2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)
Manipal University, Jaipur. Jan 19-20, 2023

2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT) | 978-1-6654-5357-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCT56969.2023.10076174

A Comparison of Neural Networks and Machine Learning Methods for Prediction of Heart Disease

Omkar Subhash Ghongade
Department of CSE,
SRM University-AP, India
omkarsubhash_ghongade@srmap.edu.in

S Kiran Sai Reddy
Department of CSE,
SRM University-AP, India
kiransaireddy_suda@srmap.edu.in

Srilatha Tokala
Department of CSE,
SRM University-AP, India
srilatha_tokala@srmap.edu.in

Koduru Hajarathaiah
Department of CSE,
SRM University-AP, India
hazarathaiah_koduru@srmap.edu.in

Murali Krishna Enduri
Department of CSE,
SRM University-AP, India
muralikrishna.e@srmap.edu.in

Satish Anamalamudi
Department of CSE,
SRM University-AP, India
satish.a@srmap.edu.in

***Abstract*: Heart disease is a major cause of death and disability across the world. Heart disease mortality and morbidity rates can be greatly decreased with early detection and treatment. Hence, the development of efficient and accurate methods for early diagnosis of heart disease has become a priority in the medical field. In this study, we did a comparative study of exiting supervised machine learning approaches for predicting heart disease diagnosis and also improved the accuracy of KNN by changing K values. We used a dataset that consists of a variety of features such as age, gender and other important indicators for heart disease diagnosis. We then explored and evaluated traditional ML algorithms such as logistic regression, decision tree, random forest and SVM for the predictive analysis. A number of criteria, including accuracy, precision, recall, and F1 Score, were used to assess the models' performance. This study provides evidence that ML algorithms can be used to forecast the diagnosis of heart disease. Healthcare providers and medical practitioners can utilize the outcomes of this study for early detection and management of cardiac disease. Further research will aim to analyse and evaluate additional machine learning algorithms to enhance precision and performance.**

***Index Terms*—Machine leaning algorithms, ML performance models. Neural Networks.**

I. Introduction

In many nations, heart disease is the main cause of mortality and a serious public health concern. Genetic, societal, and environmental variables can all be implicated in its incidence. It can interfere with the circulation of blood to the heart, leading to serious health issues including stroke, heart attack, and heart failure [1], [2]. Machine learning (ML) techniques can be used to analyze huge amounts of data quickly and accurately, allowing for more accurate diagnoses and predictions. ML algorithms can be used to identify patterns in data that can help detect and diagnose heart diseases [3]. These algorithms can also be used to predict the future risk of developing heart diseases based on data such as past medical history, lifestyle, and genetics [4]. In addition to being more accurate, ML techniques are also more cost-effective and less time-consuming than traditional methods. For example, an electrocardiogram (ECG) can be used to measure the electrical activity in the heart, but it is costly and time-consuming. On the other hand, ML algorithms can be used to analyze ECG data more quickly and accurately. ML algorithms [5] can also be used to identify patterns in other types of data including patient medical history, lifestyle, and genetics. This data can be used to identify individuals who may be at risk of developing heart diseases, allowing doctors to take preventative measures. Machine learning techniques are a more accurate and cost-effective method of diagnosing and predicting heart diseases. This technology can provide doctors with more accurate diagnoses and predictions and can help identify individuals who may be at risk. ML algorithms can also be used to analyse large amounts of data quickly, saving both time and money. Using supervised machine learning, doctors and medical researchers can develop models that will accurately predict the risk of developing heart diseases [6]. This can help to identify those who are at a higher risk of developing heart diseases and help to create appropriate prevention and treatment plans. Supervised machine learning is a powerful tool that can be used to identify patterns in data and to develop models that can accurately predict the outcome of a given situation. It can be used to develop models that can accurately predict the risk of developing heart diseases, making it a valuable tool in the fight against heart disease.

