International journal of Inventive Research in Science and Technology ISSN: 2583-5637 (Online)

Volume 4 Issue 11 November 2025

A PREDICTION OF HEART DISEASE BASED ON ARTIFICIAL INTELEGENT AND MACHINE LEARNING ALGORITHM

PRIYANKA
M.Tech Schollar,,MIT,Bhopal
er.priyanka1923@gmail.com

Prof. Preeti Mishra
Asst. Professor MIT,Bhopal
pritypreet85@gmail.com

ABSTRACT :- Artificial intelligence (AI) is simulating human intelligence processes by machines and software simulators to help humans in making accurate, informed, and fast decisions based on data analysis. The medical field can make use of such AI simulators because medical data records are enormous with many overlapping parameters. Using in-depth classification techniques and data analysis can be the first step in identifying and reducing the risk factors. In this research, we are evaluating a dataset of cardiovascular abnormalities affecting a group of potential patients. The modern era's busy schedule often leads to an unhealthy lifestyle, causing anxiety and depression. This work aims to detect heart disease symptoms at an early stage to reduce the risk of heart attacks and improve performance parameters using different machine learning techniques. As now a day's machine learning algorithm is very popular to solve real time problems in an efficient manner, mostly research uses machine learning techniques in every field like healthcare, security, and pattern recognition. In this research work, conducted experimental work using the Kaggle datasets within the Python framework. Here applied several machine learning techniques, including LR, DT, SVM, NB, and k-NN. All these classification techniques were used as binary classifiers, where the target variable had two values, 0 and 1. The simulation results indicate that the LR (Logistic Regression) demonstrated superior accuracy and performance parameter values, like accuracy and other parameters.

KEYWORDS:- Supervised Learning, Heart Disease, Machine Learning, Accuracy, Classification techniques, Neural Network, Artificial Intelligence.

1.Introduction

The human body contains various organs the heart is one of them, heart is a vital organ of the human body because it is functions to provide and pumping blood to all other remaining human body parts, therefore human life is completely dependent on the efficient operation of the heart. If the heart is not functioning properly, it will affect the other human body organs, such as the mind and kidneys with overall human body organs. Technology is increasing day by day everywhere; the healthcare sector is one of them. In the healthcare sector, artificial intelligence Approach with their subset like machine learning and, deep learning approaches are help‟s the various disease diagnosis systems for patients.

AI applications are considered to play a key role in further establishing and supporting a decentralized rehabilitation model in which intelligent connected tools will be employed to assist clinical decision-making, and health outcomes monitoring. Many AI-based methods and solutions have been proposed in recent years to support the future challenge of enabling assisted physical therapy and assessments in a minimally supervised and decentralized manner, ideally at the patient‟s home. However, to the best of the authors‟ knowledge, no published works provided a comprehensive review of machine learning methods and applications used for remote monitoring and assistance in the rehabilitation context. Some existing works in literature have provided an overview of the role of machine learning algorithms combined with specific technologies used for rehabilitation issues, such as wearable sensors and vision-based motion capture technologies.

Clinical Research on AI and Machine Learning Applications The evaluation of progress has its own set of problems. In traditional clinical research, when progress takes the form of a new drug for a definable condition, the standards for testing and accepting the drug as an advance are well established. When the intervention is an AI and machine-learning algorithm rather than a drug, the medical community expects the same level of surety, but the standards for describing and testing AI and machine-learning interventions are far from clear. What are the standards to which AI and

machine learning–based interventional research should be held, if an app is going to be accepted as the standard that will shape, reform, and improve clinical practice? That research has three components. First, the research must be structured to answer a clinically meaningful question in a way that can influence the behavior of the health professional and lead to an improvement in outcomes for a patient. Second, the intervention must be definable, scalable, and applicable to the problem at hand. It must not be influenced by factors outside the domain of the problem and must yield outcomes that can be applied to similar clinical problems across a wide range of populations and disease prevalence‟s. Can AI and machine learning–driven care meet these standards ones that we demand from a novel therapeutic intervention or laboratory-based diagnostic test or do we need to have a unique set of standards for this type of intervention? Third, when the results of the research are applied in such a way as to influence practice, the outcome must be beneficial for all patients under consideration, not just those who are similar to the ones with characteristics and findings on which the algorithm was trained. This raises the question of whether such algorithms should include consideration of public health (i.e., the use of scarce resources) when diagnostic or treatment recommendations are being made and the extent to which such considerations are part of the decision-making process of the algorithm. Such ethical considerations have engaged health professionals and the public for centuries. Use of AI and Machine-Learning Applications in Conducting Clinical Research AI and machine learning have the potential to improve and possibly simplify and speed up clinical trials through both more efficient recruitment and matching of study participants and more comprehensive analyses of the data.

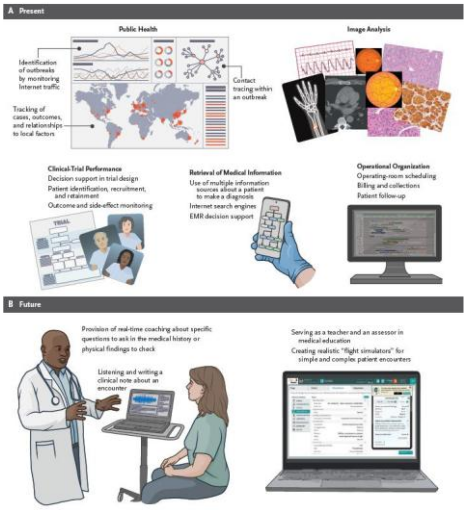

Figure 1.1: Spectrum of Artificial Intelligence (AI)

