|---|---|---|---|---|
| | | base learner layer and meta learner layer. | CVD level using a machine learning approach. | Regression, Multilayer perceptron, ET, and CatBoost | | Swizerlang, Long Beach VA and Stalog (Heart) |
| 5 | [130] | Classification Model | The proposed model is used to predict the heart disease risk predictions using classifiers. | SVM, Gaussian Naive Bayes, LR, LightGBM, XGBoost, and random forest | 88.5% | Cleveland heart disease (HD) dataset |
| 6 | [131] | Enhanced prediction model | The enhanced model is used for the detection and prevention of the CVD using classification. | Random Forest | 99% | Kaggle Heart Disease Dataset |
| 7 | [132] | Cardiovascular prediction model | The proposed model predicts the CVD and helps to reduce the fatality. | Decision Tree, multilayer perceptron,Random Forest and XGBoost (XGB) | 87.28% | Kaggle Heart Disease Dataset |

| Sr No | Ref no | Existing Models | Objective | Novel Approach | Accuracy Level | Dataset |
|---|---|---|---|---|---|---|
| 8 | [133] | Ensembles Model | The model is proposed to predict the CVD. | Data Not available | | |
| 9 | [134] | Intelligent heart disease prediction system | The proposed model is used to predict the occurrence of the heart disease. | SVM, J48, Naïve Bayes, MLP, random forest, AdaBoost, boosted tree, and binary Discriminant | 87.91%. | Heart Disease Dataset |
| 1 | [135] | An Android-based prototype software | This prototype predicts the associated risk factors with coronary heart disease | Data Not available | | |
| 1 | [136] | Intelligent cloud-based heart disease prediction system | The system is proposed for heart disease predictions. | Support vector machine | 93.33% | Cleveland, Hungary, Switzerland, and the VA Long Beach |
| 1 | [137] | IoT based m- | The model is | DT, KNN, NB | 92.8% | UCI dataset |

| Sr No | Ref no | Existing Models | Objective | Novel Approach | Accuracy Level | Dataset |
|---|---|---|---|---|---|---|
| | | health disease diagnosing system | proposed to predict and diagnose the various diseases severity levels as well as alert system is incorporated. | and SVM | | and medical sensors |
| 1 | [138] | IoT based patient monitoring system | The proposed system is used to diagnose and predict the stroke level in the patient as well as giving the alarm alert for the patient as well as caretakers. | Random Forest, KNN, Naïve Bayes, Decision Tree | 93% | Real time patient Dataset |
| 1 | [139] | Predictive analytics Model | The proposed model predicts the heart disease with its probability levels. | Naïve Bayes | 97% | UCI machine |

| Sr No | Ref no | Existing Models | Objective | Novel Approach | Accuracy Level | Dataset |
|---|---|---|---|---|---|---|
| 1 | [141] | Hybrid Random Forest with Linear model (HRFLM) | The model is proposed to predict the Cardiovascular disease using ML and IoT. | ANN, Decision Tree | 88.7% | UCI Cleveland dataset |
| 1 | [142] | Machine learning based Flask Web framework | The framework identified the different stages to predict the absence and presence of the CVD. | Random forest | | Kaggle Dataset |
| 1 | [143] | Adaptive inference system | The model is proposed for the prediction of heart disease as well as other multiple diseases using BSO algorithm. | neural network, SVM | 96.08% | Framingham and the dataset Hungarian |
| 1 | [144] | Heart disease prediction model | The model is proposed to diagnose the CAD disease based on different dataset. | decision tree and pruned C4.5 tree | 78.06% | UCI dataset |

| Sr No | Ref no | Existing Models | Objective | Novel Approach | Accuracy Level | Dataset |
|---|---|---|---|---|---|---|
| 1 | [145] | CRISP-DM | The data mining based model is proposed to predict and classify the CVD. | Random forest, LR and ANN | 80% | UCI dataset |
| 2 | [146] | Heart Prediction model | The machine learning-based model is proposed to predict and analyses the probability of heart disease development rate in the patients. | Naïve Bayes, decision tree, KNN, and random forest algorithm | 90.7% | Cleveland Dataset |
| 2 | [147] | An intelligent decision support system | A system is proposed to predict the presence and absence of CVD. | LR, KNN ANN, SVM, decision tree, Naive Bayes, and random forest | | Cleveland heart disease dataset |

## 2.4 Comparative Analysis of Existing CVD Prediction Models

The comparative analysis of existing CVD prediction models is important to classify and understand them based on various technologies. This section comprehensively covers the comparative study of the existing CVD prediction models based on different metrics such as technological platform based upon and the datasets applied. It highlights the overlapping of various prediction model implementations on multiple platforms.

### 2.4.1 Technological Platform-based Overlapping

Various digital methodologies have been used to propose the CVD prediction models which are categorized as depicted in the following overlapping diagram as shown in Figure 2.3. Proposed CVD prediction frameworks based on cloud computing[121,122,135], IoT [123], and machine learning [106,109,126-127,129-133,139,141,142,145,146,148] are analyses and survey has been performed.

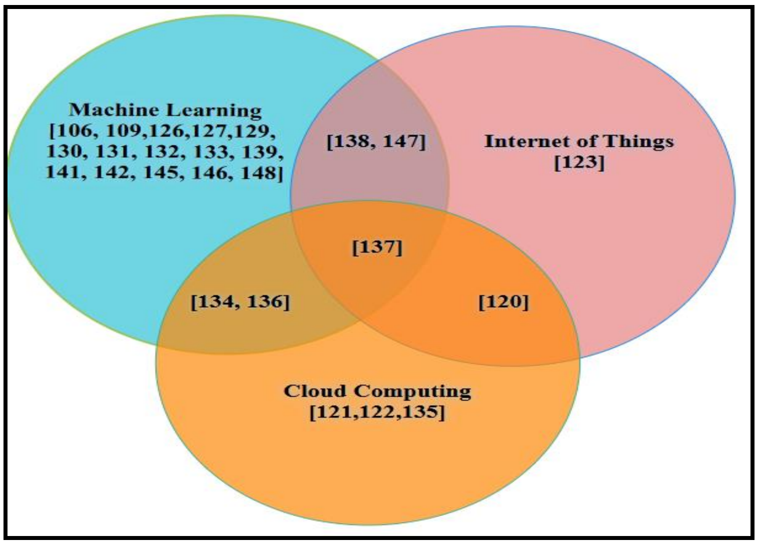



**Figure 2.3: Overlapping of technologies used for CVD Prediction**

### 2.4.2 Dataset Applied-based Comparison

There are various dataset used to proposed the CVD prediction and recommendation models which are categorized as depicted in following diagram as show in figure 2.4. Various dataset used are UCI heart disease dataset, Framingham heart disease, Cleveland heart dataset, Hungarian, and Swizerlang, Long Beach VA and Stalog (Heart), Kaggle Heart Disease Dataset as well as Real time patient Dataset.

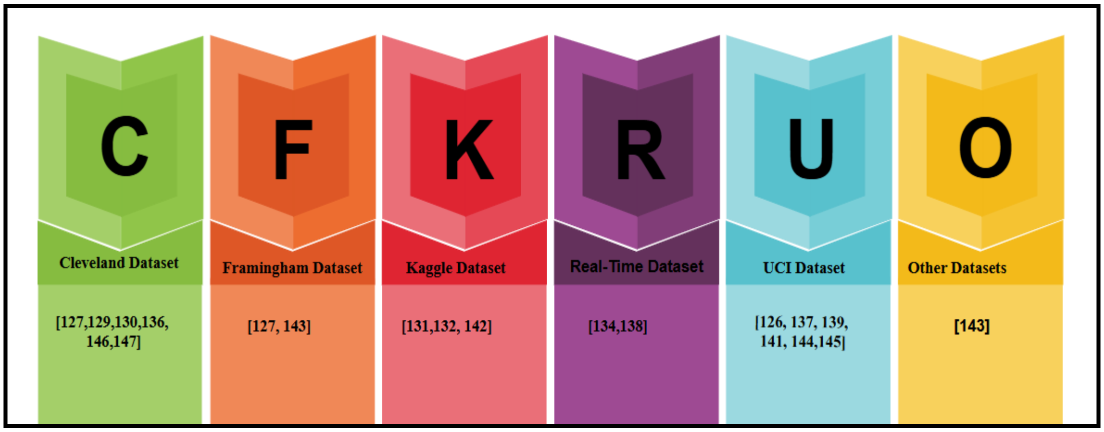



**Figure 2.4: Dataset used for CVD Prediction**

### 2.4.3 Critical Evaluation of Past Studies

Many research efforts have investigated the use of machine learning for predicting cardiovascular disease (CVD), but several methodological shortcomings remain largely unaddressed. A common issue is the reliance on small or imbalanced datasets, which limits the applicability of findings to broader, more diverse patient populations. In studies that employ ensemble methods, the rationale for choosing specific base classifiers is often unclear, raising concerns about redundancy and potential bias in predictions. Furthermore, practical considerations such as model interpretability, computational demands, and ease of integration into existing clinical workflows are frequently overlooked. While metrics like accuracy and AUC are regularly reported, they do not always reflect clinical relevance—particularly when high false positive rates could result in unnecessary tests or treatments. These concerns underscore the importance of adopting a more comprehensive and practically informed approach when developing predictive models for CVD.

## 2.5 Conclusion

Conclusively, in the literature review, CVD analysis techniques that are currently in use have been investigated and their benefits and drawbacks. It emphasizes the role that digital technology plays in healthcare, namely in medical imaging, illness diagnosis, drug development, personalized medicine, health monitoring, natural language processing

(NLP), and disease monitoring. The issues that are currently faced in CVD analysis have been discussed in this chapter. These include data availability and quality, data integration, the interpretability of machine learning models, ethical concerns, bias, clinical acceptance, validation, and resource limits. In addition to this, this chapter gives an overview of conventional methods for analyzing CVD and demonstrates the relevance of these methods in terms of clinical decision-making and patient treatment. It highlights how important it is to integrate classical learning methods with deep learning strategies in order to capitalize on their own strengths and enhance the results of CVD analyses.

# CHAPTER 3

# PCARD: AN INTELLIGENT FRAMEWORK FOR CARDIOVASCULAR DISEASE PREDICTION

Cardiovascular disease incorporates a variety of conditions or problems that affect the vessels of the heart and blood of the human body. It has become the leading cause of death globally, contributing to significant morbidity and mortality. CVD problems are complex and multifaceted, requiring comprehensive prevention, early detection, and management approaches. Adopting a heart-healthy lifestyle and working closely with healthcare providers can significantly reduce the risk and impact of CVD problems. Various tools, techniques, and methodologies identify the presence or risk of developing cardiovascular conditions. Such CVD prediction and detection tools, techniques, and methodologies include traditional clinical assessments, imaging techniques, and emerging computational approaches that have been integrated with and into various models. These proposed CVD prediction models are based on the latest IT innovations incorporated with machine learning (ML) and artificial intelligence (AI).

This chapter proposes a model, known as PCard for the prediction of the prospective occurrence and the probability level of the presence and absence of CVD using different machine learning algorithms such as k-nearest Neighbors, Support Vector Machine, Logistic Regression, Multi-Layer Perceptron, and Deep Forest. A recommender module has been integrated with the PCard model to provide valuable recommendations for patient diagnosis based on the level of CVD. The model has been trained using the CVD dataset available at the Kaggle repository and the experimental results indicate that the model achieves an accuracy of 98.5% compared to individual machine-learning approaches.

The chapter has been divided into 3 sections. Section 3.1 elaborates on the basics of the ML approach and proposed model. Section 3.2 elaborates the ensemble approach in detail with its techniques and methodologies. Section 3.3 enlightens about the proposed framework used for CVD prediction using machine learning algorithms in terms of the layered approach, its components, working in detail, and proposed algorithms.

## 3.1 Introduction

Machine learning has arisen as an advanced approach in healthcare offerings and is being used to detect CVD. Machine learning is a part of data science used for data analysis based on features to predict the outcomes and make decisions to achieve the objectives. In the healthcare field, machine learning helps medical professionals to provide better treatment facilities . It helps to collect patient data easily; manage the dataset; analyze the data; predict various health problems and offer relevant recommendations[152].

The chapter proposes a CVD prediction model that uses ML approaches for the prediction of CVD problems. It is based on an ensemble machine learning which is used to detect the absence and presence of CVD as well as predict the risk levels associated with it. The proposed model not only helps to diagnose CVD but also facilitates the level-based segregation for CVD risks such as low risk, medium risk, or high risk.

This chapter proposes a model, PCard, that is used to predict the prospective occurrence and the probability of the presence and absence of CVD using different machine learning algorithms. It uses k-Nearest Neighbors (KNN), Logistic regression (LR), Multi-Layer Perceptron (MLP), random forest (RF), and Support vector machine (SVM) for the CVD prediction. A recommendation component has been integrated with a prediction system in PCard where the patients are provided with valuable recommendations for their diagnosis based on the level of CVD predictions. The model has been trained using the CVD dataset from the Kaggle repository and the experimental results indicate that the model achieves an accuracy of 98.5% compared to individual machine-learning approaches.

Additionally, the proposed model is also centered on the development of an online CVD analysis model for better CVD diagnostics as well as the potential of ML algorithms in real-time. The purpose of using ML algorithms is for the ensemble of classifiers to utilize the capabilities of each individual method and produce a diagnostic mechanism that is more reliable and accurate. The model involves data preprocessing, feature selection, and an ensemble of various ML algorithms for CVD prediction [29]. Various ML algorithms have been used in the proposed model to analyze the result as well as performance parameters. Figure 3.1 shows the high-level view of the PCard system.

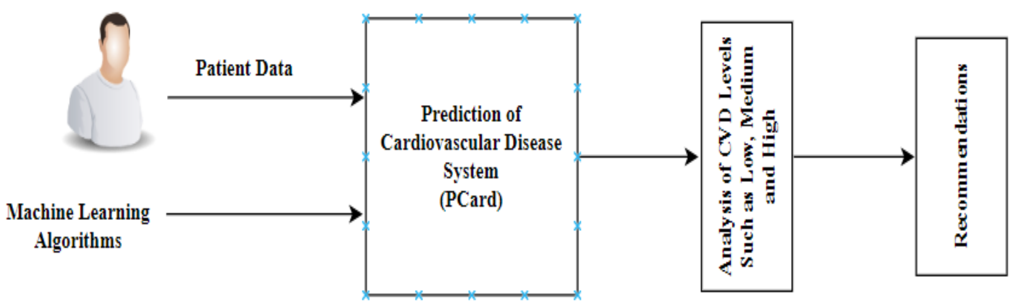



**Figure 3.1 High-level PCard Block Diagram**

## 3.2 Basic Ensemble Approach

The ensemble approach is considered as machine learning approach that is used to combine various models to propose a prediction model. The ensemble is how multiple techniques are combined to improve the accuracy as well as performance of the model [69]. The ensemble approach is categorized into two groups as shown in Figure 3.2:

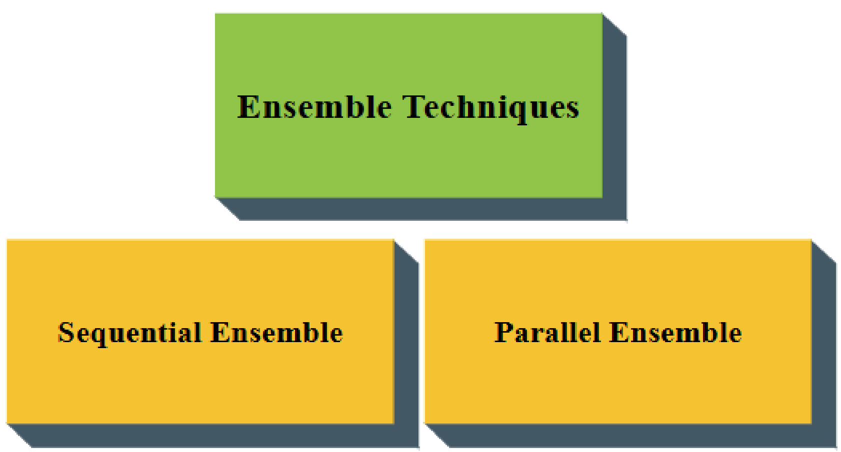



**Figure 3.2 Ensemble Techniques categories**

The sequential ensemble is the method to generate the base learners in sequence whereas in the parallel ensemble approach the base learners are being generated in a parallel way. In the sequential approach, the base learners are generated sequentially and the generation of base learners promotes the dependency between base learners however, in the parallel approach, the base learners are generated parallel and the parallel generation of base learners promotes the dependency between base learners.