The use of data and ML models to predict the risk of developing heart diseases has become increasingly prevalent in the healthcare industry [7]. These models use a variety of data such as age, lifestyle, risk factors, and medical history to develop a comprehensive risk profile for each patient. This data is then used to train the machine learning models to accurately predict the risk of developing heart diseases. This allows healthcare professionals to identify patients who are at a higher risk of developing heart diseases and to provide them with the necessary preventative measures. The accuracy of these models has been proven in numerous studies and is now being used in clinical settings to identify at-risk patients. By utilizing these models, healthcare professionals can accurately identify those who are at a higher risk of developing heart diseases and provide them with the necessary preventative care. This is important for early detection and treatment of heart diseases, which can significantly improve patient outcomes. In addition to their use in clinical settings, these models are also being used to develop personalized treatment plans for patients. By analysing a patient's risk factors and medical history, the models can determine the best course of treatment for each individual patient. This allows healthcare professionals to provide personalized and effective care to each patient. Ultimately, the use of machine learning models to predict the risk of developing heart diseases has revolutionized the healthcare industry [8]. By utilizing these models, healthcare professionals can identify at risk patients and provide them with the necessary preventative care. Additionally, these models can be used to develop personalized treatment plans for patients, allowing them to receive the most effective care possible.

Supervised machine learning models have been used to detect early signs of heart diseases [9], [10]. These models analyse data to recognize patterns that can be indicative of the disease, allowing doctors to take timely action and reduce the risk of further complications. The data used by supervised machine learning models can be collected from multiple sources, including patient records, medical scans, and vital signs. By analysing this data, the models are able to detect patterns that may indicate the early stages of heart disease. This includes changes in the heart rate, blood pressure, and other vital signs. The models can also use data from medical scans to detect signs of damage to the heart, such as blocked arteries or signs of inflammation. By combining this data with patient records, the models can accurately detect early signs of heart diseases. The models can also be used to detect conditions that may lead to heart disease [11], such as high cholesterol or elevated blood sugar levels. By detecting these conditions early, doctors can take the necessary steps to prevent the development of a more serious form of the disease. By combining data from multiple sources, the models can identify patterns that may indicate the presence of heart diseases and alert the doctor to take the necessary actions. This can help to reduce the risk of developing a more serious form of the disease.

For the diagnosis and prediction of heart problems, supervised ML approaches are being used more and more. [12], [13]. This is due to their ability to provide accurate and efficient predictions and diagnoses. These supervised machine learning techniques use data from past cases and patient records to learn patterns and trends related to illnesses. By precise predictions and diagnoses of heart diseases. Since doing this, they can detect subtle differences in the data that may indicate the presence of a disease. This allows for more the data used for predictions and diagnoses is from past cases, supervised machine learning techniques are also able to identify potential risks and warning signs that may be associated with the disease [9]. Furthermore, these techniques can also provide insights into the causes and possible treatments of a heart disease. This helps medical professionals to develop personalized treatments for their patients. The use of supervised ML techniques for diagnosing and predicting heart diseases is becoming increasingly popular and is the preferred method for many medical professionals. As a result, supervised ML techniques will continue to become increasingly important in helping to tackle the increasing prevalence of heart diseases. Various ml techniques will be utilised in this research to obtain accuracy, precision, recall, and the F1 score, on the basis of which we will conduct a comparison of the various supervised learning methods. This comparative study will evaluate the effectiveness of the ML algorithms.

A. Performance Measures:

Performance measures are essential metrics used to determine the effectiveness of a ML algorithm. Performance measures are used to analyze the results of different models on a given data set and to determine which ML model gives the good results. Performance measures can be used to identify areas in which the model can be improved and to determine the overall accuracy of the model. There are four ML classification model performance measures such as Accuracy, Precision, Recall, and F1-scores.

Accuracy Score: The most widely used metric to calculate the performance of a ML model is the Accuracy Score, which is determined by ratio of number of correct predictions made by the model to the total number of predictions. A higher Accuracy Score indicates that the model is more reliable in predicting the desired output.

$$AS = \frac{True\ Positives + True\ Negatives}{TP + FN + TN + FP}$$

Precision Score: Precision score is a performance measure used in ML to determine the accuracy of a model's predictions. The ratio of true positives (TP) to the total of false positives (FP) and true positives (TP) serves as a measure of precision. Precision score is typically expressed as a percentage, with a higher percentage indicating a better model performance.

$$PS = \frac{TP}{TP + FP}$$

Recall Score: Recall Score is a performance measure used in machine learning that determines a model's ability to correctly identify relevant instances from a data set. By dividing the total number of correctly identified relevant occurrences by the total number of relevant examples in the data set, recall is computed. The better the performance of the model, the higher the recall scores.