1.2 CLINICAL APPLICATIONS

Besides technical requirements, it is fundamental to integrate clinical knowledge into technologies to be used at home in a minimally supervised way. Hence, the role of AI in clinical practice should be to provide a combination of medical, psychological, and technical knowledge in the form of embedded algorithms analyzing and processing online the data generated by digital technologies. AI is expected to play a crucial role in clinical decision-making, the online adaptation of therapy exercises, and the monitoring of progress through the extraction of validated assessment scores. Based on the past research work, three main approaches have been introduced to manage the remote monitoring and evaluation of physical rehabilitation therapies by means of machine learning applications. These approaches could be categorized as activity recognition, movement classification, and clinical functional assessment. Machine learning algorithms address activity recognition tasks aimed to identify specific rehabilitation actions to remotely track patients‟ adherence to the prescribed therapy and sometimes to measure the treatment outcomes based on patients‟ activity in real-world life. Most of the reviewed papers covered the problem of recognizing activities for specific body parts like strength training exercises for upper and lower limb exercises to improve range of motion, flexibility and balance exercises, specific rehabilitation exercise for upper limb and lower impaired limb in stroke, and shoulder impairments. Movement classification approaches aimed to assess the quality of rehabilitation exercises performed remotely. Algorithms addressing movement classification tasks have been used to evaluate exercise performance in terms of the well and poorly-executed tasks.

1.3 IoT HELATH CARE

Health information technology is one of today‟s fastest-growing and most powerful technologies.This technology

is used predominantly for predicting illness and obtaining medications quickly because visiting a doctor and performing pathological tests can be time-consuming and expensive. This has prompted many researchers to contribute by developing new disease prediction systems or improving existing ones. In the modern era, many solutions are related to technology. The same is true for healthcare and medication. According to 2022 statistics, downloads of health and fitness mobile applications have increased from 488 million in 2019 to 656 million in recent years. As there are 5.31 billion unique mobile users worldwide, using mobile applications for medical care seems to be an efficient option. It can save time and money and reduce the percentage of health issues and sudden deaths. Today, doctors and IT developers are helping improve the health conditions of residents of developed and underdeveloped countries.

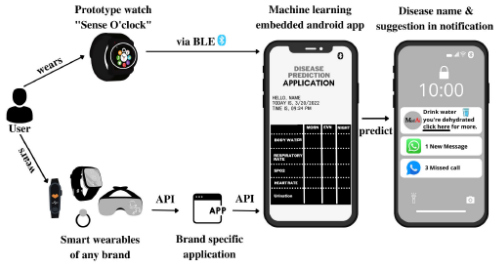

Figure 1.2: Schematic diagram of the system.

Various mobile applications developed in different parts of the world were evaluated during this research. Although a few could be found that perform prediction using a pre-trained machine learning model, most other applications provide basic health monitoring functionalities. The functional features and feasibility of these applications were tested and compared with those of the „„„MedAi"""" application. The sudden deaths of those close to us are painful and unexpected. Undiagnosed illnesses such as advanced cancer; abrupt natural causes such as heart attack, brain hemorrhage, and cot death; and major illnesses such as epilepsy are some of the leading causes of such deaths.

2. LITERATURE REVIEW

2.1 PREVIOUS WORK DONE

[1] In this chapter, the significance of AI, the typology, and the evolution of AI has been introduced and its application in healthcare are discussed with a special focus on cardiovascular management. The ML and DL techniques are explained in detail with supervised and unsupervised learning along with their subtypes ANN, RNN, CNN, and Auto-encoders as well as symbolic learning, natural language processing, reinforcement learning, cognitive and context-aware computing. With the vast amount of information, AI is expected to understand analysis and solve more complicated issues closer to actual clinical questions leading to the better decision for disease management. Researchers have started venturing AI in the entire field for achieving higher success. AI has an important role in healthcare and cardiovascular diseases solving complex problems of individuals by constant training with the data. AI contributes significantly in healthcare with an objective to improve the quality of human life and minimize errors with the support and expertise of healthcare professionals.

[2] Coronary Artery Disease (CAD) prediction is a very hard and challenging task in the medical field. The early prediction in the medical field especially the cardiovascular sector is one of the virtuosi. The prior studies about the construction of the early prediction model developed an understanding of the recent techniques to find the variation in medical imaging. The prevention of cardiovascular can be fulfilled through a diet chart prepared by the concerned physician after early prediction. Our research paper consists of the prediction of CAD by the proposed algorithm by constructing of pooled area curve (PUC) in the machine learning method. This knowledge-based identification is an important factor for accurate prediction. This significant approach provides a good impact to determine variation in medical images although weak pixels surrounding it. This pooled area construction in our machine learning algorithm is bagging shrinking veins and tissues with the help of clogging and plaque of blood vessels. Besides, the noisy type database is used in this article for better clarity about identifying the classifier. This research article provides the recent adaptive image-based classification techniques and it comparing existing classification methods to predict CAD earlier for a higher accurate value.

[3] In this paper prediction of heart disease using the UCI machine learning data set at Cleveland repository is analysed using random forest regression model and decision tree regression models. They find random forest regression model provides better accuracy than decision tree model. They have calculated the performance measures namely MAE, RMSE and MSE and it was observed that MAE, MSE and RMSE for the male is more than the female for the given attribute in the data set. They find female are more affected to heart disease than male for the values. From the analysis of random forest regression, they find the prediction model performs better for c_slope attribute value in the data set in terms of MAE, MSE and RMSE.

[4] In this paper they have developed and evaluated ML DKD screening tools using data easily accessible in the EHR, which provide a robust method of predicting DKD within a 5-year window for patients at the time of their T2DM diagnosis. Their MLA models, which use only demographics, clinical measurements, laboratory measurements, and patient history drawn from the EHR outperform the CDC CKD scoring system. Urinary albumin is commonly used for kidney disease diagnosis, however it is not routinely collected data for all patients. Therefore, to

enable screening for DKD on a broad patient population, it was not included as an input. Data for the MLAs can be automatically extracted from the EHR, removing the burden of manually calculating CKD risk assessment with the CDC CKD scoring system. These algorithms may provide warning of DKD to physicians for improved patient care by determining who is at high risk and allow for earlier detection and intervention.