### 3.2.1 Types of Ensemble Learning

Ensemble learning is also considered as the utilization of ensemble methods for the creation of more accurate classifiers by the combination of different classifiers formulating an ensemble model. The main goal of an ensemble model is that each classifier used in an ensemble model should be such that they should complement each other for a decision so that more accuracy can be achieved [153]. All the selected classifiers should then agree with different decisions. However, if all the classifiers agree with the same decisions, then the designed ensemble technique fails. Moreover, if there occurs an error with any classifier then the remaining classifiers can help to handle these errors. The following Figure 3.3 shows the working of the ensemble approach.

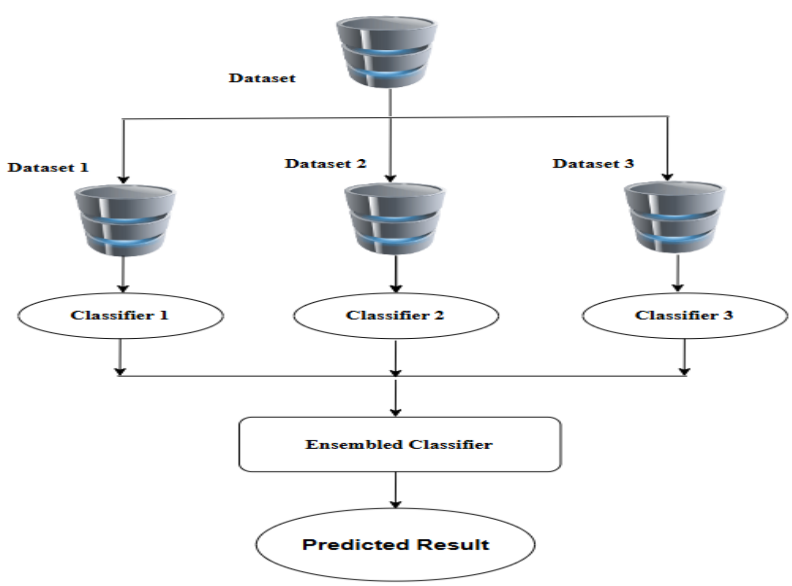



**Figure 3.3 Ensemble Approach for Classification**

Ensemble learning is classified into three main categories [153] which are as follows:

- Bagging: Bagging is an ensemble learning approach that is used to improve the performance as well as the accuracy of the ML algorithms. Bagging helps in avoiding the overfitting of the dataset. In this ensemble learning approach, the sample is randomly chosen from the dataset with its replacement, and every data subset is used for training purpose due to which an ensemble of different classifiers is obtained using average methods.
- Boosting: Boosting is an ensemble learning approach in which weak learners are converted into strong learners using weighted average methods and also combines different classifier base models. Each classifier's output works and is combined independently without any preference for any individual classifiers.
- Stacking: Stacking is an ensemble approach in which all the classifier models are aggregated based on their weights to propose a prediction model.

## 3.3Proposed PCard Model

Heart disease and associated problems are a major health concern worldwide, which significantly causes morbidity and mortality. Based on the need for CVD detection and prediction, a layered architectural model, the PCard model, has been proposed which intends to predict the presence of CVD in the patients and also analyses the risk level associated with the presence. PCard adopts a layered architectural layout to segregate the complexity and inhibit the abstraction of the tasks at different levels. The PCard model collects the patient data and uses ML algorithms to detect the CVD levels such as low, high, and medium levels. Based on the risk level, various recommendations are given to the patient. So, PCard involves diagnosis and recommendations.

The proposed PCard model is based on certain assumptions that form its basis of prediction and recommendation methodology. The assumptions enable effective prediction results in terms of accuracy and data considerations. For suitable predictions, the verification and validation of assumptions is required. The proposed PCard model has associated assumptions that include:

- The proposed PCard model is a multivariable prediction model that considers the multiple factors that collectively impact CVD occurrence and risk assessment. Thereby, the multiple risk factors are analyzed to predict the absence and presence of CVD.
- There are multiple variables considered for CVD occurrence and risk assessment that include modifiable and non-modifiable factors such as age, alcohol intake, cholesterol, diastolic blood pressure, gender, glucose, height, weight, physical activity, smoking, and systolic blood pressure.
- The model assumes that the geographical location and model evaluation environment can be different and independent. As a geographical model, patient data can be collected from any of the locations at the global level either using GUI, Google Maps, or any online repository. The proposed model should have the feature of data analysis at any level.
- The model assumes that the predictor outcomes associations may vary among ethnic groups. However, the proposed model follows internationally approved standards and parameters formulated upon extensive analysis and tests across the global population on various healthcare issues (particularly CVD problems).
- The proposed model is a layered model to create a separation of concerns. The implementation of different technologies, devices, and operations is separate from various layers, and the results of each layer are integrated collaboratively into the prediction and recommendation unit.

### 3.3.1 Layered PCard Architectural Model

PCard model is a layered architecture model that follows the bottom-up (machine learning) approach. In a layered architecture, the PCard model consists of 3 layers, namely the user and cloud layer, the Preprocessing layer, and the Prediction and Recommendation layer. These layers perform functions independent of one another. A detailed description of the layered architecture is discussed and the shown in the following Figure 3.4.

a) ***User and Cloud Layer***

The main purpose of this layer is the patient's data collection, storage, and management. It acts as an interface between the hospital professionals and patients that helps in data acquisition and communication. Sensors are being used for real-time monitoring of the patients is performed and health records are stored in a cloud database that can be accessed and monitored at the global level. At this layer, patient history can also be accessed for the prediction analysis. The collected data is not in the proper format, so it is passed to the next layer for preprocessing. The main components of the layers are user profiles, medical history, sensor records, and cloud database. These components perform the following functions:

- User profiles are related to the patient's demographic and lifestyle information, such as age, gender, ethnicity, smoking habits, consumption of alcohol, and physical activities.
- Medical history is related to chronic conditions such as diabetes, hypertension, and other related diagnoses.
- Data collection using wearable or medical devices that enable real-time health monitoring.
- All the data is stored and managed in the cloud infrastructure, where it is saved securely and can be made available for processing and analysis.

b) ***Preprocessing Layer***

Another layer is data preprocessing. In this layer, patient data is being processed for null values as well as duplicate values removal. Filter methods are also used in this layer that help to consider the features as labels to identify the dependent and independent variables. After preprocessing, the dataset is classified as a training dataset and a testing dataset for testing the model and to train the model. That preprocessed dataset is used in the prediction and recommendation layer. The steps involved are:

- **Handling of missing values:** Missing values are detected and filled in the dataset for more consistency and to avoid bias during the classification and prediction
- **Remove the duplicate values:** Repeated values are eliminated to ensure data integrity.

- **Data Filtering:** Filters are applied to retain the relevant data samples.

After preprocessing, the dataset is free from errors, redundancy, and inconsistencies, and is ready for training and use for the prediction model.

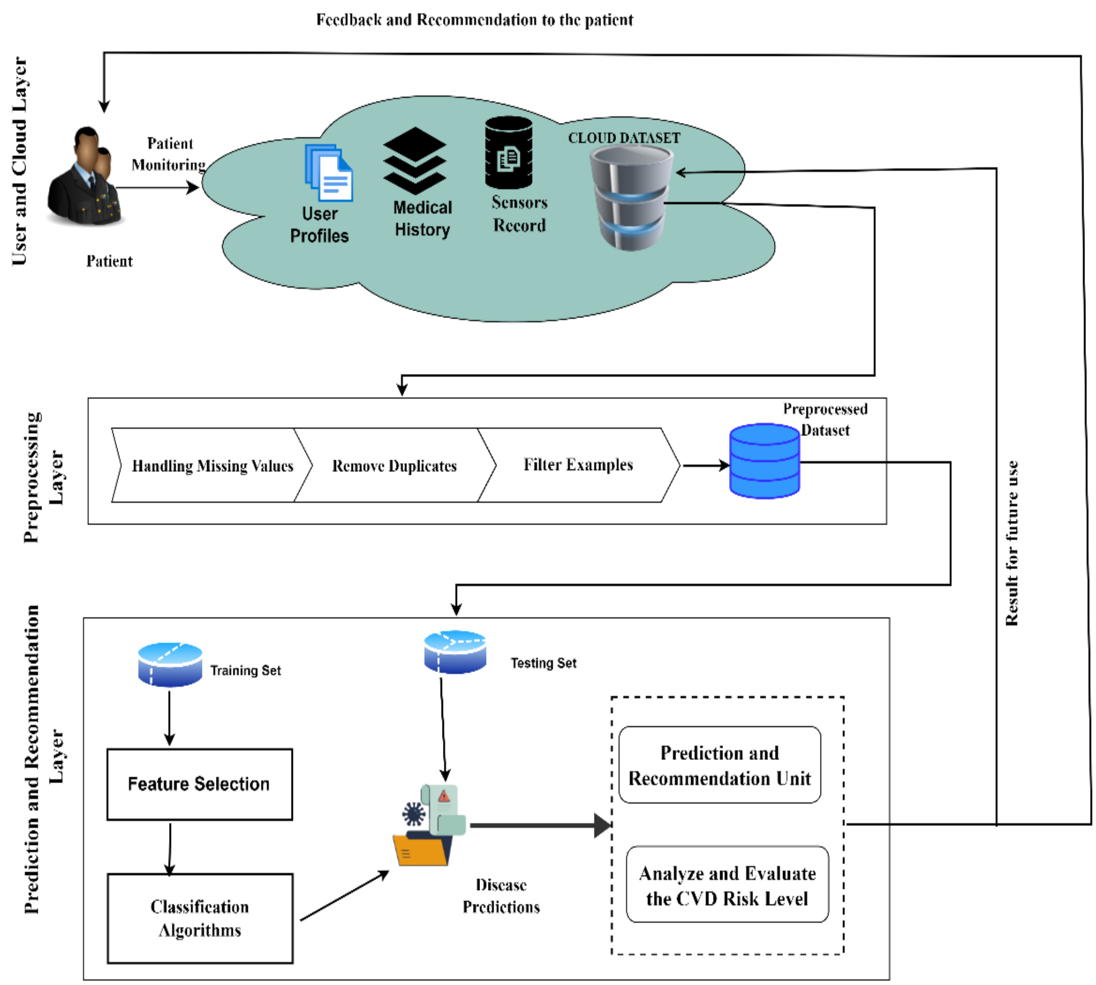



**Figure 3.4 Proposed PCard Layered Model**

c) ***Prediction and Recommendation Layer***: In this layer, the preprocessed dataset is split into training and testing datasets. The training set is used to train that model using machine learning algorithms and the testing set is used to test the performance of the model by considering various parameters such as accuracy, F1 score, Precision, recall, and AUC-ROC. The layer builds and applies the proposed ML model for the risk levels prediction of the CVD that generate the health advice. It includes:

- **Feature Selection:** Identification of the relevant attributes such as age, cholesterol level, blood pressure and lifestyle factors that helps in contributing the CVD risk levels. It reduces the dimensions for the improvement in the the model performance.
- **Classification Algorithms:** ML models is trained using the classification algorithm on the training dataset and learn the patterns and association for the known outcomes such as presence and absence of the CVD.

It performs the disease prediction based on the input features. These input features are prioritized as CVD risks as low, moderate and high and consequently personalized recommendations and the feedback to the patients are provided accordingly.

### 3.3.2 Components of PCard Model

Each layer of the proposed model is integrated with various components that are briefly discussed below:

- ***Patients***: Patients are an important entity of the proposed model. With the patient body, the sensors are attached inside or outside of the body for real-time health monitoring. The collected data is used for the CVD prediction.
- ***Doctors:*** Doctors play a crucial role in providing the healthcare services. They have vast knowledge about medicine that helps with diagnosis, detecting the symptoms, providing treatment, and monitoring the patient's health.
- ***Sensors:*** Sensors are connected to the human body to analyze different parameters such as Age, Alcohol intake, Cholesterol, Diastolic blood pressure, Gender, Glucose, Height, Weight, Smoking, Systolic blood pressure, and Physical Activity. The various sensors used for analysis of CVD parameters are discussed in Table 3.1:

**Table 3.1 Sensors Used for Analysis of Various CVD Parameters in PCard Model**

| Sr no | Parameter | Sensors | Description |
|---|---|---|---|
| 1 | Height | Laser Sensor | It is used to measure the height of the patient. |
| 2 | Weight | Weight sensor | It is used to measure the weight of the patient. |
| 3 | Systolic or Diastolic blood pressure | Sphygmomanometer | A blood pressure measuring device that uses an inflated cuff that collapses and then carefully releases the artery under the cuff, which is in a controlled manner. |
| 4 | Cholesterol | LipidPlus® Cholesterol Monitor | This device is used to measure the total cholesterol, high-density lipoprotein, low-density lipoprotein levels in the patient body. |
| 5 | Glucose | Glucometer | The device is used to check how much sugar is present in patient's blood. |

- ***Cloud Database:*** Cloud databases are crucial and disruptive innovations that can be used at the global level for data storage and by healthcare professionals for health analysis and treatment facilities.
- ***Analysis Tools***: The Pcard model is implemented in the Spyder tool. Spyder is an open-source, scientific platform for Python. Engineers and scientists use the Spyder tool for data analysis. It consists of various advanced features, including scientific packages such as numpy, Scipy, pandas, matplotlib, and more. This tool makes it easy to predict CVD and provides the model's performance.

### 3.3.3 Working of Proposed PCard Model

A PCard model is integrated with the healthcare system for patient health monitoring. It provides a way for CVD prediction using an ensemble of machine learning algorithms and provides recommendations to the patients for better treatment facilities. The proposed model uses a hybrid approach such as a top-down and bottom-up approach to predict the CVD proximity risk level in patients. The top-down approach follows the various steps from data collection, data preprocessing, data splitting, and CVD prediction with its severity levels whereas in the bottom-up approach, feedback and recommendations are given to the patients, and analysis is saved in the cloud for future use. The following Figure 3.5 shows the working of the proposed model.