$$RS = \frac{TP}{TP + FN}$$

F1-Score: The F1-score is an important performance measure in machine learning, used to determine the performance of a classification model. F1-score is the harmonic mean of recall and precision scores. The F1-score examines both false positives and negatives and is especially useful when there is an uneven class distribution or a need to weigh precision and recall equally. It is used in combination with other performance measures such as accuracy, precision, and recall.

$$F1 = \frac{2 * PS * RS}{PS + RS}$$

II. METHODOLOGY

We used some ML models like Logistic Regression [14], KNN [15], SVM [16], Naive Bayes [17], Random Forest Classifier [18], XGBoost, Decision Tree [19] and Neural Networks [20].

A. Logistic Regression (LR):

LR is a supervised learning methodology that is used to predict the output of a categorical dependent variable [14]. It is used to solve classification problems and generates values that range from 0 to 1. Using the sklearn library, we can import LR. In real-life applications, it can be used in areas such as heart disease prediction. LR is a powerful tool for decision-making and can be used in a variety of applications.

$$P = \frac{e^{a+bx}}{1 + e^{a+bx}}$$

B. K-Nearest Neighbors (KNN):

KNN is a supervised learning algorithm for both classification and regression tasks. It is considered to be a non-parametric algorithm, which means it does not assume any underlying distributions of the data. KNN works by storing the training dataset and applying the classification or regression on it during the time of prediction. In other words, the algorithm is ”lazy”, meaning it does not learn from the given training set. KNN is most effective when dealing with data sets that contain a large number of reviews, as it is robust to noisy training data. KNN is also useful when there is a need to predict values in regions where no data is available, as it is based on the closest data points. KNN is one of the simplest yet most powerful machine learning algorithms and it can be used for classification tasks.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

C. Naive Bayes (NB):

NB is a supervised ML algorithm based on applying Bayes’ theorem [17] with strong (naive) independence assumptions between the features. It is a probabilistic classification algorithm used in various applications such as text classification, sentiment analysis and medical diagnosis. It is a simple yet powerful algorithm that can be used for various classification tasks.

$$P(c|x) = \frac{P(c|x) * P(c)}{P(x)}$$

D. Support Vector Machine (SVM):

SVM is used for classification and regression [16]. Decision planes, which specify decision boundaries, serve as the foundation for SVMs. These planes are used to categorize data by separating data points into distinct classes. SVMs are effective for classification and regression, as they can accurately identify complex non-linear boundaries and handle large datasets with many features. They are also versatile and can be applied in various contexts, such as text classification and image recognition.

$$f(x) = w^T + b$$

E. Decision Tree (DT):

DT is a type of supervised ML algorithm that is used to classify and categorize data [19]. They are based on a set of rules that are used to determine the outcome of a given input. A DT is a visual representation of the conditions and possible outcomes of a decision-making process. It consists of nodes and branches, where each node represents a certain decision or outcome, and each branch represents a certain condition that must be true in order for the decision to be taken. The process of building a decision tree involves setting up a set of rules that will guide the decision-making process. These rules are based on data that has been collected, such as historical data, or data from surveys or experiments. The rules are then used to predict the outcome of a given input.

$$Gini = 1 - \sum_{i=1}^{c} {p_i}^2$$

F. XGBoost:

Extreme gradient boosting (XGBoost) is a sophisticated use of the gradient boosting technique. It is a powerful ML algorithm used for predictive modeling and classification problems. XGBoost has become popular due to its superior performance in terms of accuracy and speed compared to other ML algorithms. XGBoost works by building an ensemble of weak learners, each of which is a decision tree. The decision trees are trained using gradient descent to minimize the error of the model. The model uses a regularization technique known as regularization, which helps to prevent overfitting [21].
The model also uses a tree-based learning algorithm called boosting, which helps to improve the accuracy of the model by combining the prediction of multiple weak learners.