[5] This paper has organized a heart disease-related dataset from UCI repository. The organized dataset describes variables correlations with class-level target variables. This experiment has analyzed the variables by different machine learning algorithms. The authors have considered prediction-based previous work and finds some machine learning algorithms did not properly work or do not cover 100% classification accuracy with overfitting, underfitting, noisy data, residual errors on base level decision tree. This research has used Pearson correlation and chi-square features selection-based algorithms for heart disease attributes correlation strength. The main objective of this research to achieved highest classification accuracy with fewer errors. So, the authors have used parallel and sequential ensemble methods to reduce above drawback in prediction. The parallel and serial ensemble methods were organized by J48 algorithm, reduced error pruning, and decision stump algorithm decision tree-based algorithms. This paper has used random forest ensemble method for parallel randomly selection in prediction and various sequential ensemble methods such as AdaBoost, Gradient Boosting, and XGBoost Meta classifiers.

[6] Machine learning (ML) is a rapidly developing field in today's world. Use machine learning to extract data from a wide variety of sources. ML can solve various problems based on complex data sets. The prediction of heart disease is the most complex task in the medical field. It cannot be observed with the naked eye, it can appear immediately anywhere, anytime. Many ML algorithms are more capable of handling various algorithms. Due to complexity, the processing of massive data sets is more complicated. By improving these systems, the quality of medical diagnosis decisions can be improved. They can find patterns hidden in large amounts of data that will avoid the use of traditional statistical methods for analysis. In this article, An Enhanced New Dynamic Data Processing (ENDDP) Algorithm is developed to predict the early stages of heart disease. The results prove the performance of the proposed system.

[7] The heart, like a pump, is an organ about the size of a fist, mainly composed of muscle and connective tissue that functions to distribute blood to tissues. The heart is located under the rib cage, above the diaphragm between the lungs, slightly closer to the left. Sometimes a small, unexpected problem with the veins or the valves that supply the heart affects a person's life and can lead to death. Early diagnosis is essential to predict diseases that affect the human heart and lead people to live another period of life. In this context, the authors introduce two methods for early diagnosis of heart disease, the support vector machine and artificial neural network (ANN). The medical data is taken from the University of California Irvine (UCI) Machine Learning Repository database, and it contains reports of 170 people. The investigation results confirm that the optimal execution is the support vector machine technique. It gives high-accuracy prediction results. As for the performance of the forward propagation artificial neural networks technique is acceptable.

8] A stroke is caused when blood flow to a part of the brain is stopped abruptly. Without the blood supply, the brain cells gradually die, and disability occurs depending on the area of the brain affected. Early recognition of symptoms can significantly carry valuable information for the prediction of stroke and promoting a healthy life. In this research work, with the aid of machine learning (ML), several models are developed and evaluated to design a robust framework for the longterm risk prediction of stroke occurrence. The main contribution of this study is a stacking method that achieves a high performance that is validated by various metrics, such as AUC, precision, recall, F-measure and accuracy.

[9] As cardiovascular is one of the major causes for death there are some data analytical techniques that predicts the occurrence of cardiovascular disease. It can be achieved through selecting a correct combination of prediction models and features. Prediction models were developed using different classification techniques based on feature selection and there are certain algorithms which provide varied and improved accuracy. Here prediction model is developed using Random Forest classification technique - Method for classification, regression by constructing a multitude of decision trees at training time. Developed by aggregating tree Avoids over fitting can deal with large number of features.

[10] Major chronic diseases such as cardiovascular disease (CVD), diabetes, and cancer impose a significant burden on people and health care systems around the globe. Recently, deep learning (DL) has shown great potential for the development of intelligent mobile health (mHealth) interventions for chronic diseases that could revolutionize the delivery of health care anytime, anywhere. The aim of this study is to present a systematic review of studies that have used DL based on mHealth data for the diagnosis, prognosis, management, and treatment of major chronic diseases and advance our understanding of the progress made in this rapidly developing field. A search was conducted on the bibliographic databases Scopus and

PubMed to identify papers with a focus on the deployment of DL algorithms that used data captured from mobile devices (eg, smartphones, smartwatches, and other wearable devices) targeting CVD, diabetes, or cancer. The identified studies were synthesized according to the target disease, the number of enrolled participants and their age, and the study period as well as the DL algorithm used, the main DL outcome, the data set used, the features selected, and the achieved performance.

2.2 EXISTING WORK

Machine learning plays a vital role in different application nowadays, the healthcare sector is one of them, in this section we present the literature review work for healthcare and disease diagnosis system using machine learning and other techniques, here below table presents comparative literature work based on the machine learning techniques and different performance parameter evaluation.

2.3 PROBLEM STATEMENT

This study deals with the application of machine learning techniques in medical domain. The main objective of this study is to extract pattern and generate precious rules from critical disease patient data. Here we improve the classification rate in the terms of accuracy for each dataset; therefore we can enhance the results which are generated from previous techniques in the medical science domain. To generate precious classification from real world medical data which are may be useful to decision making for doctors and patients, To create a pattern or model from past data of patient, To identify issues and challenges in data mining as applied to the medical practice, To evaluate data mining methods for medical database, Build and test models using python software and used reputed dataset form UCI machine learning Repository/Kaggle datasets.

3. WIRELESS COMMUNICATION

3.1 OVERVIEW

AI is an innovation in the replication of human intelligence in machine form. AI requires customized hardware and software to develop and train ML algorithms using programming languages like Python, R, and Java. This system works by considering labeled training data, analyzing it for correlations, recognizing, patterns, and making predictions. AI focuses on three cognitive skills: learning acquire data for developing algorithms, reasoning to obtain the desired outcome by selecting the correct algorithm, and self-correction-continuously optimizing the algorithms. Some applications of AI are machine vision, speech recognition, natural language processing, and expert systems.