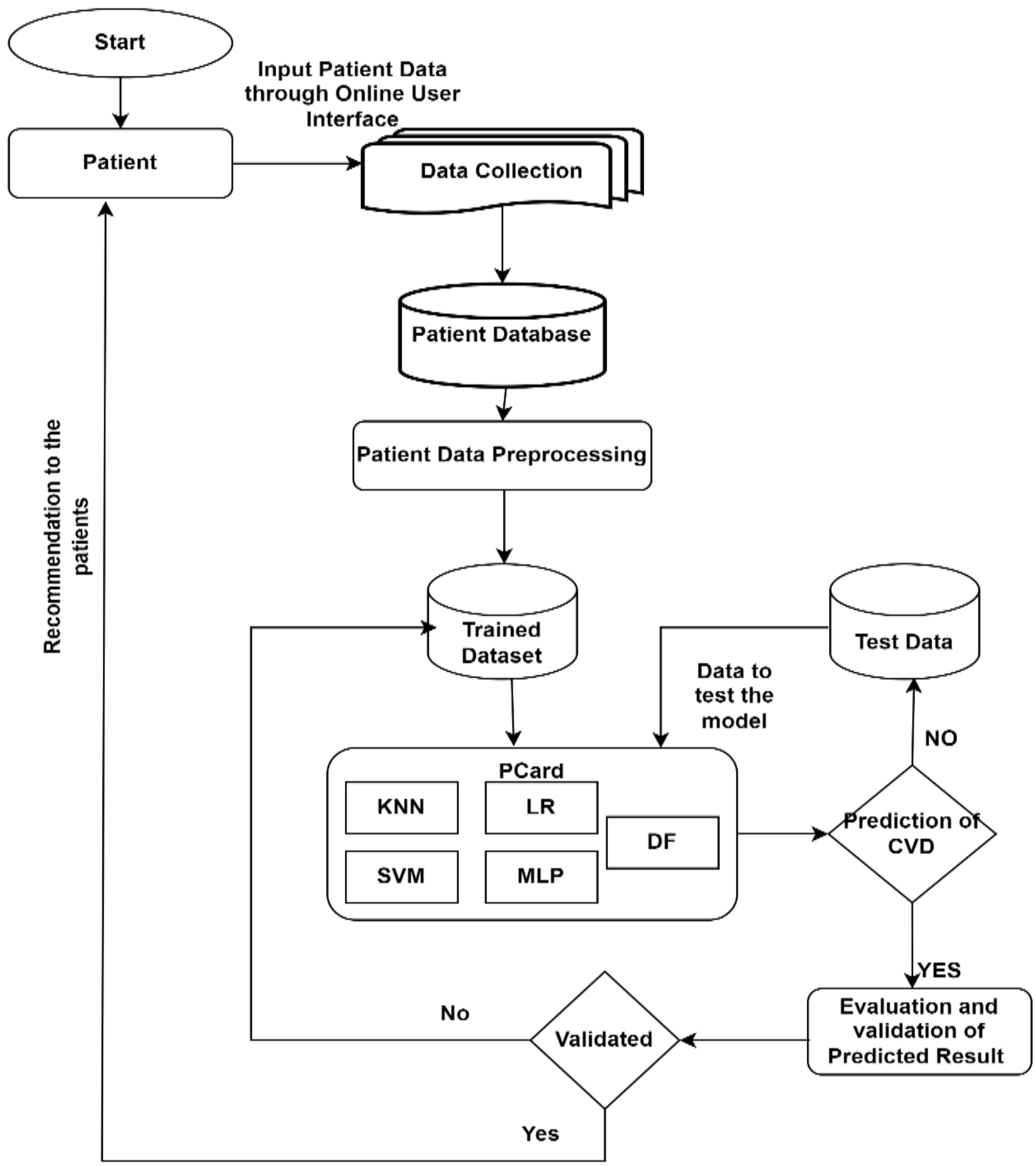



**Figure 3.5 Flow Diagram of PCard Model**

PCard model has phases such as data collection from GUI as well as Kaggle data source, data preprocessing such as removal of irrelevant data, filling of NULL values and classification of CVD levels based on occurrence of the different parameters using machine learning algorithms. The PCard model is proposed for the prediction of CVD which is advantageous for healthcare field and patients throughout the world. To achieve the objective, various machine learning algorithms are employed as ensemble technique on the dataset and results are presented in this report. This section highlights the in detail description of every layer with the used methodology. The various phases of this model are as following:

### a) Data Collection

In this work, two methods have been used for the data collection. First, a cloud interface has been deployed that can be used at global level and another is Kaggle as a data source is used.

**Cloud UI and Cloud of Things Service**

To facilitate efficient operations and provide a cloud of things service to patients, a cloud user interface (UI) has been developed. The cloud UI allows authorized healthcare professionals to access and analyze patient data securely from any device with an internet connection. It provides a user-friendly interface for visualizing and interpreting the results of the heart disease analysis.

Furthermore, the cloud UI enables the integration of a cloud of things service for heart disease management. This service utilizes Internet of Things (IoT) devices, such as wearable sensors and monitoring equipment, to collect real-time data from patients. The collected data can be securely transmitted to the cloud UI, where it is analyzed using the heart disease classification model. This makes it possible to monitor patients' health state continuously, identify any problems early, implement tailored therapies, and personalized interventions.

**Kaggle Data Source**

The Kaggle dataset used for this analysis that is a popular platform of data science competitions. It consists of various features that can help in the diagnosis of heart disease, including demographic information, medical test results, and lifestyle factors. The dataset has undergone pre-processing, which includes scaling operations to ensure uniformity and comparability among the features.

The dataset cardio_train [154]. used in this study consists of 70,000 records of the patients and 12 unique features. The following Table 3.2 describes the details of features with range values

**Table 3.2 Dataset Features**

| Feature | Variable | Range |
|---|---|---|
| Age | Age | 10,798 to 23,713 (in Days) |
| Height | Height | 55-250 (in cm) |
| Weight | Weight | 10-200 (in kg) |
| Gender | Gender | 1 - women, 2 – men |
| Syst blood pressure | Ap_hi | 150- 16,020 (Categorical between 0-2) |
| Diast blood pressure | Ap_lo | 70-11,000 |
| Cholesterol | Chol | Categorical Value between 0-2 |
| Glucose | Gluc | Categorical Value between 0-2 |
| Smoking Habit | Smoke | 1: Yes, 0: No |
| Alcohol intake | Alco | 1: Yes, 0: No |
| Physical Activity | Active | 1: Yes, 0: No |

**b) Data Preprocessing**

Data Preprocessing is considered as the major step for the data analysis. In this work, we have imported the StandardScaler package from sklearn. preprocessing for data scaling. The basic functions such as fit(data), transform(data), and split() used for data scaling have been discussed in this section.

The fit(data) is used to calculate the mean value and standard deviation of the features. Then transform(data) is used for data scaling by calculating the mean and std using fit() where it ensures that the mean is 0 and the standard deviation value is 1 and splits the dataset into train and test datasets. For the selection of the special features from the dataset, feature selection function is used to identify the specific the label and to increase the efficiency of the predicted model using different machine learning algorithms such as Decision tree, KNN, logistic regression, MLP and SVM. Machine learning algorithms are compatible with numerical values only so feature encoding process is used that helps to convert the categorical data into numerical values and normalization process is performed to convert the numerical values in specific scale range to apply the analysis on the dataset. To check the performance of the predicted model data splitting() is used to split the data set into training and testing sets such as 80% training and 20% as testing dataset.

**c) Classification Phase**

In the classification phase, the proposed model splits the data into training and testing datasets. The training dataset is being trained by various machine learning algorithms such as KNN, Logistic regression, Multi-Layer Perceptron, Random forest and Support vector machine. This section defines the classification algorithms used in the work as follow:

- ***Classification Algorithm***

To classify heart disease, a fusion of multiple classification algorithms has been employed. The following algorithms were used:

  - k-Nearest Neighbors (KNN): A non-parametric supervised classification approach called kNN. In this, the majority of the votes are used for N nearest neighbors in the feature space to classify a sample dataset.

- Support Vector Machine (SVM): SVM is a supervised learning ML algorithm that constructs hyperplanes to separate different classes in the feature space.
- Logistic Regression (LR): A logistic function is used in logistic regression. a linear regression framework forecasts the likelihood of a binary outcome.
- Multi-Layer Perceptron (MLP): MLP is an artificial neural network with multiple layers of nodes that can learn complex patterns in the data.
- Random Forest: In Random Forest, Several decision trees are used as an ensemble learning technique to generate predictions.

The fusion of these algorithms aims to leverage the strengths of each method and improve the overall accuracy of heart disease classification.

### 3.3.4 Model Training and Tuning Parameters

In ML, model training is the process that helps ML-based algorithms to learn the hidden patterns and predict the required output. The provided dataset is preprocessed using a preprocessing layer and then used for model training and testing. For training and testing the model, the dataset is being split into training and testing datasets using the train_test_split method, and the library used is scikit-learn. Feature values are normalized using StandardScaler in the training dataset, and multiple classification algorithms such as kNN, Random forest, SVM, LR, and MLP are used to normalize the training data. After classification, each method is applied on the test dataset to evaluate the model performance in the form of a confusion matrix, accuracy, precision, recall, and F1-score. Various tuning parameters are considered for the model as discussed in the Table 3.3 below. Finally, a confidence interval for accuracy is estimated using bootstrapping to provide a reliable measure of model performance. The following table 3.3 shows how different classification algorithms are used as classifiers by using the ‘method’ parameter in the code:

Table 3.3: Model Training and Tuning Parameters

| Method ID | Classifier | Training Parameters | Tuning Parameters |
|---|---|---|---|
| 1 | KNN | neigh1.fit(X_train, Y) | KNeighborsClassifier(n_neighbors=1) |
| 2 | Randon Forest | neigh2.fit(X_train, Y) | RandomForestClassifier(n_estimators=100, max_depth=2,random_state=0) |
| 3 | Support Vector Machine | neigh3.fit(X_train, Y) | LinearSVC(random_state=0, tol=1e-5) |
| 4 | Logistic Regression | logisticRegressionClassifier .fit(X_train, Y) | LogisticRegression(random_state=0,multi_class='auto', solver='lbfgs',max_iter=1000) |
| 5 | MLP | mlp.fit(X_train, Y) | MLPClassifier(hidden_layer_sizes=100, alpha=2, random_state=1) |

## 3.4 Proposed PCard Algorithm

The Proposed PCard algorithm categorizes patients with CVD risk levels as low, Borderline, Intermediate, and high risk. Identifying risk levels helps make recommendations to patients based on the risk score. Mathematically, the risk score is calculated by adding the weights of each attribute. Each attribute weight is considered as shown in Table 3.4.

**Table 3.4 Scoring System Table**

| Risk Factor | Range Value | Weight |
|---|---|---|
| Age | <45 Year | 0 |
| | 45-54 Years | 1 |
| | 55-64 years | 2 |
| | >64 | 3 |
| Total Cholestrol | <200 mg/dl | 0 |
| | 200-239 mg/dl | 1 |
| | >239 mg/dl | 2 |
| HDL Cholestrol | >59 mg/dl (High) | 0 |
| | 40-59 mg/dl | 1 |
| | <40 mg/dl (Low) | 2 |
| Blood Pressure | <120/80 mmHg | 0 |
| | 120/80-139/89 mmHg | 1 |
| | >139/89 | 2 |
| Diabetes | No | 0 |
| | Yes | 2 |
| Physical Activity | Active | 0 |
| | Inactive | 1 |
| Smoking | Non-smoker | 0 |

| Risk Factor | Range Value | Weight |
|---|---|---|
| | Active Smoker | 2 |
| Alcohol intake | Moderate(1-2 Drinks/day) | 0 |
| | Heavy (3+Drink/day) | 2 |
| Family History of CVD | No | 0 |
| | Yes | 1 |

**Assumption for the PCard Model:**

1. The risk level is calculated by considering the weight of each attribute, which is regarded as 0,1,2, or 3.
2. After adding all the weights, the risk score is calculated which is further used for the recommendations.
3. For CVD prediction, the risk levels are considered as Low, Borderline, Intermediate, and high risk.

**Algorithm 1: Cardiovascular Disease Prediction Process**

Input: Import the patient's Dataset (Cardio_train.csv)

Output: Patient's classification into Low risk level, medium risk level, and High Risk level of CVD. Special recommendations for patients having high-risk level of CVD.

#Step 1: Load the Dataset

#Step 2:Perform preprocessing of the loaded dataset.

#Step 3: Split the dataset as training and testing datasets.

#Step 4: Use the pre-processed data as a training dataset. The training dataset is used to train the model for generations of the rules.

Use CVD_Pred(Pti , Rfj, Levij → CVD) function to predict the patient having CVD. (Algorithm 2)

#Step 5: Evaluate the Model to check the performance. The split testing dataset is used for the performance of the prediction using a trained model.

#Step 6: The result being used by the cardiologist for the recommendations based on the levels.

**Algorithm 2: Calculation of CVD prediction levels**

**Objective function: CVD_Pred($Pt_i$ , $Rf_j$, $Lev_{ij}$ →CVD)**

where: Pt :Patients,

$R_{at}$: Risk attributes,

$Lev_{ij}$ Risk level of $j^{th}$ value associated with $i^{th}$ patient and

CVD: Cardiovascular Disease

$i$: *index of the Patients*

$i \epsilon S$ *where* $S = \{1 \leq i \leq n\}$ S= Set of patients and n is the maximum number of patients

J:Index for Risk attributes where $j \in R$ *where* $R = \{1 \leq j \leq m\}$

$R_a$=Set of risk attributes and a is the maximum number of risk attributes

$Lev_{ij}$ : Value of each factor such as $i \in S$ *and* $j \in R$

*where* $S = \{1 \leq i \leq n\}$ *and* $R = \{1 \leq j \leq m\}$

**Calculate the risk level Using Risk Score table 3.3 and function 1**

Risk_Score=calculate_cvd_risk_score($R_{at1}$, $R_{at2}$,$R_{at3}$ ,$R_{at4}$, $R_{at5}$,…………..,$R_{atn}$)

If(Risk_Score<5%) then Low-CVD Risk

else if(Risk_Score>5% AND Risk_Scorel<8%) then Borderline risk

else if (Risk_Score >8 AND Risk_Score<20) then Intermediate Risk

else High Risk of CVD

**#Function 1 Definition**

calculate_cvd_risk_score($R_{at1}$, $R_{at2}$,$R_{at3}$ ,$R_{at4}$, $R_{at5}$,…………..,$R_{atn}$)

Risk_Score = $\sum R_{a*w}$

return Risk_Score

CVD_Pred() is used for prediction of the risk level of CVD in a patient by considering the different parameters and their range. Here, the CVD_Pred is the function that maps the $i^{th}$ patient into the different risk levels such as high, low, and medium by considering the different range of risk factors.

The proposed model is consist of set S={pat1,pat2,pat2………patn} of n patients $i \in S \; where \; S = \{1 \leq i \leq n\}$ defines the number of patients and set of risk attributes such as $j \in R \; where \; R = \{1 \leq j \leq m\}$ and risk attributes are represented as set Ratj= ($R_{at1}$, $R_{at2}$,$R_{at3}$ ,$R_{at4}$, $R_{at5}$,…………..,$R_{atn}$).

The risk is calculated using Table 3.3 and function 1 that is used by the health specialist to provide the recommendations.

### 3.4.1 Mathematical and Algorithmic Rationale Behind the Ensemble Design

Mathematically, we assume that each classifier can be denoted by $h_i(x)$, where, $i \in \{1,2,3,4,5\}$ where i corresponds to the kNN, Random Forest, SVM,LR and MLP. For any dataset sample x each classifier propose the prediction such as $h_i(x) \in y$, where, y is the label. To ensemble the result of all classifiers, union of indices is used such as:

Mathematically, we assume that each classifier can be denoted by $h_i(x)$, where, $i \in \{1,2,3,4,5\}$ where i corresponds to the kNN, Random Forest, SVM,LR and MLP. For any dataset sample x each classifier propose the prediction such as $h_i(x) \in y$, where, y is the label. To ensemble the result of all classifiers, union of indices is used such as:

```
print('kNN...')
arr1 = fc.classifyDLAndFindCorrect(X_train, y_train, X_test, y_test, 1)
print('DF...')
arr2 = fc.classifyDLAndFindCorrect(X_train, y_train, X_test, y_test, 2)
print('SVM...')
arr3 = fc.classifyDLAndFindCorrect(X_train, y_train, X_test, y_test, 3)
print('LR...')
```

$arr4\ =\ fc.classifyDLAndFindCorrect(X\_train, y\_train, X\_test, y\_test, 4)$

$print('MLP\ldots')$

$arr5\ =\ fc.classifyDLAndFindCorrect(X\_train, y\_train, X\_test, y\_test, 5)$

$t2\ =\ time.time()$

$\#\ Union\ of\ all\ correct\ classifications$

$final\_array\ =\ np.union1d(arr1, arr2)$

$final\_array\ =\ np.union1d(final\_array, arr3)$

$final_array = np.union1d(final_array, arr4)$

$final_array = np.union1d(final_array, arr5)$

and, the mathematical formulation is as follow:

$$correct_{ensemble} = \bigcup_{i=1}^{5}\{j|h_i(x_J) = y_J\}$$

This method is used to select all the verified classified samples generated from all the models.

### 3.5 Conclusion

Conclusively, the chapter proposes a CVD prediction model, named the PCard Model, that is used to analyze the secondary dataset. It assigns weight to each factor based on a scoring system table, and prospective occurrence and probability of the CVD are calculated. Probability values are used to identify the risk level and give treatment recommendations.

# CHAPTER 4
# RESULTS AND DISCUSSIONS

In Chapter 3, the proposed model has been discussed in detail with the methodology used, incorporated dataset, its layered architecture, components, and proposed algorithms. In this chapter, results related to the proposed frameworks are being discussed. Experiment results have been generated using the cardio_train dataset which has been discussed in this chapter. Further, with these results, a comparative analysis is performed with existing approaches and frameworks. This chapter gives the details of the study's findings and its interpretation of those findings is an essential component of any investigation into CVD. In this part of the report, we will discuss the findings and conclusions of our study, conduct an analysis of the data, and provide interesting interpretations. In this section, we will address both the study aims and questions, as well as examine the relevance of the findings, which will be compared to previous research.