$$f(x) = f(a) + f'(a)(x - a) + 0.5(f''(a)(x - a)^2)$$

G. Random Forest (RF):

Random Forest [18] is an ensemble machine learning algorithm that utilizes multiple decision trees to make predictions. It is an effective method for both classification and regression tasks. The algorithm works by randomly selecting a subset of features from the training dataset and then building individual decision trees using these features. The individual trees are then combined to create a "forest" of trees that are then combined to make a prediction. The predictions are made by averaging the predictions of each individual tree. The main benefit of RF is that it is not prone to the overfitting problem found in other algorithms. This is because of its ensemble nature, in which it combines the results of multiple models to make the final prediction. Additionally, it is able to handle large data sets with high dimensionality. RF is also a fast algorithm, requiring only a few lines of code to implement. This makes it highly suitable for large-scale tasks as it can be trained quickly and efficiently.

$$RFfi_i = \frac{\sum_j normf_{ij}}{\sum_{j \in all\ features,\ k \in all\ trees} normfi_{jk}}$$

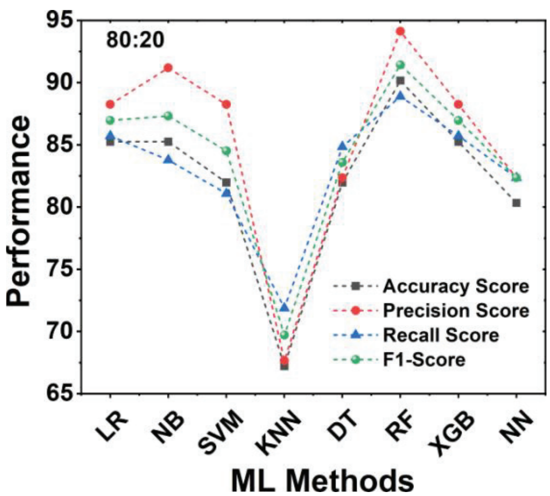
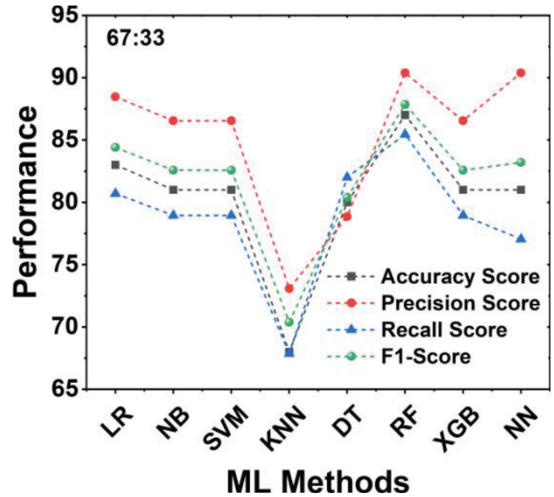


Fig. 1. Performance measures with training and testing ratio 80 : 20.

H. Neural Networks (NN):

Neural Networks are mainly of artificial intelligence that simulates the process of human brain. They are controlled of neurons, or nodes, that are associated together in layers and process information through a series of mathematical equations. Every neuron considers in input from other neurons in the network and processes it according to a set of weights and biases. The output of each neuron is then sent to the next layer of neurons. Based on the input, each neuron's output is updated, creating a chain of interconnected "neurons" that work together to process information. By "training" the network with large amounts of data, the neurons can learn to recognize patterns and make predictions. NN is used in a variety of applications such as image recognition, and medical diagnosis. They are useful for complex tasks that require the processing of huge amounts of data and are often more accurate than traditional algorithms.



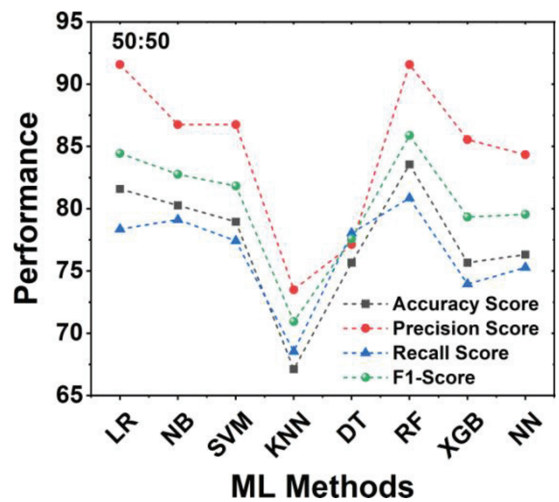


Fig. 3. Performance measures with training and testing ratio 50 : 50.