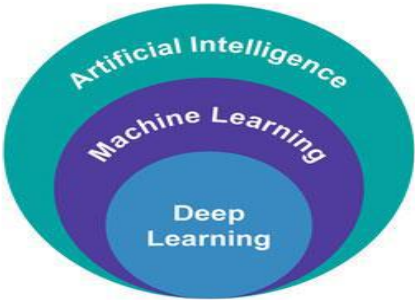

Figure 3.1: Different hierarchies of artificial intelligence techniques.

The market value of AI was globally calculated as USD 62.35 billion (2020) and is projected to increase at a compound annual growth rate (CAGR) of 40.2% (2021–2028). AI is the technological advancement of computing systems that are capable of performing tasks that include decision-making, speech recognition, visual perception, and language translation. It utilizes multiple layers of algorithms to understand, visualize, and process information that are involved in data processing, calculation, and automated reasoning. The algorithms are improved on a daily aspect, as conventional algorithms are doubted on their accuracy and efficiency which had led to the development of standard algorithms. The continuous research and innovation by various industry rulers had bought in advanced technologies in the area of automotive, healthcare, retail, finance, and manufacturing. AI has owned its place with self-driving vehicles to crucial life-saving medical gear by being infused virtually into every machine and program. Leading technology companies like Amazon, Google, Apple, Facebook, International Business Machines Corporation (IBM), and Microsoft are capitalizing on AI and are working intensively to create AI more accessible for enterprise applications. The main purpose of AI is to help humans with human capabilities to make advanced decisions with better consequences. From a philosophical perspective, AI makes humans life meaningful, free from hard labor, and help to manage the complex connection of individuals, companies, states, and nations to function in a beneficial way. Currently, AI is employed to improve process efficiencies, automate resource heavy tasks, and make business predictions. The research and development cost related to the system components of AI must be subsidized by the corporations and government agencies for enhancing the development of AI in different applications.

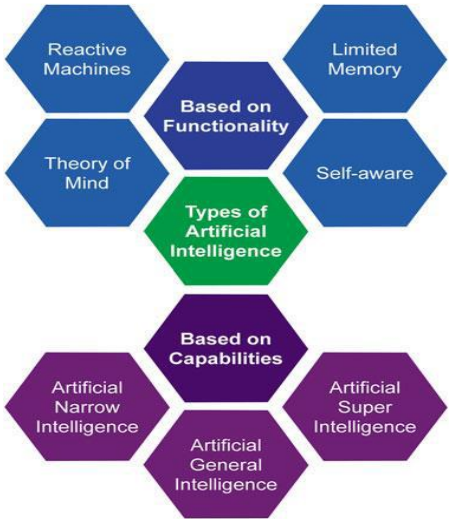

Figure 3.2: Different types of artificial intelligence

Artificial Intelligence vs Traditional Risk Calculators

Predicting the likelihood of a postoperative complication accurately is critical for optimizing patient selection before surgery, directing perioperative decision-making, gauging the threshold for concern in the postoperative period, and guiding early intervention. To stratify patients‟ postoperative morbidity and mortality risk, various risk stratification and predictive models have been developed, including the American College of Surgeons Surgical Risk Calculator (ACS-SRC), the American Society of Anesthesiologists (ASA) score, and the Physiologic and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM). AI techniques often have better predictive performance than do traditional statistical models limited logistic regression. ML algorithms allow for the evaluation of a higher number of clinical variables than do traditional modeling approaches and may help identify weak predictors or interactions between variables that may improve prediction accuracy. By developing nonlinear models that use multiple data sources, such as diagnoses, treatments, and laboratory values, ML has outperformed logistic regression for predicting postoperative outcomes. Recent work from our group, for example, showed that, compared with multivariable logistic regression, ML demonstrated higher predictive discriminatory performance and identified more predictors of complications in both abdominal wall reconstruction and reconstruction following mastectomy. And in contrast to conventional statistical approaches, incremental learning enables ML to improve continuously as new data are added.

4. PROPOSED MODEL

4.1 OVERVIEW

Machine learning is one of the most promising tools in classification. In essence; machine learning is a model that aims to discover the unknown function, dependence, or structure between input and output variables. Usually, these relations are difficult to be existed by explicit algorithms via automated learning process. Machine-learning methods are applied to predict possible confirmed cases and mortality numbers for the upcoming. Machine learning can be divided into two parts. The first part is to define the optimal weight of data fusion of multi-node perception outcomes and eliminate unusable nodes based on the genetic algorithm, while the second part is to find fault nodes through a fault recognition neural network. Machine learning is a subsection of Artificial Intelligence (AI), and it involves several learning paradigms, such as Supervised Learning (SL), Un-supervised Learning (UL), and Reinforcement Learning (RL). Typical ML models consist of classification, regression, clustering, anomaly detection, dimensionality reduction, and reward maximization. The ML algorithms are trained in the SL paradigm, on labeled data sets, meaning that they exist to a ground-truth output (continuous or discrete) for every input. Conversely, in UL there is no ground-truth output, and the algorithms normally attempt to discover patterns in the data. Reinforcement Learning aims to raise the cumulative reward so that it is more suitable for sequential decision-making tasks. Supervised learning has regression and classification; unsupervised learning includes cluster analysis and dimensionally reduction, also Reinforcement Learning (RL) includes classification and control, as illustrated in below figure.

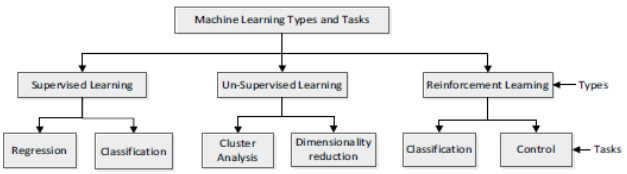

Figure 4.1: Overview of machine learning types and tasks.