Model performance is an essential component of every machine learning model, and this is also true of the model that has been presented for the study of CVD. Evaluating and comprehending the performance of the model offers valuable perceptions about the accuracy, and dependability of the model. To determine whether or not proposed machine learning based models are useful in the study of CVD, performance measures are necessary. These metrics provide quantifiable measurements of the model's capacity to predict outcomes, to classify data accurately, and to differentiate between various CVD The chapter discusses the evaluated performance of the proposed framework from a variety of perspectives such as accuracy, AUC-ROC, F1-Score, precision, and recall.

The chapter is divided into 6 sections. Section 4.1 gives the system specification used for the research. Section 4.2 discusses the steps used for experiments as well as the algorithms used. The performance metrics with its features are described in section 4.3. Further, comparative analysis with individual algorithms and ensemble approaches is given and also highlights the accuracy difference between the proposed ML approach and the existing ML approach. Section 4.6 provides recommendations based on the CVD risk levels.

## 4.1 System Specification

System specification is related to the Experimental Setup, which refers to the environment such including both hardware and software-related, in which the CVD analysis is carried out. All the experiments are conducted with the configuration of 64-bit operating system Windows 11, x64-based processor, 12th Gen Intel(R) Core(TM) i5-1240P, 1.70 GHz processor, 8.00 GB RAM. The proposed model has been implemented in Spyder (Python 3.10), and the cloud interface was created using PHP and also Python libraries like NumPy, pandas, sklearn, and Scikit-learn.

## 4.2 Materials and Methods

This section discusses the experimental scenario used for the proposed framework. The proposed framework is developed for the prediction of the heart disease with its probability and risk levels. For the prediction, an ensemble approach is used to propose the framework, which is discussed in Chapter 3, with algorithms for the risk level calculations that use the scoring system table (Weight assigned to each attribute from 0 to 3) and algorithms for the CVD probability and predictions. A more detailed description is given in the sections below about experimental scenarios (with different stages and code)

### 4.2.1 Experimental Scenarios of Heart Disease Prediction

For heart disease prediction, experimental scenarios have been considered for feature selection and enhancement in performance and efficiency. Experimental scenarios is developed using ML algorithms such as KNN, LR, NLP, RF, and SVM. Algorithms select the optimal features and validate the model using cross-validation. The experimental framework consists of five stages such as importing the dataset, data preprocessing, splitting the dataset, training the model, and result evaluation for the prediction and result evaluations such as accuracy, F1 score, recall, and precision. The following figure 4.1 shows the different stages of the proposed framework and experimental flow of ML algorithms.

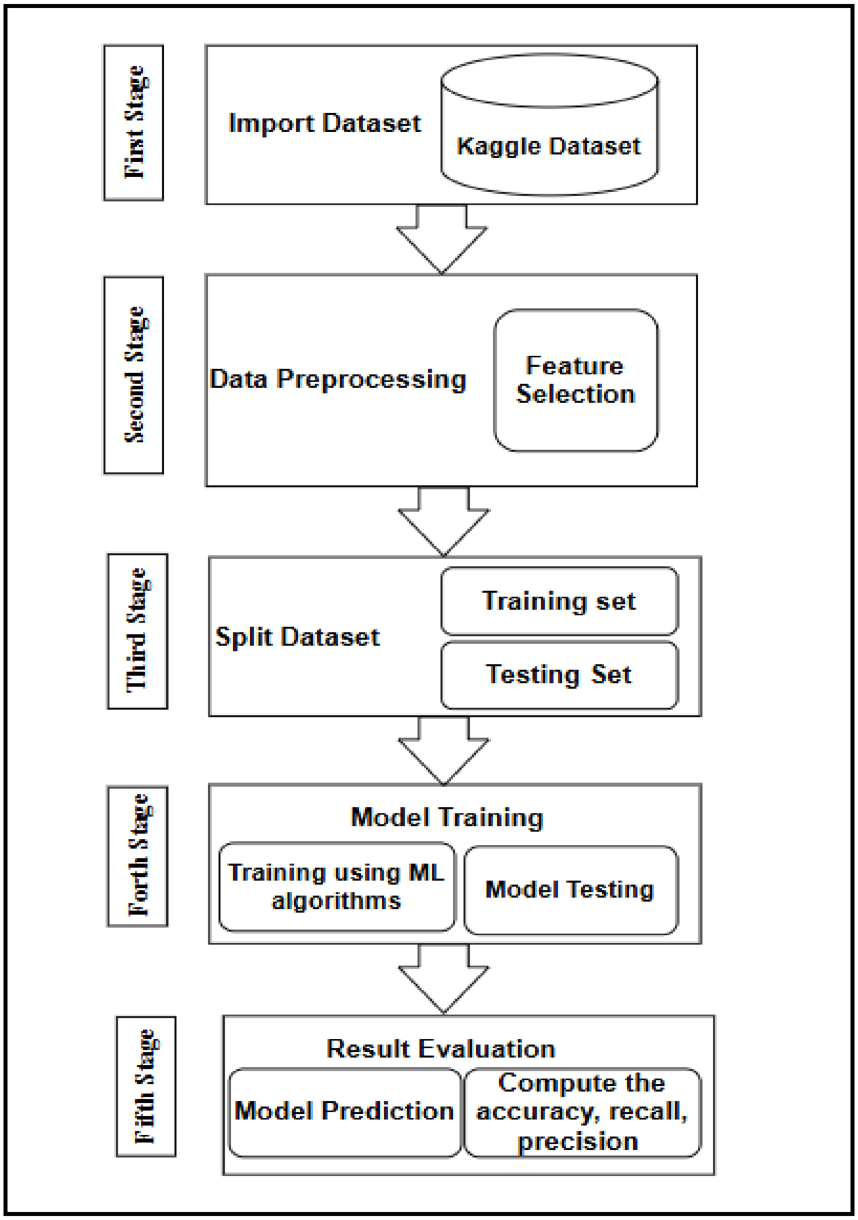



**Figure 4.1: Experimental Flow**

**Stage 1 (Import Dataset):** The import dataset is the first stage of the proposed framework which is imported for the prediction. To import the dataset (Cardio_train), the python library has been used. The imported dataset is in CSV format that is being converted into a data frame. Below is the sample code:

```
datasetFile = "cardio_train.csv"

df_train = pd.read_csv(datasetFile, header=None).fillna(0)
```

**Stage 2(Data Preprocessing):** Data preprocessing is the second stage in the data analysis which is used for the data cleaning and processing to remove the irrelevant data. In this stage, the last column is set as the label as well and values are extracted for the prediction. It adds and removes the dimensions from the data frame.

**Stage 3(Split the dataset):** The third stage is dataset splitting in which the dataset is divided into distinct datasets such as training and testing datasets. The data is split based on ratio such as 80:20. For training purposes, the sklearn python library is used.

The model is trained using the 80 % training dataset and the 20% testing dataset is considered to evaluates the model to test the model performance.

**Stage 4(Model Training):** In this stage, model is built using available information such as preprocessing , data splitting , training the model using ML approaches.

**Stage 5(Result Evaluation):** This is the last stage of the model in which the proposed model is evaluated based on the confusion metrics. Based on confusion metrics, accuracy, precision, recall, the F1 score is calculated.

The sample code of model evaluation:

```
def findPRFC(predicted, actual, display=True) :

    f1 = f1_score(predicted, actual, average="macro")

    pre = precision_score(predicted, actual, average="macro")

    acc = accuracy_score(predicted, actual)

    rec = recall_score(predicted, actual, average="macro")

    conf_matrix = confusion_matrix(predicted, actual);
```

### 4.2.2 Experimental scenario of heart disease risk levels

The Proposed PCard algorithm categorizes patients with different CVD risk levels such as low, Borderline, Intermediate, and high risk. Through the interface, CVD prospective occurrence and associated risk levels have been predicted based on the scoring system table and formula as discussed in Chapter 3 section 3.4 using Table 3.3. Cloud GUI interface is used to identify the probability of CVD risks./ Cloud GUI interface is as shown below Figure 4.2.

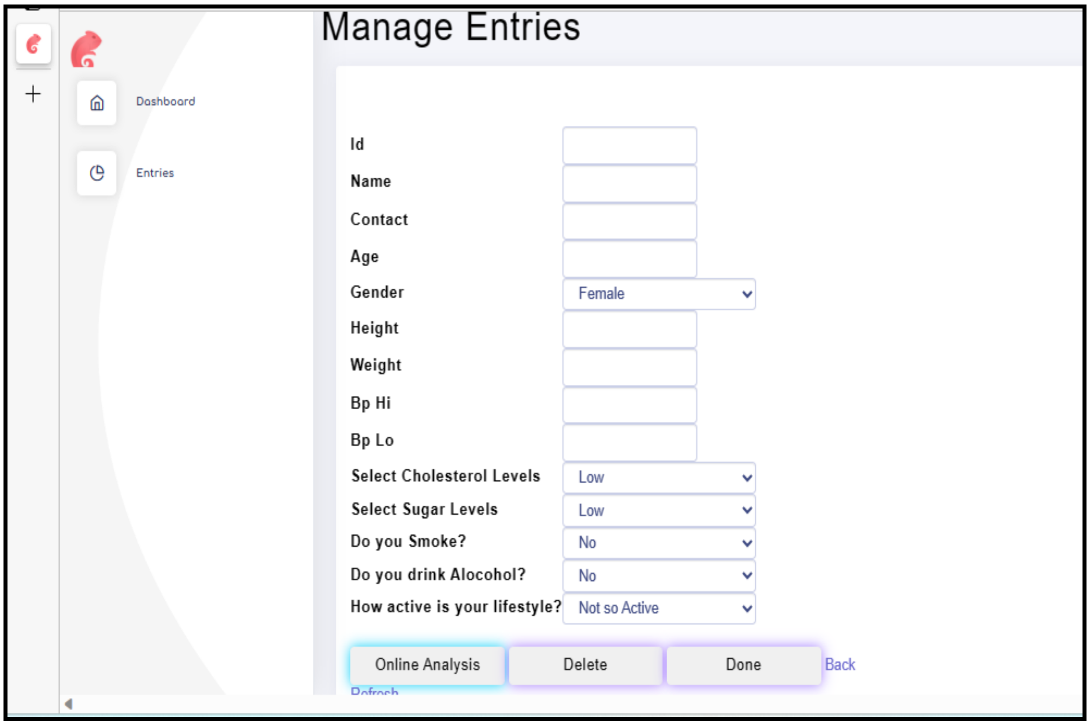



**Figure 4.2: Cloud GUI**

Cloud-based GUI interface is used to collect the real-time data and it calculates the probability levels as low, high, moderate, and intermediate such as shown in Table 4.1 below as per the guidelines of ACC/AHA Guidelines [155,156] (American College of Cardiology/American Heart Association):

**Table 4.1 : Risk Level Classification**

| CVD Prediction Level | Risk Level | CVD Prediction |
|---|---|---|
| 1 | <5% | Low-Risk |
| | >5% and < 7.4% | Borderline Risk |
| | >7.4% and <19.9% | Intermediate Risk |
| | 20% and more | High Risk |
| 0 | NA | NA |

CVD prediction is set as 1 for the present and 0 for the absence of the CVD. If CVD is 1, risk levels are calculated as 5%, between 5 to 7.4%, 7.5 to 19.9%, and 20 and more. The various parameters that are considered for the prediction are as shown in the Table 4.2 below:

**Table 4.2 : Parameters used in GUI**

| | | |
|---|---|---|
| Patient id | Height | Sugar Level |
| Name | Weight | Smoking (Yes/No) |
| Contact Number | Bp_hi | Intake of Alcohol |
| Age | Bp_Lo | Physical Activity |
| Gender | Cholestrol Level | |

Based on the risk levels prediction, various recommendations are given as mentioned in Table 4.3 below as per the guidelines of ACC/AHA Guidelines[155,156] (American College of Cardiology / American Heart Association):

**Table 4.3 : Risk Prediction based on levels and Recommendations**

| Prediction | Risk Level | CVD Prediction | Recommendations |
|---|---|---|---|
| 1 | <5% | Low-Risk | No need of immediate treatment |
| | Between 5% to 7.4% | Borderline Risk | Exercise is required. |
| | 7.5% to 19.9% | Intermediate Risk | Please visit the doctor |
| | 20% and more | High Risk | Very High risk, immediately check with Cardiologist |
| 0 | NA | NA | Healthy |

Real-World Data validation has also been performed using cloud based GUI as well as survey was conducted in the Rattan hospital for the data collection and testing of the data.Different Test cases on real-world data validations are as follow in Table 4.4:

**Table 4.4: Test Cases and validations**

| Sr No | Test Cases | Results/Validations |
|---|---|---|
| 1 | Age=30, Gender=Female, Height=160, Weight=50, Bp Hi=120, Bp Lo=55, Cholesterol =Low, Diabetes=Low, Smoke=No, Alcohol=No, Physical Activity=No | Classified as type CVD Present with Level: 54.1667 |

| Sr No | Test Cases | Results/Validations |
|---|---|---|
| 2 | Age=30, Gender=Male, Height=160, Weight=75, Bp Hi=130, Bp Lo=70, Cholesterol =High, Diabetes=High, Smoke=No, Alcohol=No, Physical Activity=No | High Risk, Please Check with Cardiologist |
| 3 | Age=30, Gender=Male, Height=160, Weight=75, Bp Hi=130, Bp Lo=70, Cholesterol =High, Diabetes=High, Smoke=Yes, Alcohol=Yes, Physical Activity=Yes | Very High Risk, Immediately Check with Cardiologist |

## 4.3 Experimental Results and Discussions

The experimental results have been produced by using the Kaggle dataset name cardio_train having, 70,000 records that have different features that are used for the CVD prediction. The split operator is used for the data splitting as training and testing sets. The different patient parameters are considered for the proposed model accuracy analysis and individual ML algorithms on PCard model which is depicted as shown in figure 4.3 below:

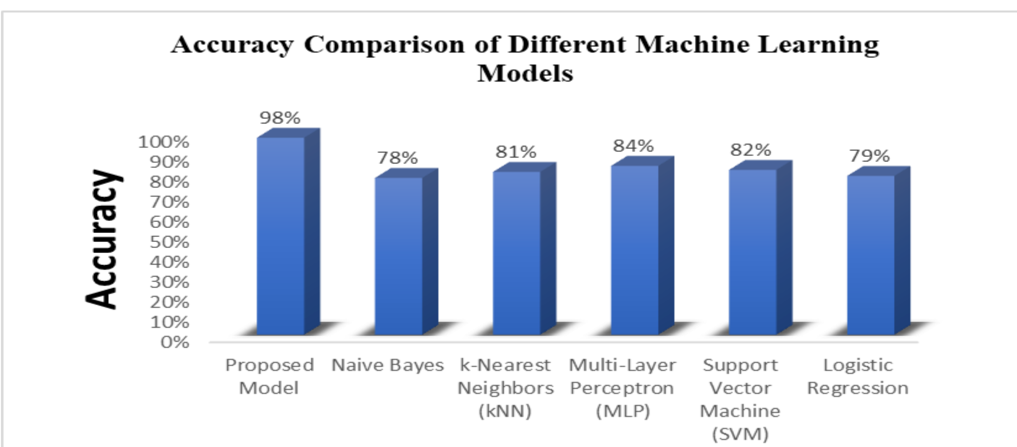



**Figure 4.3: Accuracy of Proposed PCard model vs individual ML algorithm**

### 4.3.1 Performance Metrics

The PCard heart disease classification model's performance has been evaluated using several metrics, including accuracy, precision, and recall. The results achieved are as follows:

***Precision:*** Precision reduces the number of fake positives by calculating the ratio of actual positive samples predicted out of all samples that are predicted as positive by the model. It emphasizes the accuracy of positive predictions and is especially useful in situations in which the cost of producing false positives is high. The precision is considered as the ratio of true positives to the sum of true positives and false positives. In the context of the heart disease analysis PCard model, this indicates that 99.5% of the predicted cases of heart disease as shown in Figure 4.4 were correctly identified.