III. Data sets

The given data is stored in comma-separated value (CSV) files. Each line of the file contains an individual data record. For this Research, we trained on datasets containing 303 people's information about their heart disease issues. Each Data Set contain 14 attributes [22]. They are

1) Age: The duration of something from its beginning up to the present. This attribute contains age of each person which ranges from 0 to 100. It plays a major role in heart disease prediction because majority of those who succumb to heart disease are aged 60 and above.
2) Sex: Sex refers to the social and cultural distinctions between people who are male or female. This attribute contains sex of each person where 1 indicates male and 0 indicates female. It attributes plays a role in 2 different ways.
   - Diabetes: Heart problems are more common in women than in men.
   - No Diabetes: In contrast to women, men are more likely to get heart problems.
3) Chest Pain: Chest pain is discomfort or pain felt in the chest area. It can be a symptom of various medical conditions, such as heart attack, angina, or pneumonia. The pain may be sharp or dull, and may be felt in the center of the chest or in a specific area. This attribute contains the kind of chest pain experienced by every person. Four types of chest pains are there which are as follows:
   - Typical Angina - indicates 0
   - Atypical Angina - indicates 1
   - Non-Anginal pain - indicates 2
   - Asymptotic - indicates 3

   So, these attribute ranges from 0 to 3.
4) Resting Blood Pressure: Having high BP can be detrimental, as it can harm the arteries that supply blood to the heart. This attribute contains the resting blood pressure value of every person in mmHg(unit).
5) Serum Cholesterol: Having a high amount of LDL cholesterol can lead to clogged arteries, whereas having a high amount of triglycerides, increases the possibility of a heart attack. On the other hand, having a high amount of HDL cholesterol, which is beneficial, reduces the possibility of a heart attack. This attribute contains the serum Cholesterol of every person in mg/dl(unit).
6) Fasting Blood Sugar: Having insufficient insulin secretion or an inability to effectively utilize insulin from your pancreas can cause your blood sugar levels to rise, putting you at higher possibility of suffering with a heart attack. This attribute compares value of every person with 120mg/dl. This attribute contains $0's$ and $1's$, where 1 indicates fasting blood sugar greater than 120mg/dl and 0 indicates fasting blood sugar less than or equal to 120mg/dl.
7) Resting ECG: This attribute contains the results of Resting ECG of every person which range from 0 to 2 where (a) Normal - indicates 0, (b) having ST-T wave abnormality - indicates 1, (c) left ventricular hypertrophy - indicates 2. Max heart rate achieved: This attribute contains the maximum heart rate attain by every person.
8) Exercise induced angina: Angina is a kind of chest pain that is usually experienced as a gripping, tight, or squeezing sensation. This attribute says whether person has Exercise induced angina or not. It contains $0'$s and $1'$s were
   - Having Exercise induced angina - indicates 1
   - No Exercise induced angina - indicates 0
9) ST depression induced by exercise relative to rest: It refers to a decrease in the ST segment of the ECG (electrocardiogram) compared to the ST segment of the ECG at rest. This decrease is typically seen during exercise stress tests and is a sign of myocardial ischemia, which is a decrease in blood flow to the heart caused by an occlusion in the coronary arteries.
10) Peak exercise ST segment: It is a measurement of electrical activity in the heart during exercise. It is measured by an electrocardiogram and is used to detect signs of ischemia, or decreased blood flow to the heart muscle. During exercise, the ST segment may become elevated, flattened, or depressed, which can indicate a decrease in blood flow to the heart.
    - Upsloping – indicates 0
    - flat – indicates 1
    - downsloping – indicates 2
11) Number of major vessels $(0-3)$ colored by fluoros copy: It is used to diagnose and treat diseases and medical conditions. It is also used to guide and monitor the progress of medical procedures such as catheter placement and biopsies.
    - Normal – indicates 0
    - Mild Abnormality – indicates 1
    - Moderate Abnormality – indicates 2
    - Severe Abnormality – indicates 3
12) Thal: Hemoglobin, a protein found in red blood cells that carries oxygen to other regions of the body, cannot be produced by the body properly due to the hereditary disorder known as thalassemia. This can lead to a decrease in red blood cell count, also called anemia, and other serious health problems. Those affected by thalassemia either make none or too little hemoglobin, which can cause various issues including fatigue, bone deformity, shortness of breath, and even heart failure. It is important to be aware of the signs and symptoms of thalassemia, and to seek medical advice if any of them are present.
    - No Thalassemia – indicates 0
    - Mild Thalassemia – indicates 1
    - Moderate Thalassemia – indicates 2
    - Severe Thalassemia – indicates 3
13) Target: This attribute contains 0's and 1's where 1 indicates a person is suffering with heart disease and 0 indicates absence of heart disease.