4.2 WORK FLOW OF MACHINE LEARNING DATA COLLECTION

This part of the project was slightly challenging, as open source datasets could not be found for all the diseases „that we intended to predict. Moreover, the collection of observational data from patients wearing smart watches would not have been feasible, as the system involves the use of many sensor data for the prediction of several diseases that current smart watches do not provide. Because of this concern, records for approximately 303 patients with specific diagnoses were obtained from Hospital, with written consent from the patients. Each patient was provided with a copy of the ethical consent form. The reports were also checked and approved by the legal authority of the hospital. The health records and diagnoses from the report were taken as symptoms for specific diseases and used to train our model, and the test results obtained were further validated by expert doctors from the hospital. The final

validated dataset consists of around 260 data that have been used for training and testing machine learning models is discussed in the latter part of this section.

5. EXPERIMENTAL RESULT ANALYSIS

5.1 DATA SOURCE

Individuals with a history of cardiac problems or other medical conditions were selected from a structured dataset. Heart disease refers to a variety of conditions that affect the heart. Cardiovascular diseases are the leading cause of death among middle-aged people, according to the World Health Organization (WHO) [28]. The data set we use contains the medical histories of 304 people ranging in age from infants to elderly. This dataset contains crucial information about the patient, such as his or her age, resting blood pressure, fasting blood sugar level, and other medical features.

S. no.	Attributes	Description	Value
1	age	Age of a patient	Age of a patient start from 29 and maximum age is 77
2	sex	Sex of a patient	1 signifies male, and 0 signifies the female
3	cp	Type of chest pain	1 signifies typical angina; 2 signifies atypical angina; 3 signifies non-anginal pain; 4 signifies asymptomatic
4	trestbps	Blood pressure in resting phase	Blood pressure of a patient start from 94 and maximum blood pressure is 200
5	chol	Serum cholesterol	Cholesterol of a patient start from 126 and maximum cholesterol is 564
6	fbs	Blood pressure in fasting phase	1 signifies true, and 0 signifies false
7	restecg	Electrocardiographic during resting	0 signifies normal; 1 signifies patients having ST-T wave abnormality; 2 signifies left ventricular hyperthrophy
8	thalach	Maximum heart rate	Heart rate ranges from 71 to 202
9	exang	Angina due to exercise	1 means true, and 0 means false
10	oldpeak	ST depression due to exercise	displays the value as float or integer
11	slope	exercise slope relative to rest	1 signifies upsloping; 2 signifies flat; 3 signifies down sloping
12	Ca	Number of vessels (0–3) colored by fluoroscopy	displays the value as float or integer
13	thal	Thalassemia	3 signifies normal; 6 signifies fixed defect, and 7 signifies reversable defect
14	target	Whether a person is having CVD or not	0 signifies that patient is not having CVD and 1 signifies that patient is having CVD

Table 5.1: Above table shows dataset features value.

5.2 SIMULATION PARAMETERS Performance Analysis

This is accomplished by examining a range of performance metrics. A classifier‟s precision is defined by its accuracy across all classifications. Precision is a performance statistic that is used to evaluate the results of planned activities. This measure contrasts the overall number of TPs and TNs with the total number of correctly classified occurrences (True Negatives).

Table 5.2: The below table shows performance parameters value.

Metric	Definition	Formula
Accuracy	The overall truly predicted samples divided by overall samples	(TP + TN)/N
Recall	The percentage of actual positive samples that were predicted as positive	TP/(FN + TP)
Precision	How many of the positively classified samples were actually positive	TP/(TP + FP)
F1 Score	The harmonic means of both recall and precision	2(recall * precision)/(recall + precision)

Recall is the next performance statistic that is determined for performance analysis. Recall is a metric that indicates the completeness of all classifiers. The mathematical equation is used to determine the value of Recall. In general, the F-score is a performance metric that considers both precision and recall. It has a minimum value of 0 and a maximum value of 1. It is calculated as the harmonic mean of both the recall and precision numbers. Accuracy is a measure of how close the suggested model is to the target value. It is a measure that indicates the number of forecasts made in relation to the total number of predictions. The accuracy of the system is calculated using the following mathematical representation.

5.3 EXPERIMENTAL RESULT ANALYSIS

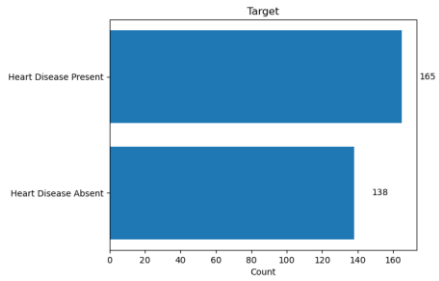

Figure 5.1: This picture shows dataset distribution

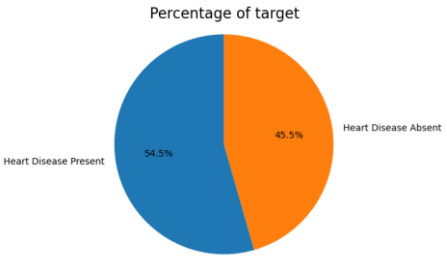

Figure 5.2: This picture shows dataset distribution in percentage value with diseases present or absent.

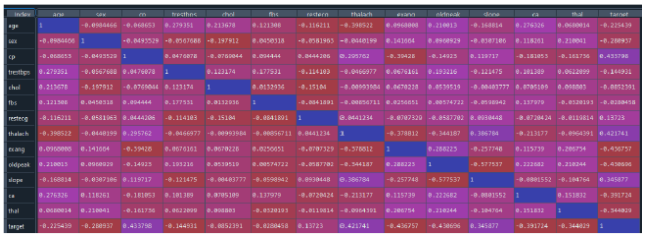

Figure 5.3: This picture shows dataset distribution in data frame.