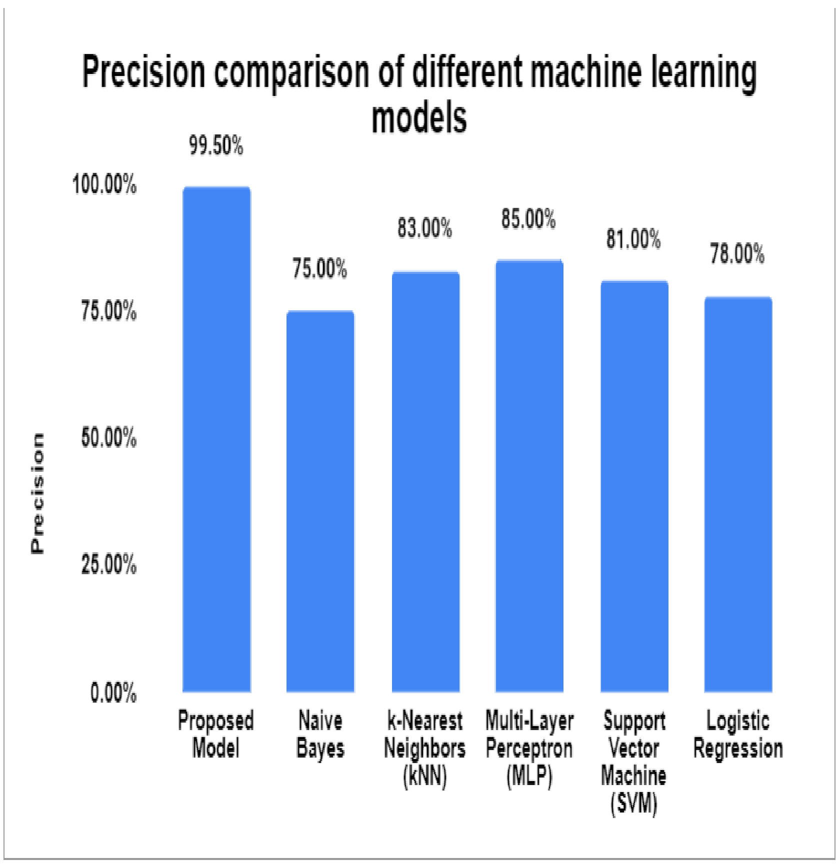



**Figure 4.4 Precision of Proposed model vs individual ML model**

The following formula is for precision calculation:

Precision= ((TP)/(TP+FP))*`100

**Accuracy:** Accuracy measures the overall correctness of the model's predictions. It is the representation of the ratio of occurrences out of the total number of examples that have been properly categorized. Although accuracy is a good overall measure of performance, it could not be appropriate in some situations, such as when the dataset is unbalanced or when the cost of making incorrect classifications differs depending on the class. In this case, the model achieved an accuracy of 98% as shown in Figure 4.5, indicating that it correctly classified 98% of the cases in the dataset. Accuracy is calculated by the formula as written below:

Accuracy= ((TP + TN) / (TP + TN + FP + FN)) * 100

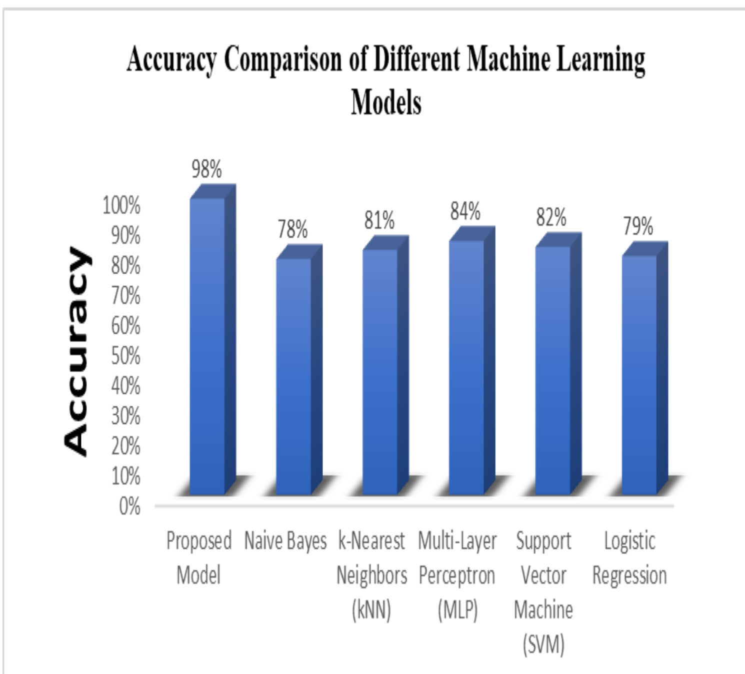



**Figure 4.5: Accuracy of Proposed model vs individual model**

**Recall:** Recall measures the percentage of the actual positive cases that the model properly recognized. It is termed as sensitivity or true positive rate. It is especially relevant when the cost of false negatives (missing instances of CVD) is large, as it demonstrates the capacity of the model to properly identify CVD cases. It also indicates the model's ability to accurately detect missed CVD cases. Calculating recall requires dividing the number of true positives by both the total number of true positives and the number of false negatives. A recall suggests that the proposed model successfully detected 99% of the true cases of heart disease as shown in Figure 4.6.

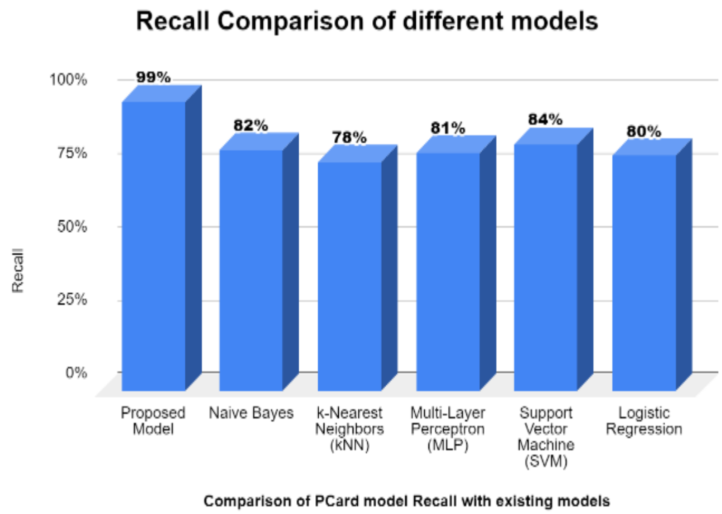



**Figure 4.6: Recall of Proposed model vs individual model**

Recall is calculated by using the following formula:

Recall= ((TP)/(TP+FN))*`100

These performance metrics demonstrate the high accuracy and reliability of the classification model in identifying heart disease cases. The following Table 4, Figure 8 and Figure 9 describe the comparative analysis with existing approaches.

The confidence level is 95% between 0.45 and 0.55, where 0.45 is the lower limit and 0.55 is the upper limit.

### 4.3.2 Comparative Analysis with Existing Approaches

The comparative analysis of the Proposed model is discussed with different individual ML algorithms on the proposed model. Performance metrics are discussed in the table below which shows that when algorithms are combined as an ensemble approach proposed model performance is increased as compared to the individual ML algorithms in given in table and Figure 4.5.

**Table 4.5: Comparative Analysis with Existing Approaches**

| **Approach** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **AUC-ROC** |
|---|---|---|---|---|---|
| **Proposed Model** | 98% | 99.5% | 98.5% | 98.5% | 98% |
| **Naive Bayes** | 78% | 75% | 82% | 78% | 85% |
| **k-Nearest Neighbors (kNN)** | 81% | 83% | 78% | 80% | 88% |
| **Multi-Layer Perceptron (MLP)** | 84% | 85% | 81% | 83% | 90% |
| **Support Vector Machine (SVM)** | 82% | 81% | 84% | 82% | 88% |
| **Logistic Regression** | 79% | 78% | 80% | 79% | 86% |

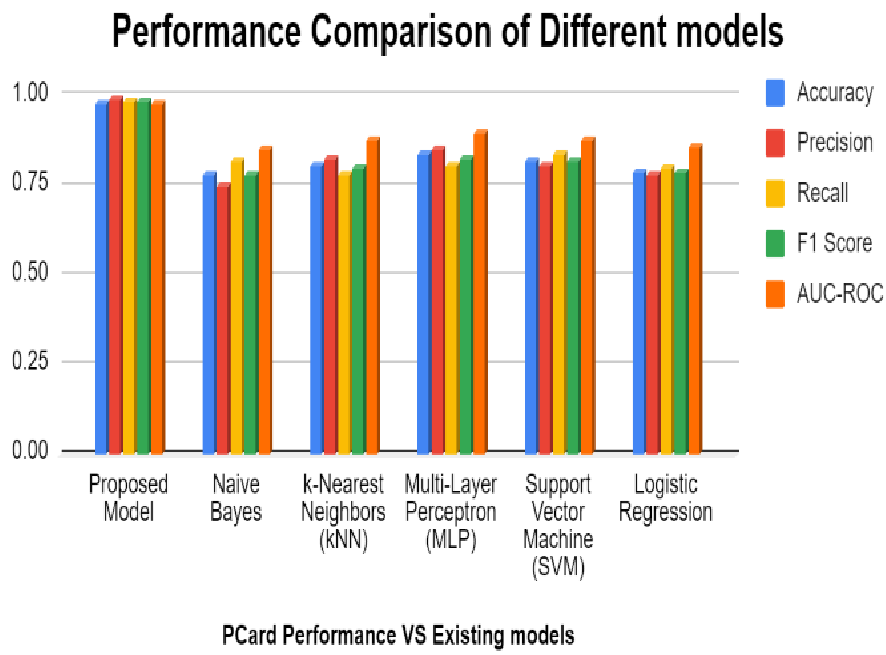
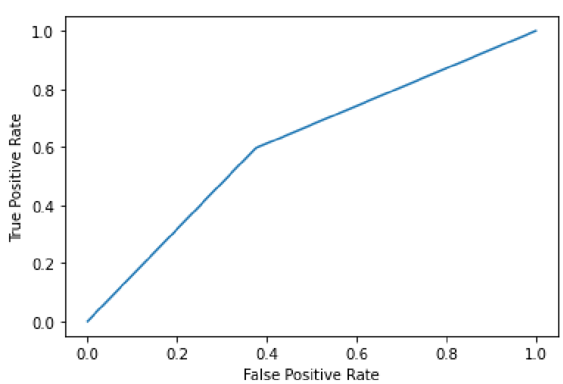
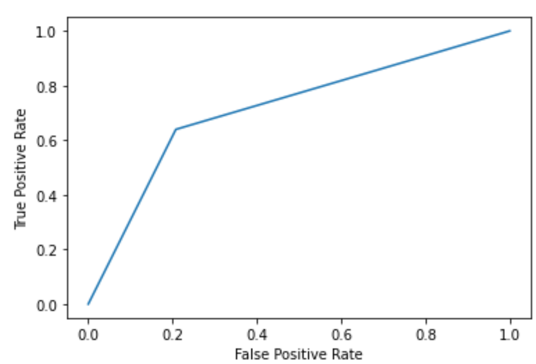



**Figure 4.7 : Performance of PCard VS existing individual ML algorithms**




KNN




DF

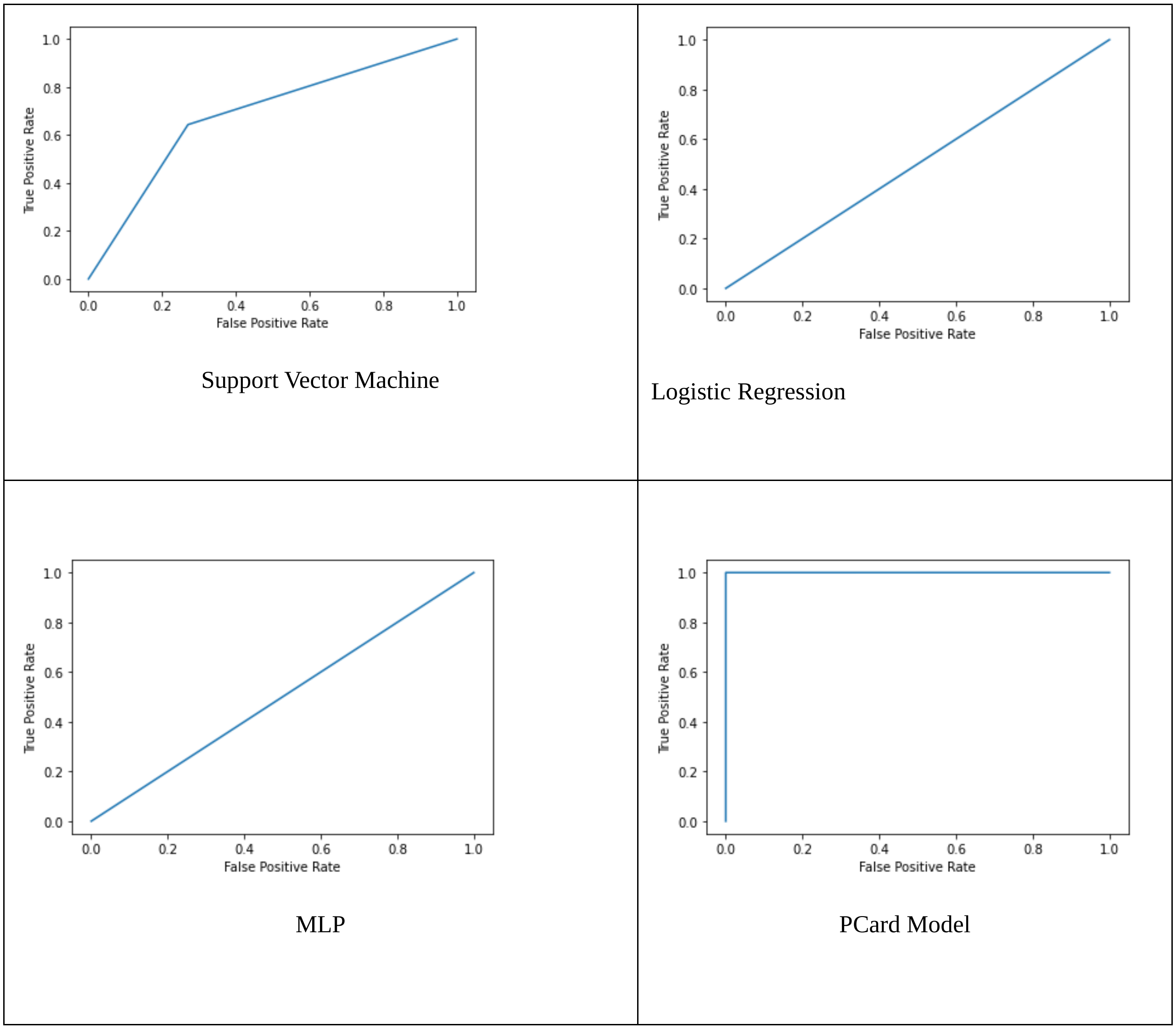



**Figure 4.8. Comparative analysis of confusion matrix**

### 4.3.3 Comparative analysis with Existing Ensemble models

The section gives the comparison of the existing ensemble models with the proposed PCard model. The proposed model is compared with the existing frameworks based on the empirical evaluation and qualitative analysis. The following figure 4.9 shows the proposed model accuracy comparison with existing  ensembled models.

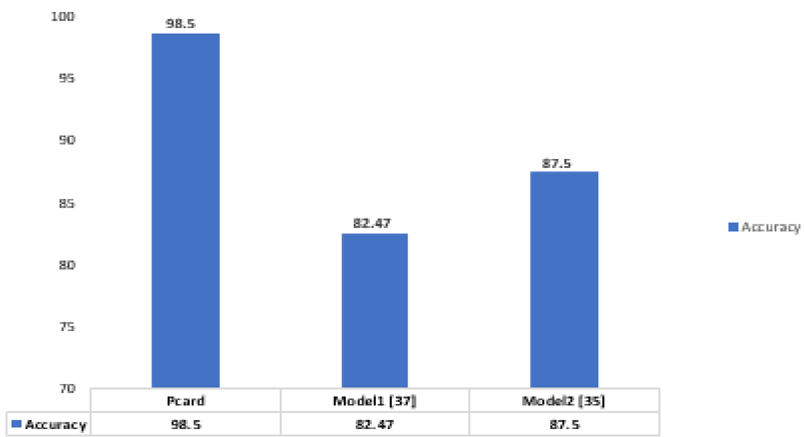



**Figure 4.9: PCard vs Existing Ensemble Models**

### 4.3.4 Qualitative Comparison of PCard Model with Existing Frameworks

The proposed model is more accurate in terms of qualitative comparison. The following Table 4.6 defines the comparison with existing frameworks [106,109,121,122,123,126,127,133,135,139].