As we are applying supervised Machine Learning models, here my Target attribute acts as a dependent parameter and the remaining of the attributes are independent parameters. Here, we are making a Comparative study on each performance measure i.e., Recall, Precision, Accuracy, and F1-Score by taking different Train-Test Split Ratios into Data Sets. This allows us to test the accuracy of each model by using the seven ML methods and neural networks that have been developed.

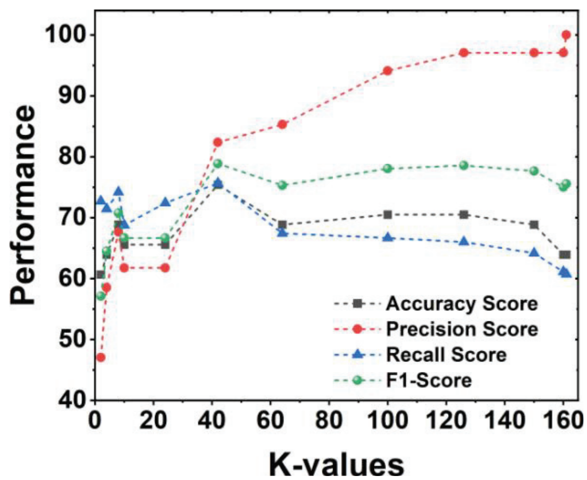


Fig. 4. Different K-values for a data set using seven ML and NN algorithms.

IV. Results

In this section, after training the seven machine learning algorithms and Neural Network, we observed that Random Forest Classifier is giving high values for all the performance measures, i.e., Recall, Precision, Accuracy, and F1-Score, shown in the Fig. 1, 2, and 3. We also observed that, when the training ratio decreases or the test ratio increases, then the scores of all performance measures decrease. In Random Forest, when constructing trees, a randomly chosen subset of features is used instead of searching for the most important parameter when slicing a node. This adds a level of randomness to the model, which reduces the overfitting problem of decision trees and decreases the variance, thereby boosting the accuracy. When the training data decreases, the model will have less information to learn from, resulting in a lower accuracy on unseen data (the test set). On the other hand, when the test set increases, the model will need to be more precise in its predictions, resulting in a lower accuracy score. This is because the model will now have more data to make predictions on, meaning that it needs to be more precise in its predictions in order to correctly classify the new data. After Random Forest, XGBoost and Na¨ıve Bayes are giving better accuracy compared to Neural Networks. For predicting heart disease, KNN is giving poor accuracy compared to all other Machine Learning Models and Neural Networks. Finally, we can observe that Random Forest Classifier is provides the best Scores (Accuracy, Precision, Recall, F1) for predicting heart disease when we compared with other ML Models and Neural Networks. For our data set, which contains 303 reviews on heart disease prediction, we trained model using KNN. We got different accuracies based on the K value. We observed that, as K increases then Precision score also increases, which is shown in the Fig. 4. If K is more than 160 then our Precision Score is 100%. As shown in the Fig. 4, we can conclude that KNN is giving good Accuracy Scores and other Performance measures at $K = 42$ for heart disease prediction.

V. Conclusion and Future Work

As the next step in this field of research, some promising directions could be explored. Investigating the impact of incorporating different types of data sources on the accuracy of heart disease prediction. For instance, incorporating environmental and lifestyle factors, such as air pollution, physical activity, and diet, into the predictive models could lead to more accurate predictions. Developing more sophisticated machine learning techniques, such as ensemble methods, to improve the accuracy of predictions of heart disease. Developing deep learning architectures for predictions of heart disease. Enhancing the interpretability of predictive models by using techniques, such as feature importance and sensitivity analysis. Investigating the potential of using natural language processing techniques to extract and analyse medical records for predictions of heart disease. Exploring the use of reinforcement learning techniques for predictions of heart disease. Investigating the impact of utilizing large-scale medical records and electronic health records on the accuracy of heart disease predictions. It would also be beneficial to consider the use of unsupervised learning methods to find hidden patterns in the data, such as clustering or anomaly detection. Another, it would be interesting to explore the use of transfer learning, and convolutional neural networks, which could enable the model to leverage knowledge from existing models in the domain to improve its prediction accuracy. Finally, it would be beneficial to investigate the impact of incorporating additional sources of data, such as medical images and patient records, into the model to further improve its accuracy.