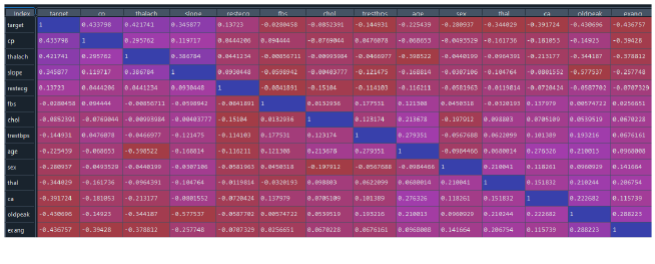

Figure 5.4: This picture shows dataset distribution in data frame where row indicate actual value and column indicate predicted value for features.

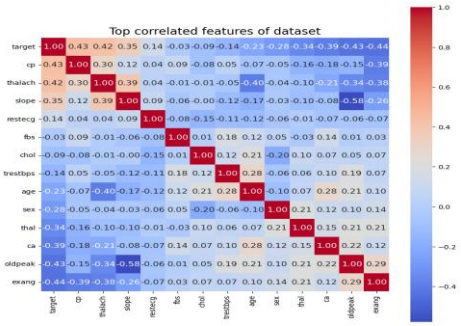

Figure 5.5: This picture shows dataset distribution correlated features.

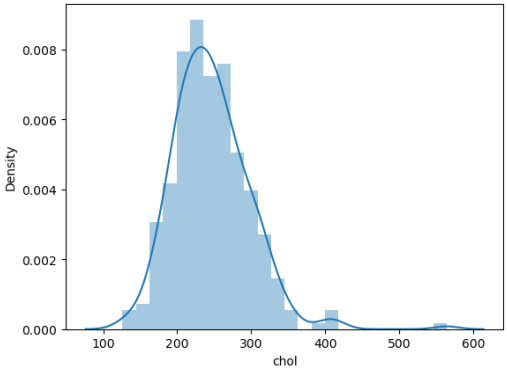

Figure 5.6: This picture shows dataset chol features

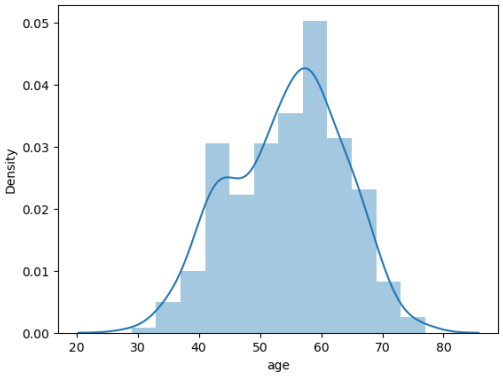

Figure 5.7: This picture shows dataset age features

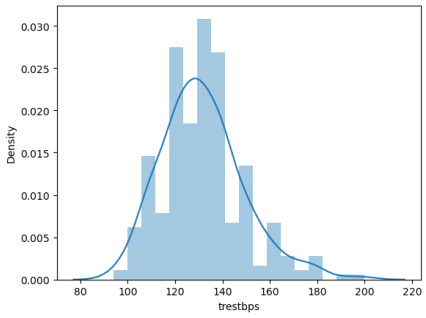

Figure 5.8: This picture shows dataset trestbps features.

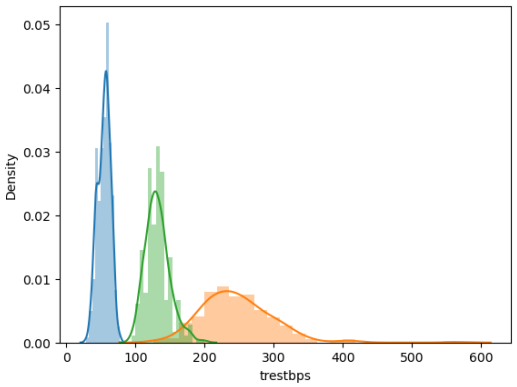

Figure 5.9: This picture shows dataset chol, age and, trestbps histogram features.

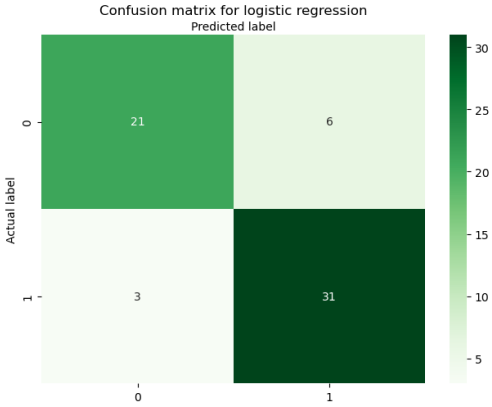

Figure 5.10: This picture shows confusion matrix for logistic regression model.

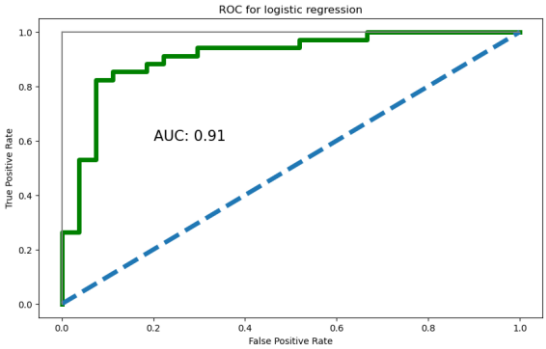

Figure 5.11: This picture shows roc for logistic regression model based on true positive rate and false positive rate.

```
    ...: from sklearn.metrics import confusion_matrix, classification_report
    ...: #from sklearn.metrics import precision_score, recall_score, f1_score
    ...: print(classification_report(Y_test, Y_pred_lr))
The accuracy score achieved using Logistic Regression is: 87.1 %
              precision    recall  f1-score   support

           0       0.88      0.88      0.88        17
           1       0.86      0.86      0.86        14

    accuracy                           0.87        31
   macro avg       0.87      0.87      0.87        31
weighted avg       0.87      0.87      0.87        31
```

Figure 5.12: This picture shows accuracy and other performance parameters for logistic regression model.