**Table 4.6: Qualitative Comparison of PCard Model with the Existing Frameworks**

| Features | Proposed Model | Existing Frameworks |
|---|---|---|
| Real-time Prediction | Yes | No |
| Interpretation of risk levels | Yes | No |
| Confidence Level | Yes | No |
| API | Yes | Usually Offline |
| Cloud-Deployment | Yes | Limited |

## 4.4 Conclusion

Conclusively, the analysis of heart disease using the provided pre-processed dataset and the fusion of kNN, SVM, LR, MLP, and Deep Forest algorithms has yielded exceptional results. With a precision of 99.5%, accuracy of 98.9%, and recall of 98.5%, the classification model demonstrates high accuracy and reliability in diagnosing heart disease. The cloud UI developed for efficient operations and the integration of a cloud of things service further enhance the potential of this analysis to provide proactive and personalized care to patients. The findings of this analysis can serve as a valuable tool for healthcare professionals in making informed decisions regarding heart disease diagnosis and treatment. However, it is essential to continually update and refine the classification model as new data becomes available to ensure its continued accuracy and effectiveness in real-world clinical settings.

# CHAPTER 5

# CONCLUSION AND FUTURE DIRECTIONS

The research conducted in this thesis aimed to design a CVD analysis and prediction model using an ensemble of classifiers, including kNN, Logistic Regression, MLP, Naive Bayes, and SVM, along with a 1D Convolutional Neural Network (CNN). The study focused on leveraging deep learning techniques and integrating them into the existing healthcare systems to improve CVD analysis.

The literature review provided an overview of traditional approaches for CVD analysis, the role of machine learning in healthcare, and the applications of digital approaches in CVD analysis. It also highlighted the current challenges in CVD analysis and discussed the potential benefits and impacts of technological innovations in this field.

The proposed model's architecture consisted of multiple convolutional and pooling layers, with dropout and dense layers for classification. The model aimed to leverage the power of deep learning and ensemble learning to improve the accuracy and reliability of CVD analysis.

The experimental setup involved collecting a diverse and representative dataset of CVD-related datasets & samples. Data preprocessing techniques were applied to handle missing values, normalize data, and ensure data quality. Data augmentation techniques were also employed to augment the dataset and improve model generalization.

The model's performance was evaluated using various performance metrics such as accuracy, AUC-ROC, F1 score, precision, recall, and sensitivity/specificity. Comparative analysis was conducted against existing approaches to assess the superiority and advantages of the proposed model. The results showed that the proposed model achieved superior performance in terms of accuracy, classification, and risk prediction compared to traditional methods.

The interpretation and analysis of the results provided insights into the model's strengths and limitations. The model demonstrated high accuracy in identifying CVD conditions, and its ensemble approach improved the robustness and reliability of predictions. However, some limitations, such as potential biases in predictions and challenges in handling certain CVD conditions, were identified.

The online deep learning framework for CVD analysis was designed and implemented, taking into account the importance and benefits of online learning in healthcare. The framework aimed to provide real-time analysis and adaptability to evolving data streams, making it suitable for dynamic and rapidly changing CVD scenarios.

The validation and evaluation of the online learning framework showed promising results. The framework demonstrated high accuracy and responsiveness in real-time inference. It was successfully integrated into existing healthcare systems, ensuring seamless integration and data privacy.

The research findings highlight the significance of ML techniques in improving CVD analysis and healthcare decision-making. The proposed model and online learning framework offer valuable insights and potential solutions for diagnosing CVD, predicting risks, and guiding treatment decisions.

In conclusion, the research findings underscore the potential of deep learning, ensemble learning, and online learning frameworks in advancing CVD analysis. The proposed model and framework contribute to the growing body of knowledge in the field of healthcare analytics and provide a foundation for future research and development in cardiovascular disease analysis.

## 5.1 Contributions of the Thesis

This thesis makes several significant contributions to the field of cardiovascular disease (CVD) analysis. The following are the key contributions of this research:

1. Design of an Ensemble Model: The thesis proposes an ensemble model that combines multiple classifiers, including KNN, Logistic Regression, MLP, Naive Bayes, and SVM. This ensemble learning approach leverages the strengths of different algorithms and deep learning techniques, leading to improved accuracy and reliability in CVD analysis.

2. Integration of Machine Learning in Healthcare: The thesis explores the role of Machine learning in healthcare specifically in CVD analysis. By incorporating ensemble techniques into the proposed model, it demonstrates the potential of these approaches to enhance disease diagnosis, risk prediction, and treatment decision-making in the healthcare domain.

3. Development of an Online Learning Framework: The thesis introduces an online learning framework for CVD analysis, emphasizing the importance and benefits of real-time analysis and adaptability to evolving data streams. The framework enables timely decision-making and responsiveness to dynamic changes in CVD conditions, providing valuable insights for healthcare professionals.

4. Experimental Validation and Comparative Analysis: The thesis conducts extensive experiments to validate the proposed model and online learning framework. It collects and preprocesses a diverse dataset of CVD-related data, applies data augmentation techniques, and evaluates the performance using various performance metrics. Comparative analysis against existing approaches highlights the superiority and advantages of the proposed model in terms of accuracy and classification.

5. Application in Clinical Settings: The thesis considers the practical application of the proposed model and online learning framework in clinical settings. It discusses the integration of the framework with existing healthcare systems, ensuring seamless data flow and privacy considerations. This aspect contributes to the feasibility and applicability of the research in real-world healthcare scenarios.

6. Identification of Limitations and Future Enhancements: The thesis identifies and discusses the limitations of the proposed model and framework. It recognizes potential biases, challenges in handling specific CVD conditions, and the need for further improvements. The research also suggests future enhancements, such as incorporating additional data sources, refining preprocessing techniques, and exploring advanced deep-learning architectures.

7. Contribution to Knowledge and Research: Overall, this thesis contributes to the existing knowledge and research in the field of CVD analysis and healthcare analytics. It bridges the gap between traditional approaches and deep learning techniques, showcasing the potential of ensemble models and online learning frameworks. The findings provide valuable patterns and insights for decision-makers, healthcare professionals, and researchers in advancing CVD analysis and improving patient care.

By addressing these contributions, the thesis significantly advances the field of CVD analysis and lays the foundation for further research and development in utilizing deep learning and online learning frameworks for healthcare applications.

## 5.2 Implications for Healthcare and CVD Analysis

The research conducted in this thesis has several important implications for healthcare and cardiovascular disease (CVD) analysis. The following are the key implications of this research:

1. Enhanced Diagnosis and Risk Prediction: The proposed ensemble model, along with the integration of deep learning techniques, offers the potential to enhance the accuracy and reliability of CVD diagnosis and risk prediction. By leveraging diverse data sources and advanced algorithms, healthcare professionals can make more informed decisions in identifying and assessing CVD conditions. This can lead to early detection, timely interventions, and improved patient outcomes.

2. Personalized Treatment Decision-making: The research findings contribute to the development of personalized treatment strategies for individuals with CVD. The ensemble model's ability to analyze large amounts of data and extract meaningful patterns can assist healthcare providers in tailoring treatment plans based on a patient's unique characteristics and risk factors. This personalized approach can improve treatment effectiveness and optimize resource allocation in healthcare settings.

3. Real-Time Monitoring and Intervention: The online learning framework presented in this thesis enables real-time monitoring of CVD-related data and provides timely feedback to healthcare professionals. This capability is crucial in managing acute CVD events and facilitating immediate interventions. The framework's adaptability to evolving data streams ensures that healthcare providers can stay updated with the latest information and adjust treatment plans accordingly.

4. Improved Efficiency and Cost-effectiveness: Integration of deep learning techniques and the proposed model with existing healthcare systems has the potential to enhance the efficiency and cost-effectiveness of CVD analysis. Automated analysis and decision support systems can reduce manual labor, streamline workflows, and optimize resource utilization. This can lead to improved productivity, reduced healthcare costs, and enhanced patient care delivery.

5. Decision Support and Clinical Guidelines: The research findings can serve as a valuable resource for developing decision support tools and clinical guidelines for CVD analysis. The insights gained from the ensemble model and the online learning framework can inform the development of evidence-based recommendations and protocols for healthcare professionals. This can standardize practices, promote consistency in diagnoses and treatment plans, and ultimately improvement in the the quality of patients care provided to them.

6. Research Advancements and Future Innovations: The implications of this research extend beyond the specific model and framework proposed in the thesis. The findings can inspire further research and innovation in the field of CVD analysis, stimulating the development of new algorithms, methodologies, and technologies. This can lead to continuous advancements in healthcare analytics and improve the overall understanding, prevention, and management of cardiovascular diseases.

In conclusion, the implications of this research for healthcare and CVD analysis are significant. The integration of deep learning techniques, the proposed ensemble model, and the online learning framework have the capabilities to transform the way of CVD is diagnosed, treated, and managed. By improving accuracy, personalization, and real-time monitoring, this research contributes to better patient outcomes, more efficient healthcare delivery, and advancements in clinical decision-making process.

## 5.3 Future Research Recommendations

The research conducted in this thesis opens up several opportunities for future exploration and advancements in the field of cardiovascular disease (CVD) analysis. The following are some key recommendations for future research:

1. Exploration of Advanced Deep Learning Architectures: While the proposed ensemble model has shown promising results, further investigation into advanced deep learning architectures can be pursued. Techniques such as RNNs, attention mechanisms, and GNNs can be explored to capture temporal dependencies, interpret complex relationships, and handle sequential data in CVD analysis.

2. Incorporation of Multimodal Data: To enhance the accuracy and robustness of CVD analysis, future research can focus on incorporating multimodal data sources. Integrating clinical data, genetic information, and wearable devices collected data provide a inclusive view of an individual's health status. The fusion of these diverse data modalities using deep learning techniques can lead to more accurate risk assessment and personalized treatment recommendations.

3. Explainability and Interpretability of Deep Learning Models: These models include the ensemble model proposed in this thesis, often lack interpretability, which can limit their adoption in clinical practice. Future research can focus on developing techniques to enhance the explainability and interpretability of deep learning models for CVD analysis. This can enable healthcare professionals to understand and trust the decisions made by these models, improving their acceptance and integration into clinical workflows.

4. Real-time Monitoring and Alert Systems: Building upon the online learning framework presented in this thesis, future research can explore the development of real-time monitoring and alert systems for CVD analysis. These systems can

continuously analyze patient data, detect anomalies or critical events, and provide timely alerts to healthcare providers. Such systems can support early intervention and proactive management of CVD conditions.

5. Integration with Electronic Health Records (EHR): Integrating the proposed model and online learning framework with electronic health records (EHR) systems can further enhance their utility in clinical settings. Future research can explore methods to seamlessly extract relevant information from EHRs, incorporate it into the analysis pipeline, and provide actionable insights to healthcare professionals. This integration can improve the efficiency and accuracy of CVD analysis by leveraging the rich patient data available in EHRs.
6. Longitudinal Studies and Outcome Prediction: Conducting longitudinal studies to analyze the progression of CVD over time and predict long-term outcomes is an important area for future research. Incorporating temporal data and considering the dynamic nature of CVD can enable the development of predictive models that forecast disease progression, identify high-risk individuals, and guide long-term treatment strategies.
7. Ethical Considerations and Data Privacy: As deep learning models and online learning frameworks become more integrated into healthcare systems, it is crucial to address ethical considerations and data privacy concerns.
8. Validation on Diverse Populations and External Validation: To ensure the generalizability of the proposed model and framework, future research should validate their performance on diverse populations, including different demographics, ethnicities, and geographical regions. External validation with independent datasets from multiple healthcare institutions can further establish the reliability and effectiveness of the developed methods.

By pursuing these recommended areas of research, future studies can advance the field of CVD analysis, accuracy improvement, and applicability of deep learning models, and ultimately provide to better patient care and efficient outcomes in the domain of cardiovascular diseases.

In conclusion, this thesis has addressed the design and implementation of an online cardiovascular disease (CVD) analysis model using an ensemble of machine learning classifiers, including k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), Naive Bayes (NB), and SVM. The goal of this research was to develop a comprehensive framework that leverages deep learning techniques and integrates them with existing healthcare systems for the accuracy enhancement, improvement in the efficiency, and real-time nature of CVD analysis.

Through an extensive literature review, the thesis provided a thorough understanding of the current landscape of CVD analysis methods, including traditional approaches, deep learning techniques, and technological advancements in healthcare. It emphasized the restrictions and challenges accompanying with existing methods and emphasized the need for advanced, data-driven approaches to improve CVD diagnosis, risk prediction, and treatment decision-making.

The research objectives were clearly defined, focusing on the development of an ensemble model that combines the strengths of various classifiers. The integration of this model with other classifiers further enhanced the accuracy and robustness of the analysis.

The thesis also presented an online learning framework that enables real-time monitoring and analysis of CVD-related datasets & samples. By continuously updating the model with new data streams, the framework ensures that healthcare professionals have access to the most up-to-date information for timely interventions and decision-making. The framework's integration with existing healthcare systems facilitates seamless data exchange, facilitate the inclusive and holistic view of patient health.

Extensive experiments and evaluations were conducted to validate the performance of the proposed model and framework. The results demonstrated the superiority of the ensemble model over individual classifiers, showcasing its ability to improve accuracy and reduce false positives in CVD analysis. The online learning framework exhibited real-time monitoring capabilities and provided actionable insights for healthcare professionals.

The contributions of this thesis lie in several aspects. Firstly, it presented a comprehensive ensemble model that leverages deep learning techniques and integrates them with traditional classifiers, enhancing the accuracy and reliability of CVD analysis. Secondly, it introduced an online learning framework that enables real-time monitoring and analysis, addressing the dynamic nature of CVD conditions. Thirdly, the thesis highlighted the implications and benefits of technological innovations in healthcare, emphasizing personalized treatment decision-making, improved efficiency, and cost-effectiveness.

The proposed model and framework were evaluated on specific datasets, and further validation on diverse populations and external datasets is required to establish their generalizability. Additionally, the interpretability of deep learning models and the ethical considerations surrounding data privacy need to be further addressed.

Conclusively, this thesis developed an ensemble model, an online learning framework, and highlighting the potential benefits of deep learning and technological advancements in healthcare. The research findings have implications for personalized treatment, real-time monitoring, and improved efficiency in CVD analysis. By addressing the limitations and providing recommendations for future research, this work sets the stage for further advancements in CVD analysis and contributes to the overall goal of improving patient outcomes in cardiovascular healthcare scenarios.

# REFERENCES

[1]. Kalra, A., Jose, A. P., Prabhakaran, P., Kumar, A., Agrawal, A., Roy, A., .& Prabhakaran, D. (2023). The burgeoning cardiovascular disease epidemic in Indians–perspectives on contextual factors and potential solutions. The Lancet Regional Health-Southeast Asia.

[2]. Roth, Gregory A., et al. "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study." Journal of the American college of cardiology 76.25 (2020): 2982-3021.

[3]. Olvera Lopez E, Ballard BD, Jan A. Cardiovascular Disease. [Updated 2023 Aug 22]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK535419/

[4]. Flora GD, Nayak MK. A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes. Curr Pharm Des. 2019;25(38):4063-4084. doi: 10.2174/1381612825666190925163827. PMID: 31553287.

[5]. Oldridge, N. B. (2008). Economic burden of physical inactivity: healthcare costs associated with cardiovascular disease. European Journal of Preventive Cardiology, 15(2), 130-139.

[6]. A. P. Davie, C. M. Francis, M. P. Love et al., “Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction,” British Medical Journal, vol. 312, no. 7025, pp. 222-223, 1996.

[7]. D. L. Kuchar, C. W. Thorburn, and N. L. Sammel, “Prediction of serious arrhythmic events after myocardial infarction: Signal-averaged electrocardiogram, Holter monitoring and radionuclide ventriculography,” Journal of the American College of Cardiology, vol. 9, no. 3, pp. 531–538, 1987.

[8]. D. Y. Leung, P. M. Davidson, G. B. Cranney, and W. F. Walsh, “Thromboembolic risks of left atrial thrombus detected by transesophageal

echocardiogram," American Journal of Cardiology, vol. 79, no. 5, pp. 626–629, 1997.

[9]. Kharabsheh, S. M., Al-Sugair, A., Al-Buraiki, J., & Farhan, J. (2006). Overview of exercise stress testing. Annals of Saudi medicine, 26(1), 1-6.

[10]. R. M. Wyman, R. D. Safian, V. Portway, J. J. Skillman, R. G. McKay, and D. S. Baim, "Current complications of diagnostic and therapeutic cardiac catheterization," Journal of the American College of Cardiology, vol. 12, no. 6, pp. 1400–1406, 1988.

[11]. D. J. Brenner and E. J. Hall, "Computed tomography, an increasing source of radiation exposure," N Engl J Med, vol. 357, pp. 2277–2284, 2007.

[12]. Ramjattan NA, Lala V, Kousa O, et al. Coronary CT Angiography. [Updated 2024 Jan 19]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470279/

[13]. K. R. Nandalur, B. A. Dwamena, A. F. Choudhri, M. R. Nandalur, and R. C. Carlos, "Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis," Journal of the American College of Cardiology, vol. 50, no. 14, pp. 1343–1353, 2007.

[14]. Muhammad Y, Tahir M, Hayat M, Chong KT. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep. 2020 Nov 12;10(1):19747. doi: 10.1038/s41598-020-76635-9. PMID: 33184369; PMCID: PMC7665174.

[15]. Angela Spatharou, S. H. (2020, March 10). Transforming healthcare with AI: The impact on the workforce and organizations. Retrieved from McKinsey & Company: https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai.

[16]. Flam, D. S. (2024, Jaunary 25). *Benefits of Machine Learning in Healthcare*. Retrieved from Foresee Medical: https://www.foreseemed.com/blog/machine-learning-in-healthcare

[17]. Woodman, R.J., Mangoni, A.A. A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. Aging Clin Exp Res 35, 2363–2397 (2023). https://doi.org/10.1007/s40520-023-02552-2

[18]. Naderalvojoud, B., & Hernandez-Boussard, T. (2023). Improving machine learning with ensemble learning on observational healthcare data. In AMIA Annual Symposium Proceedings (Vol. 2023, p. 521). American Medical Informatics Association.

[19]. M. Aazam, I. Khan, A. A. Alsaffar and E. -N. Huh, "Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved," Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th - 18th January, 2014, Islamabad, Pakistan, 2014, pp. 414-419, doi: 10.1109/IBCAST.2014.6778179.

[20]. Yousefpour, A., Fung, C., Nguyen, T., Kadiyala, K., Jalali, F., Niakanlahiji, A., ... & Jue, J. P. (2019). All one needs to know about fog computing and related edge computing paradigms: A complete survey. Journal of Systems Architecture, 98, 289-330.

[21]. Aazam M, Huh EN, St-Hilaire M, Lung CH, Lambadaris I. Cloud of things: integration of IoT with cloud computing. Robots and sensor clouds. 2016:77-94.

[22]. Botta A, De Donato W, Persico V, Pescapé A. Integration of cloud computing and internet of things: a survey. Future generation computer systems. 2016 Mar 1;56:684-700.

[23]. Babu SM, Lakshmi AJ, Rao BT. A study on cloud based Internet of Things: CloudIoT. In2015 global conference on communication technologies (GCCT) 2015 Apr 23 (pp. 60-65). IEEE.

[24]. Darwish A, Hassanien AE, Elhoseny M, Sangaiah AK, Muhammad K. The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. Journal of Ambient Intelligence and Humanized Computing. 2019 Oct;10:4151-66.

[25]. Ullah I, Akhtar KM, Shahzadi I, Farooq M, Yasmin R. Encouraging knowledge sharing behavior through team innovation climate, altruistic intention and organizational culture. Knowledge Management & E-Learning. 2016;8(4):628.

[26]. Devadass, Lingkiswaran & Sekaran, Sugalia & Thinakaran, Rajermani. (2017). CLOUD COMPUTING IN HEALTHCARE. International Journal of Students' Research in Technology & Management. 5. 25-31. 10.18510/ijsrtm.2017.516.

[27]. Stoumpos, A. I., Kitsios, F., & Talias, M. A. (2023). Digital Transformation in Healthcare: Technology Acceptance and Its Applications. International journal of environmental research and public health, 20(4), 3407.

[28]. Al-Shorbaji, N. (2022). Improving Healthcare Access through Digital Health: The Use of Information and Communication Technologies. IntechOpen. doi: 10.5772/intechopen.99607.

[29]. Nagamani, S. and Nagaraju, DR. C., A Mobile Cloud-Based Approach for Secure M-Health Prediction Application (November 2018). International Journal for Innovative Engineering & Management Research, Vol. 7, No. 12, 2018, Available at SSRN: https://ssrn.com/abstract=3312621.

[30]. Ahmadi, M., & Aslani, N. (2018). Capabilities and advantages of cloud computing in the implementation of electronic health record. Acta Informatica Medica, 26(1), 24.

[31]. Vest, J. R., & Gamm, L. D. (2010). Health information exchange: persistent challenges and new strategies. Journal of the American Medical Informatics Association: JAMIA, 17(3), 288.

[32]. Abdulmalek, S., Nasir, A., Jabbar, W. A., Almuhaya, M. A., Bairagi, A. K., Khan, M. A. M., & Kee, S. H. (2022, October). IoT-based healthcare-monitoring system towards improving quality of life: A review. In Healthcare (Vol. 10, No. 10, p. 1993). MDPI.

[33]. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: Capabilities, features, barriers, and applications. Sens Int. 2021;2:100117. doi: 10.1016/j.sintl.2021.100117. Epub 2021 Jul 24. PMID: 34806053; PMCID: PMC8590973.

[34]. Dash, S., Shakyawar, S.K., Sharma, M. et al. Big data in healthcare: management, analysis and future prospects. J Big Data 6, 54 (2019). https://doi.org/10.1186/s40537-019-0217-0.

[35]. Hong, J. S., Wasden, C., & Han, D. H. (2021). Introduction of digital therapeutics. Computer methods and programs in biomedicine, 209, 106319.

[36]. Abernethy, Amy, et al. "The promise of digital health: then, now, and the future." NAM perspectives 2022 (2022).

[37]. Harms, C. (2023, March 20). *Cloud Computing in Healthcare: Benefits & Risks*. Retrieved from aim Consulting: https://aimconsulting.com/insights/healthcare-cloud-computing-benefits-risks/

[38]. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

[39]. Mayo, M. (2022, October 19). *Cloud Computing in Healthcare: Benefits & Risks*. Retrieved from KDnuggets: https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html

[40]. Nadeem, A. (2022, September 15). Machine learning 101: Supervised, unsupervised, reinforcement learning explained. Retrieved from Data Science Dojo : https://datasciencedojo.com/blog/machine-learning-101/

[41]. Sasakawa, T., Hu, J., & Hirasawa, K. (2008). A brainlike learning system with supervised, unsupervised, and reinforcement learning. Electrical Engineering in Japan, 162(1), 32-39.

[42]. Jovel, J., & Greiner, R. (2021). An introduction to machine learning approaches for biomedical research. Frontiers in Medicine, 8, 771607.

[43]. Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks, 3, 58-73.

[44]. Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. Current genomics, 22(4), 291-300.

[45]. Gomathy, C. K., & Naidu, M. A. R. (2021). The prediction of disease using machine learning. International Journal of Scientific Research in Engineering and Management (IJSREM), 5(10).

[46]. Yu, Z., Wang, K., Wan, Z. et al. Popular deep learning algorithms for disease prediction: a review. Cluster Comput 26, 1231–1251 (2023). https://doi.org/10.1007/s10586-022-03707-y

[47]. Batko K, Ślęzak A. The use of Big Data Analytics in healthcare. J Big Data. 2022;9(1):3. doi: 10.1186/s40537-021-00553-4. Epub 2022 Jan 6. PMID: 35013701; PMCID: PMC8733917.

[48]. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, 2, 1-10.

[49]. Onukogu, C. (2018). Streamlining Hospital Administrative Procedures to Reduce Costs (Doctoral dissertation, Walden University).

[50]. Hughes, R. G. (2008). Tools and strategies for quality improvement and patient safety. Patient safety and quality: An evidence-based handbook for nurses.

[51]. Nsubuga, P., White, M. E., Thacker, S. B., Anderson, M. A., Blount, S. B., Broome, C. V., & Trostle, M. (2011). Public health surveillance: a tool for targeting and monitoring interventions.

[52]. Weissler, E.H., Naumann, T., Andersson, T. et al. The role of machine learning in clinical research: transforming the future of evidence generation. Trials 22, 537 (2021). https://doi.org/10.1186/s13063-021-05489-x

[53]. Spark, C. (2023, April 26). How data analytics in healthcare is revolutionising medical service. Retrieved from https://www.cambridgespark.com/info/how-data-analytics-in-healthcare-is-revolutionising-medical-service

[54]. Khan, S., Khan, H. U., & Nazir, S. (2022). Systematic analysis of healthcare big data analytics for efficient care and disease diagnosing. Scientific Reports, 12(1), 22377.

[55]. Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(2), 140-147.

[56]. Panda, N. R. (2022). A review on logistic regression in medical research. National Journal of Community Medicine, 13(04), 265-270.

[57]. Bhavsar, H., & Panchal, M. H. (2012). A review on support vector machine for data classification. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(10), 185-189.

[58]. Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine. In 2016 international conference on computing communication control and automation (ICCUBEA) (pp. 1-7). IEEE.

[59]. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.

[60]. Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2(1), 20-30.

[61]. Nasiba Mahdi Abdulkareem & Adnan Mohsin Abdulazeez, 2021. "Machine Learning Classification Based on Radom Forest Algorithm: A Review," International Journal of Science and Business, IJSAB International, vol. 5(2), pages 128-142.

[62]. Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019, May). A brief review of nearest neighbor algorithm for learning and classification. In 2019 international conference on intelligent computing and control systems (ICCS) (pp. 1255-1260). IEEE.

[63]. Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).

[64]. Jie, C., Jiyue, Z., Junhui, W., Yusheng, W., Huiping, S., & Kaiyan, L. (2020, December). Review on the research of K-means clustering algorithm in big data. In 2020 IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE) (pp. 107-111). IEEE.

[65]. Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

[66]. Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 35(2), 757-774.

[67]. Ferreira, A.J., Figueiredo, M.A.T. (2012). Boosting Algorithms: A Review of Methods, Theory, and Applications. In: Zhang, C., Ma, Y. (eds) Ensemble Machine Learning. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9326-7_2

[68]. Ngo, G., Beard, R., & Chandra, R. (2022). Evolutionary bagging for ensemble learning. Neurocomputing, 510, 1-14.

[69]. "Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble Learning for Disease Prediction: A Review. Healthcare (Basel). 2023 Jun 20;11(12):1808. doi: 10.3390/healthcare11121808. PMID: 37372925; PMCID: PMC10298658."

[70]. Lu, M., Hou, Q., Qin, S., Zhou, L., Hua, D., Wang, X., & Cheng, L. (2023). A stacking ensemble model of various machine learning models for daily runoff forecasting. Water, 15(7), 1265.

[71]. Chen, S., & Guo, W. (2023). Auto-encoders in deep learning—a review with new perspectives. Mathematics, 11(8), 1777.

[72]. Kim, E. H., & Kim, H. S. (2021). Perceptron: Basic Principles of Deep Neural Networks. Cardiovascular Prevention and Pharmacotherapy, 3(3), 64-72.

[73]. Shreyas, D. K., Joshi, S. N., Kumar, V. H., Venkataramanan, V., & Kaliprasad, C. S. (2023). A Review on Neural Networks and its Applications. Journal of Computer Technology & Applications, 14(2), 59-70p.

[74]. Gonog, L., & Zhou, Y. (2019, June). A review: generative adversarial networks. In 2019 14th IEEE conference on industrial electronics and applications (ICIEA) (pp. 505-510). IEEE.

[75]. Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H. I. K., Frisky, A. Z. K., & ul Haq, Q. M. (2022). Understanding of convolutional neural network (cnn): A review. International Journal of Robotics and Control Systems, 2(4), 739-748.

[76]. Flora GD, Nayak MK. A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes. Curr Pharm Des. 2019;25(38):4063-4084. doi: 10.2174/1381612825666190925163827. PMID: 31553287.

[77]. Kundu, J., & Kundu, S. (2022). Cardiovascular disease (CVD) and its associated risk factors among older adults in India: evidence from LASI Wave 1. Clinical Epidemiology and Global Health, 13, 100937.

[78]. Olvera Lopez E, Ballard BD, Jan A. Cardiovascular Disease. [Updated 2023 Aug 22]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK535419/

[79]. Vaccarino, V., Bremner, J.D. Stress and cardiovascular disease: an update. Nat Rev Cardiol 21, 603–616 (2024). https://doi.org/10.1038/s41569-024-01024-y

[80]. Roth, G, Mensah, G, Johnson, C. et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. JACC. 2020 Dec, 76 (25) 2982–3021.https://doi.org/10.1016/j.jacc.2020.11.010

[81]. Guttmacher, Alan E.; Collins, Francis S.; Nabel, Elizabeth G. . (2003). Cardiovascular Disease. New England Journal of Medicine, 349(1), 60–72. doi:10.1056/NEJMra035098

[82]. Desai DS, Hajouli S. Arrhythmias. [Updated 2023 Jun 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK558923/

[83]. Ciarambino T, Menna G, Sansone G, Giordano M. Cardiomyopathies: An Overview. Int J Mol Sci. 2021 Jul 19;22(14):7722. doi: 10.3390/ijms22147722. PMID: 34299342; PMCID: PMC8303989.

[84]. Dudley, P. (2003). Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease. Arterioscler Thromb Vasc Biol, 23, 1319-1321. Chicago

[85]. Virani SS, Alonso A, Benjamin EJ, et al.: Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation. 2020, 141:e139-596. 10.1161/CIR.0000000000000757

[86]. Yuyun MF, Sliwa K, Kengne AP, Mocumbi AO, Bukhman G: Cardiovascular diseases in Sub-Saharan Africa compared to high-income countries: an epidemiological perspective. Glob Heart. 2020, 15:15. 10.5334/gh.403.

[87]. Khosravi A, Akhavan Tabib A, Golshadi I, et al.: The relationship between weight and CVD risk factors in a sample population from central Iran (based on IHHP). ARYA Atheroscler. 2012, 8:82-89 .

[88]. Klakk H, Kristensen PL, Andersen LB, Froberg K, Møller NC, Grøntved A: Symptoms of depression in young adulthood is associated with unfavorable clinical- and behavioral cardiovascular disease risk factors. Prev Med Rep. 2018, 11:209-15. 10.1016/j.pmedr.2018.05.017

[89]. Hayes L, White M, Unwin N, Bhopal R, Fischbacher C, Harland J, Alberti KG: Patterns of physical activity and relationship with risk markers for cardiovascular disease and diabetes in Indian, Pakistani, Bangladeshi and European adults in a UK population. J Public Health Med. 2002, 24:170-8. 10.1093/pubmed/24.3.170

[90]. Sonel AF, Good CB, Mulgund J, et al.: Racial variations in treatment and outcomes of black and white patients with high-risk non-ST-elevation acute coronary syndromes: insights from CRUSADE (Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes With Early Implementation of the ACC/AHA Guidelines?). Circulation. 2005, 111:1225-32. 10.1161/01.CIR.0000157732.03358.64

[91]. Xavier D, Pais P, Devereaux PJ, et al.: Treatment and outcomes of acute coronary syndromes in India (CREATE): a prospective analysis of registry data. Lancet. 2008, 371:1435-42. 10.1016/S0140-6736(08)60623-6

[92]. Nadeem M, Ahmed SS, Mansoor S, Farooq S: Risk factors for coronary heart disease in patients below 45 years of age. Pak J Med Sci. 2013, 29:91-6. 10.12669/pjms.291.2828

[93]. Farrag A, Eraky AE, Aroussy WE, et al.: Obesity and other cardiovascular risk factors in Egyptian University students: magnitude of the problem. Epidemiology. 2015, 5:181.

[94]. Gupta R, Deedwania PC, Sharma K, et al.: Association of educational, occupational and socioeconomic status with cardiovascular risk factors in Asian Indians: a cross-sectional study. PLoS One. 2012, 7:e44098. 10.1371/journal.pone.0044098

[95]. Sani MU, Wahab KW, Yusuf BO, Gbadamosi M, Johnson OV, Gbadamosi A: Modifiable cardiovascular risk factors among apparently healthy adult Nigerian population - a cross sectional study. BMC Res Notes. 2010, 3:11. 10.1186/1756-0500-3-11

[96]. May AL, Kuklina EV, Yoon PW: Prevalence of cardiovascular disease risk factors among US adolescents, 1999-2008. Pediatrics. 2012, 129:1035-41. 10.1542/peds.2011-1082

[97]. Weiss E, Japie C, Balahura AM, Bartos D, Badila E: Cardiovascular risk factors in a Roma sample population from Romania. Rom J Intern Med. 2018, 56:193-202. 10.2478/rjim-2018-0010

[98]. Lee S, Shafe A, Cowie MR: UK stroke incidence, mortality and cardiovascular risk management 1999e2008: time-trend analysis from the General Practice Research Database. BMJ Open. 2011, 1:e000269.