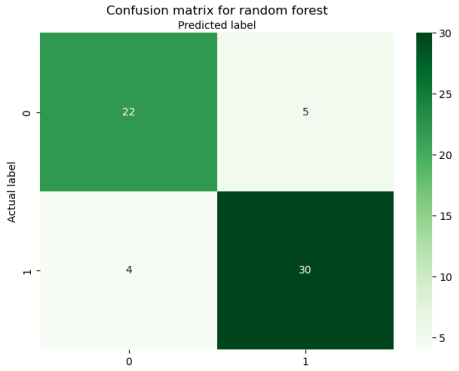

Figure 5.13: This picture shows confusion matrix for random forest model

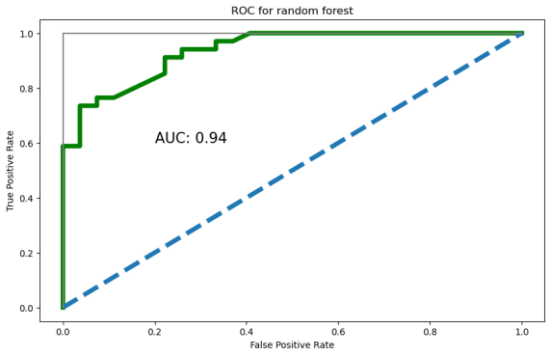

Figure 5.14: This picture shows roc for random forest model based on true positive rate and false positive rate.

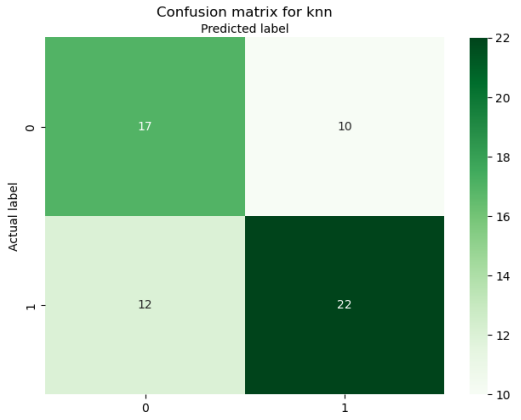

Figure 5.15: This picture shows confusion matrix for knn model.

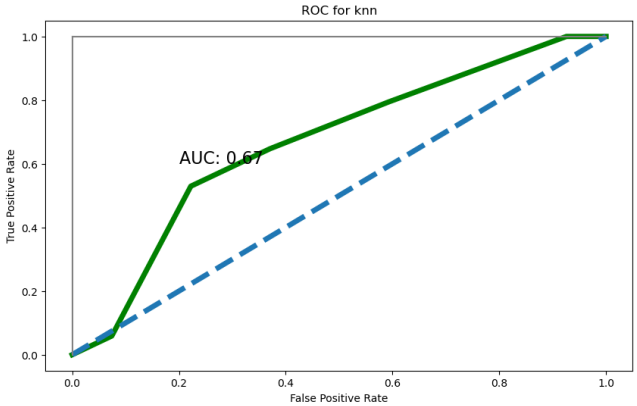

Figure 5.16: This picture shows roc for knn model based on true positive rate and false positive rate.

Table 5.3: The comparative experimental study between the different machine learning classifiers

Dataset name	Machine learning techniques	Accuracy in percentage (%)	Precision in percentage (%)	Recall in percentage (%)	F1-score in percentage (%)
Heart disease	LR	87.1	88	86	88
	NB	80.65	87	76	81
	SVM	83.87	88	86	85
	DT	70.97	84	86	83
	k-NN	74.19	88	82	85

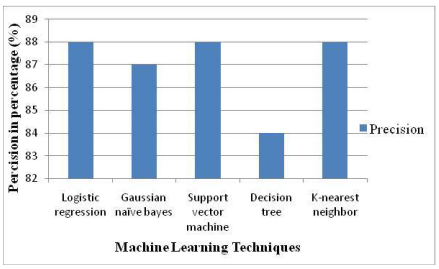

Figure 5.17: The above picture shows that comparative performance graph for precision.

The fig.5.17, presents that comparative results graph for precision performance parameter, and study shows that proposed model logistic regression, support vector machine, and K-nearest neighbor classifier gives the same precision value while naive bayes, and decision tree lower precision value.

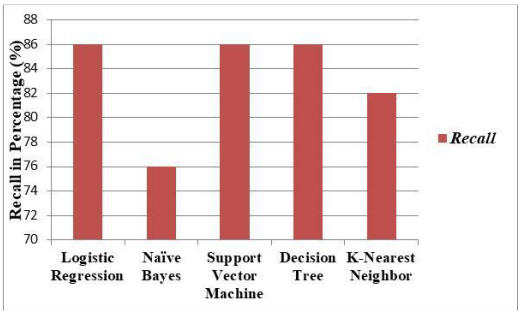

Figure 5.18: The above picture shows that comparative performance graph for recall.

The fig. 5.18, presents that comparative results graph for precision performance parameter, and study shows that proposed model logistic regression, support vector machine, and K-neares neighbor classifier gives the same precision value while naive bayes, and decision tree lower precision value.

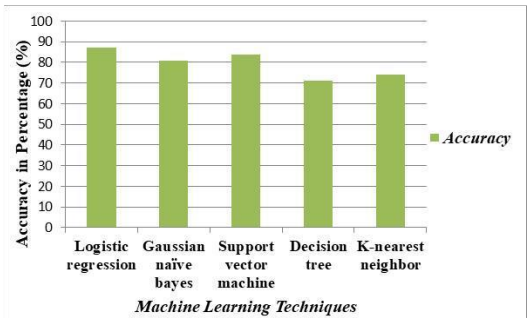

Figure 5.19: The above picture shows that comparative performance graph for accuracy. The fig. 5.19, presents that comparative results graph for accuracy performance parameter, and study shows that proposed model logistic regression gives highest accuracy than other models like, support vector machine, K-nearest neighbor, naive bayes, and decision tree models. In this research works, the comparative experimental study logistic regression model gives highest accuracy, i.e. (87.10) than other models.