[99]. Serhani MA, T El Kassabi H, Ismail H, Nujum Navaz A. ECG Monitoring Systems: Review, Architecture, Processes, and Key Challenges. Sensors (Basel). 2020 Mar 24;20(6):1796. doi: 10.3390/s20061796. PMID: 32213969; PMCID: PMC7147367.

[100]. Omerovic S, Jain A. Echocardiogram. [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK558940/

[101]. Lima, J. A., & Desai, M. Y. (2004). Cardiovascular magnetic resonance imaging: current and emerging applications. *Journal of the American College of Cardiology*, *44*(6), 1164-1171.

[102]. Sureshkumar A, Hansen B, Ersahin D. Role of Nuclear Medicine in Imaging. Semin Ultrasound CT MR. 2020 Feb;41(1):10-19. doi: 10.1053/j.sult.2019.10.005. Epub 2019 Oct 14. PMID: 31964489.

[103]. Kharabsheh SM, Al-Sugair A, Al-Buraiki J, Al-Farhan J. Overview of exercise stress testing. Ann Saudi Med. 2006 Jan-Feb;26(1):1-6. doi: 10.5144/0256-4947.2006.1. PMID: 16521867; PMCID: PMC6078558.

[104]. Tavakol M, Ashraf S, Brener SJ. Risks and complications of coronary angiography: a comprehensive review. Glob J Health Sci. 2012 Jan 1;4(1):65-93. doi: 10.5539/gjhs.v4n1p65. PMID: 22980117; PMCID: PMC4777042.

[105]. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010 Nov;5(6):463-6. doi: 10.1097/COH.0b013e32833ed177. PMID: 20978388; PMCID: PMC3078627.

[106]. Naser, M.A.; Majeed, A.A.; Alsabah, M.; Al-Shaikhli, T.R.; Kaky, K.M. A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. Algorithms 2024, 17, 78. https://doi.org/10.3390/a17020078

[107]. Kumamaru KK, Hoppel BE, Mather RT, Rybicki FJ. CT angiography: current technology and clinical use. Radiol Clin North Am. 2010 Mar;48(2):213-35, vii. doi: 10.1016/j.rcl.2010.02.006. PMID: 20609871; PMCID: PMC2901244.

[108]. Galli A, Ambrosini F, Lombardi F. Holter Monitoring and Loop Recorders: From Research to Clinical Practice. Arrhythm Electrophysiol Rev. 2016 Aug;5(2):136-43. doi: 10.15420/AER.2016.17.2. PMID: 27617093; PMCID: PMC5013174.

[109]. Ismail SNA, Nayan NA, Jaafar R, May Z. Recent Advances in Non-Invasive Blood Pressure Monitoring and Prediction Using a Machine Learning Approach. Sensors (Basel). 2022 Aug 18;22(16):6195. doi: 10.3390/s22166195. PMID: 36015956; PMCID: PMC9412312.

[110]. Jacob R, Khan M. Cardiac Biomarkers: What Is and What Can Be. Indian J Cardiovasc Dis Women WINCARS. 2018 Dec;3(4):240-244. doi: 10.1055/s-0039-1679104. PMID: 31934672; PMCID: PMC6957084.

[111]. Tiwari R, Kumar R, Malik S, Raj T, Kumar P. Analysis of Heart Rate Variability and Implication of Different Factors on Heart Rate Variability. Curr Cardiol Rev. 2021;17(5):e160721189770. doi: 10.2174/1573403X16999201231203854. PMID: 33390146; PMCID: PMC8950456.

[112]. Cirino AL, Ho CY. Genetic testing for inherited heart disease. Circulation. 2013 Jul 2;128(1):e4-8. doi: 10.1161/CIRCULATIONAHA.113.002252. PMID: 23817488; PMCID: PMC3769178.

[113]. Manda YR, Baradhi KM. Cardiac Catheterization Risks and Complications. [Updated 2023 Jun 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK531461/

[114]. Albouaini K, Egred M, Alahmar A, Wright DJ. Cardiopulmonary exercise testing and its application. Postgrad Med J. 2007 Nov;83(985):675-82. doi: 10.1136/hrt.2007.121558. PMID: 17989266; PMCID: PMC2734442.

[115]. O'Rourke MF, Pauca A, Jiang XJ. Pulse wave analysis. Br J Clin Pharmacol. 2001 Jun;51(6):507-22. doi: 10.1046/j.0306-5251.2001.01400.x. PMID: 11422010; PMCID: PMC2014492.

[116]. Chowdhury MEH, Khandakar A, Alzoubi K, Mansoor S, M Tahir A, Reaz MBI, Al-Emadi N. Real-Time Smart-Digital Stethoscope System for Heart Diseases Monitoring. Sensors (Basel). 2019 Jun 20;19(12):2781. doi: 10.3390/s19122781. PMID: 31226869; PMCID: PMC6630694.

[117]. Reddy, M. (2024, June 1). *Digital Transformation in Healthcare in 2025: 7 Key Trends*. Retrieved from Digitalauthority: https://www.digitalauthority.me/resources/state-of-digital-transformation-healthcare/

[118]. Thimbleby H. Technology and the future of healthcare. J Public Health Res. 2013 Dec 1;2(3):e28. doi: 10.4081/jphr.2013.e28. PMID: 25170499; PMCID: PMC4147743.

[119]. Griebel, L., Prokosch, HU., Köpcke, F. et al. A scoping review of cloud computing in healthcare. BMC Med Inform Decis Mak 15, 17 (2015). https://doi.org/10.1186/s12911-015-0145-7

[120]. Dang LM, Piran MJ, Han D, Min K, Moon H. A Survey on Internet of Things and Cloud Computing for healthcare electronics 2019 8(7):768. https://doi.org/10.3390/electronics8070768

[121]. Mekawie, N., & Yehia, K. (2021). Challenges of deploying cloud computing in eHealth. Procedia Computer Science, 181, 1049-1057.

[122]. Ali, O., Shrestha, A., Soar, J., & Wamba, S. F. (2018). Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review. International Journal of Information Management, 43, 146-158.

[123]. S. M. Riazul Islam, Daehan Kwak, M. Humaun Kabir, M. Hossain, and Kyung-Sup Kwak, "The internet of things for health care: a comprehensive survey," IEEE Access, vol. 3, pp. 678–708, 2015.

[124]. Sinha, S. (2024, September 3). *State of IoT 2024: Number of connected IoT devices growing 13% to 18.8 billion globally*. Retrieved from IoT Analytics: https://iot-analytics.com/number-connected-iot-devices/

[125]. Dash, S., Shakyawar, S.K., Sharma, M. et al. Big data in healthcare: management, analysis and future prospects. J Big Data 6, 54 (2019). https://doi.org/10.1186/s40537-019-0217-0

[126]. Desai, F., Chowdhury, D., Kaur, R., Peeters, M., Arya, R. C., Wander, G. S., ... & Buyya, R. (2022). HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing. Internet of Things, 17, 100485.

[127]. Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. IEEE Access, 9, 106575-106588.

[128]. Nancy, A. A., Ravindran, D., Vincent, D. R., Srinivasan, K., & Chang, C. Y. (2023). Fog-Based Smart Cardiovascular Disease Prediction System Powered by Modified Gated Recurrent Unit. Diagnostics, 13(12), 2071.

[129]. Subramani, S., Varshney, N., Anand, M. V., Soudagar, M. E. M., Al-Keridis, L. A., Upadhyay, T. K., ... & Rohini, K. (2023). Cardiovascular diseases prediction by machine learning incorporation with deep learning. Frontiers in Medicine, 10, 1150933.

[130]. K. Karthick, S. K. Aruna, Ravi Samikannu, Ramya Kuppusamy, Yuvaraja Teekaraman, Amruth Ramesh Thelkar, "Implementation of a Heart Disease Risk Prediction Model Using Machine Learning", Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 6517716, 14 pages, 2022. https://doi.org/10.1155/2022/6517716.

[131]. Sumwiza, K., Twizere, C., Rushingabigwi, G., Bakunzibake, P., & Bamurigire, P. (2023). Enhanced cardiovascular disease prediction model using random forest algorithm. Informatics in Medicine Unlocked, 41, 101316.

[132]. Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, 16(2), 88.

[133]. Abdullah Alqahtani, Shtwai Alsubai, Mohemmed Sha, Lucia Vilcekova, Talha Javed, "Cardiovascular Disease Detection using Ensemble Learning", Computational Intelligence and Neuroscience, vol. 2022, Article ID 5267498, 9 pages, 2022. https://doi.org/10.1155/2022/5267498

[134]. Gupta, Nishant; Ahuja, Naman; Malhotra, Shikhar; Bala, Anju; Kaur, Gurleen (2017). Intelligent heart disease prediction in cloud environment through ensembling. Expert Systems, (), e12207–. doi:10.1111/exsy.12207

[135]. N. Sahanaa Sree and N. Banupriya, “A cloud based risk prediction of coronary heart disease,” Int. J. Appl. Eng., vol. 13, no. 5, pp. 2786–2790, 2018.

[136]. M. A. Khan, S. Abbas, A. Atta, A. Ditta, H. Alquhayz, M. F. Khan, et al., “Intelligent cloud based heart disease prediction system empowered with supervised machine learning,” Comput. Mater. Continua, vol. 65, no. 1, pp. 139–151, 2020.10.32604/cmc.2020.011416.

[137]. P. Verma and S. K. Sood, “Cloud-centric IoT based disease diagnosis healthcare framework,” J. Parallel Distrib. Comput., vol. 116, pp. 27–38, 2018. 10.1016/j.jpdc.2017.11.018.

[138]. R. Ani, S. Krishna, N. Anju, M. S. Aslam, and O. S. Deepa, “Iot based patient monitoring and diagnostic prediction tool using ensemble classifier,” 2017 International Conference on Advances in Computing, Communication, 2017.10.1109/ICACCI.2017.8126068

[139]. R. Venkatesh, C. Balasubramanian, and M. Kaliappan, “Development of big data predictive analytics model for disease prediction using machine learning technique,” J. Med. Syst., vol. 43, p. 272, 2019.

[140]. Adhikary D, Barman S, Ranjan R, Stone H. A Systematic Review of Major Cardiovascular Risk Factors: A Growing Global Health Concern. Cureus. 2022 Oct 10;14(10):e30119. doi: 10.7759/cureus.30119.

[141]. S. Mohan, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” IEEE Access, vol. 7, pp. 81542–81554, 2019.10.1109/ACCESS.2019.2923707.

[142]. A. Mohammed, R. Basa, A. K. Kuchuru, S. P. Nandigama, and M. Gangolla, “Random forest machine learning technique to predict heart disease,” Eur. J. Mol. Clin. Med., vol. 7, no. 4. p. 2020, 2020.

[143]. P. Singh, A. Kaur, R. S. Batth, S. Kaur, and G. Gianini, “Multi-disease big data analysis using beetle swarm optimization and an adaptive neuro-fuzzy inference

system," Neural Comput. Appl., vol. 33, no. 16, pp. 1–12, 2021.10.1007/s00521-021-05798-x.

[144]. R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, "Feature analysis of coronary artery heart disease data sets," Procedia Comput. Sci., vol. 65, pp. 459–468, 2015. 10.1016/j.procs.2015.09.132.

[145]. S. Ghosh, Application of various data mining techniques to classify heart diseases, Doctoral dissertation, Dublin, National College of Ireland, 2017.

[146]. D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," SN Comput. Sci., vol. 1, no. 6, pp. 1–6, 2020.10.1007/s42979-020-00365-y.

[147]. Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, vol. 2018, Article ID 3860146, 21 pages, 2018. https://doi.org/10.1155/2018/3860146.

[148]. "How machine learning is driving innovation in cardiovascular disease research?" "https://www.broadinstitute.org/how-machine-learning-driving-innovation-cardiovascular-disease-research ", Accessed on: 22 April, 2024.

[149]. Australian Institute of Health and Welfare. (2024). Heart, stroke and vascular disease: Australian facts. Retrieved from https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/hsvd-facts

[150]. Staff, M. C. (2023, August 17). *Strategies to prevent heart disease*. Retrieved from Mayoclinic: https://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/heart-disease-prevention/art-20046502.

[151]. Bays, H. E., Taub, P. R., Epstein, E., Michos, E. D., Ferraro, R. A., Bailey, A. L., ... & Toth, P. P. (2021). Ten things to know about ten cardiovascular disease risk factors. American journal of preventive cardiology, 5, 100149.

[152]. Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics, 14(2), 144.

[153]. Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.

[154]. https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

[155]. https://tools.acc.org/ascvd-risk-estimator-plus/#!/calculate/estimate/

[156]. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary. Journal of the American College of Cardiology, 2019;74(10):1376–1414.DOI: 10.1016/j.jacc.2019.03.009

# LIST OF PUBLICATIONS

## Journal Publications


1. Shilpa, Kaur, Tarandeep, "Data Analysis using Association Rule in Data Mining", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.5, Issue 11, page no.316-319, November 2018.
2. Shilpa, Kaur, Tarandeep and Garg, Rachit. "Digital healthcare: A topical and futuristic review of technological and robotic revolution" Paladyn, Journal of Behavioral Robotics, vol. 14, no. 1, 2023, pp. 20220108. https://doi.org/10.1515/pjbr-2022-0108. (Published July 26, 2023) (SCOPUS indexed)


## Conference Publications


1. Shilpa, Kaur, T. (2022). Digital Healthcare: Current Trends, Challenges and Future Perspectives. Proceedings of the Future Technologies Conference (FTC) 2021, Volume 2. FTC 2021. Lecture Notes in Networks and Systems, vol 359. Springer, pp 645–661. https://doi.org/10.1007/978-3-030-89880-9_48 (Published 04 November 2021, Springer) (Scopus Indexed).
2. Shilpa, Kaur, T. (2022). Blockchain and Cloud Technology: Leading the ICT Innovations. ICT Systems and Sustainability. Lecture Notes in Networks and Systems, vol 321. Springer, Singapore. https://doi.org/10.1007/978-981-16-5987-4_41. (Published 04 January 2022, Springer) (Scopus Indexed).


## Accepted Papers


1. HBCIoT: A Healthcare Provisioning Model Using Blockchain and Cloud Internet of Things (Conference: Digital Transformation for Business Sustainability & Growth ) , Mittal School of Business, Lovely Professional University. (August, 2024)

## Communicated Papers

1. Cardiovascular Disease and Covid-19: Unveiling Shared Risk Factors and the Role of Technology in Advancing Healthcare Innovation for Sustainable Development. (Discover Sustainability) Revision Submitted.
2. Healthcare-as-a-Service Provisioning using cloud-of-things: A Contemporary Review of Existing Frameworks Based on Tools, Services and Diseases ( Scalable computing Indexed: Scopus ) Revision Submitted.
3. PCard: A Machine Learning Model for Predicting Cardiovascular Diseases and Associated Advisories (Journal: International Journal of Information Engineering and Electronic Business). December 2024 (Revision Submitted)

## COPYRIGHT

Shilpa, Dr. Tarandeep Kaur, Lovely Professional University, “A cardiovascular disease prediction and recommendation model using an ensemble approach.” (Published Oct, 2024)

## Book Chapter

Kaur, Harjinder & Shilpa. (2025). AI-driven healthcare: Transformation and advancement. In Citizen-centric artificial intelligence for smart cities (pp. 139–164). IGI Global. https://doi.org/10.4018/979-8-3693-7832-8.ch006

# List of Workshops and FDPs

- Two-Day Faculty Development Programme on Technology and Healthcare: Approaches and Applications from 19-20 December 2022, organized by the Department of Computer Applications, Chitkara University, Punjab.
- Faculty Development Program on Cloud Computing: From Basics to Implementation, organized by Lovely Professional University w.e.f. December 20, 2023, to January 03, 2024 (30 hours).
- Short Term Course on Bits 'n Bytes: Python-Powered Business Analytics organized by Lovely Professional University w.e.f September 23, 2024 to September 28, 2024