6. CONCLUSIONS AND FUTURE SCOPE

6.1 CONCLUSION

Technology is an integral part of our life. Wherever we are, we are accompanied with technology as a companion. People have been replaced with technology to live with. As the expectations are increasing day by day, it had marked itself a standard of development from its early stage. From the invention of the wheel in the past to the autonomous machines and robots now, it had established a form of necessity and technology has earned a place in the world. Future technology which is in accordance with the term "emerging technology" changes with the timeline running in past, present, and future. With the evolving trend on technology, people no more update technology for their needs, it updates accordingly with the implementation in various fields. This research work aims to analyze the effects of ML techniques in detecting the heart diseases. The heart diseases dataset contain the a total of 303 person information, for the machine learning classifier, Here find the logistic regression model provides the best accuracy among all of them. In this research work proposed model which is based on ML techniques is an artificial intelligence technique and subfield of machine learning techniques, this model works with the concept of binary classification. In this study a present that model this is logistic regression give highest accuracy i.e. 87 % than other models. In future work, extend this research work using some optimization techniques, like honey bee classification, ant colony optimization, and particle swarm optimization to improve the enhancement of current work, and also improve the accuracy and other performance parameters value.

REFERENCES

[1] Karna Vishnu Vardhana Reddy, Irraivan Elamvazuthi, "Heart Disease Risk Prediction Using Machine Learning Classifiers with Attribute Evaluators", Appl. Sci. 2021, pp. 1-22.

[2] Ashir Javeed, Muhammad Asim Saleem, "Decision Support System for Predicting Mortality in Cardiac Patients Based on Machine Learning", Appl. Sci. 2023, pp. 1-14.

[3] Vigneshwaran Namasivayam, Nithyashree Senguttuvan, "Artificial Intelligence and Its Application in Cardiovascular Disease Management", 2022, pp. 189-207.

[4] Rustem Yilmaz, Fatma Hilal Yagın, "Early Detection of Coronary Heart Disease Based on Machine Learning Methods", MEDICAL RECORDS-International Medical Journal, 2022, pp. 1-6.

[5] Peng Wang, Zihuai Lin, Xucun Yan, Zijiao Chen, Ming Ding, Yang Song, Lu Meng, "A Wearable ECG Monitor for Deep Learning Based Real-Time Cardiovascular Disease Detection", IEEE 2022, pp. 1-9.

[6] Sandeep Chandra Bollepalli, Ashish Kumar Sahani, "An Optimized Machine Learning Model Accurately Predicts In-Hospital Outcomes at Admission to a Cardiac Unit", Diagnostics 2022, pp. 1-14.

[7] Jasjit S. Suri, Mrinalini Bhagawati, "Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review", Computers in Biology and Medicine, 2022, pp. 1-24.

[8] Sarra RR, Dinar AM, Mohammed MA, Ghani MKA, Albahar MA, "A Robust Framework for Data Generative and Heart Disease Prediction Based on Efficient Deep Learning Models", Diagnostics (Basel). 2022 Nov 22;12(12):2899. doi: 10.3390/diagnostics12122899. PMID: 36552906; PMCID: PMC9777498.

[9] Ahmed, Sumaira, " Prediction of Cardiovascular Disease on Self‐Augmented Datasets of Heart Patients Using Multiple Machine Learning Models", Journal of Sensors 2022.1 (2022): 3730303.

[10] Raniya R. Sarra 1 , Ahmed M. Dinar, "A Robust Framework for Data Generative and Heart Disease Prediction Based on Efficient Deep Learning Models", Diagnostics 2022, 12, 2899. https://doi.org/10.3390/diagnostics12122899, pp. 1-13.

[11] Li Li1, Xi Chen, and Sanjun Hu, "Application of an end-to-end model with self-attention mechanism in cardiac disease prediction", 2024, Front. Physiol. 14:1308774. doi: 10.3389/fphys.2023.1308774.

[12] Hengame Ahmadi Golilar, Alireza Azadbar, Roohallah Alizadehsani, "GAN‐MD: A myocarditis detection using multi‐channel convolutional neural networks and

generative adversarial network‐based data augmentation”, CAAI Trans. Intell. Technol. 2024;9:866‐878.

[13] Zhou, “Risk prediction of heart diseases in patients with breast cancer: A deep learning approach with longitudinal electronic health records data”, 2024, Published by Elsevier Inc. https://doi.org/10.1016/ j.isci.2024.110329.

[14] Anna Markella Antoniadi, Yuhan Du, “Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review”, Appl. Sci. 2021, pp. 1-23.

[15] Awais Mehmood, Munwar Iqbal, “Prediction of Heart Disease Using Deep Convolutional Neural Networks”, Arabian Journal for Science and Engineering, 2021, pp. 1-15.

[16] Baptiste Vasey, MMed; Stephan Ursprung, “Association of Clinician Diagnostic Performance With Machine Learning–Based Decision Support Systems A Systematic Review”, JAMA Network, 2021, pp. 1-15.

[17] JoungOuk Kim, Yong-Suk Jeong, “Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database”, Diagnostics 2021, pp. 1-12.

[18] K. Arumugam, MohdNaved, “Multiple disease prediction using Machine learning algorithms”, Elsevier, 2021, pp. 1-4.

[19] Krishan Kumar Goyal, Sandeep Kumar Jain, “Design and Analysis of Improved Machine Learning Model for Heart Disease Prediction”, Journal of Algebraic Statistics, 2022, pp. 461-472.

[20] Lamido Yahaya, Nathaniel David Oye, “A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques”, American Journal of Artificial Intelligence, 2020, pp. 20-29.