References

[1] A. K. Pagare and V. K. Verma, "Heart attack prediction using data mining classification techniques: A study," *International Journal of Technology Research and Management ISSN (Online)*, pp. 2348–9006.

[2] S. U. Amin, K. Agarwal, and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," in *2013 IEEE Conference on Information & Communication Technologies*. IEEE, 2013, pp. 1227–1231.

[3] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in *2018 second international conference on electronics, communication and aerospace technology (ICECA)*. IEEE, 2018, pp. 1275–1278.

[4] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine learning: algorithms and applications*. Crc Press, 2016.

[5] A. Singh, N. Thakur, and A. Sharma, “A review of supervised machine learning algorithms,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. Ieee, 2016, pp. 1310–1315.

[6] V. Krishnaiah, G. Narsimha, and N. S. Chandra, “Heart disease prediction system using data mining technique by fuzzy k-nn approach,” in *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1*. Springer, 2015, pp. 371–384.

[7] S. Krishnan and S. Geetha, “Prediction of heart disease using machine learning algorithms.” in *2019 1st international conference on innovations in information and communication technology (ICIICT)*. IEEE, 2019, pp. 1–5.

[8] A. Rajdhan, A. Agarwal, M. Sai, D. Ravi, and P. Ghuli, “Heart disease prediction using machine learning,” *International Journal of Research and Technology*, vol. 9, no. 04, pp. 659–662, 2020.

[9] S. Goel, A. Deep, S. Srivastava, and A. Tripathi, “Comparative analysis of various techniques for heart disease prediction,” in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*. IEEE, 2019, pp. 88–94.

[10] S. K. Jonnavithula, A. K. Jha, M. Kavitha, and S. Srinivasulu, “Role of machine learning algorithms over heart diseases prediction,” in *AIP Conference Proceedings*, vol. 2292, no. 1. AIP Publishing LLC, 2020, p. 040013.

[11] C. Kalaiselvi, “Diagnosing of heart diseases using average k-nearest neighbor algorithm of data mining,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2016, pp. 3099–3103.

[12] F. Fatima, A. Jaiswal, and N. Sachdeva, “Heart disease prediction using supervised classifiers,” *Available at SSRN 4121817*, 2022.

[13] M. Fatima, M. Pasha *et al.*, “Survey of machine learning algorithms for disease diagnostic,” *Journal of Intelligent Learning Systems and Applications*, vol. 9, no. 01, p. 1, 2017.

[14] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, “Credit card churn forecasting by logistic regression and decision tree,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 15 273–15 285, 2011.

[15] L. Jiang, Z. Cai, D. Wang, and S. Jiang, “Survey of improving k-nearest-neighbor for classification,” in *Fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007)*, vol. 1. IEEE, 2007, pp. 679–683.

[16] B. Mahesh, “Machine learning algorithms-a review,” *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, pp. 381–386, 2020.

[17] K. Vembandasamy, R. Sasipriya, and E. Deepa, “Heart diseases detection using naive bayes algorithm,” *International Journal of Innovative Science, Engineering & Technology*, vol. 2, no. 9, pp. 441–444, 2015.

[18] G. Biau, “Analysis of a random forests model,” *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 1063–1095, 2012.

[19] C. Liu, K. Sun, Z. H. Rather, Z. Chen, C. L. Bak, P. Thøgersen, and P. Lund, “A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees,” *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 717–730, 2013.

[20] W. G. Baxt, “Use of an artificial neural network for data analysis in clinical decision-making: the diagnosis of acute coronary occlusion,” *Neural computation*, vol. 2, no. 4, pp. 480–489, 1990.

[21] S. Doki, S. Devella, S. Tallam, S. S. R. Gangannagari, P. S. Reddy, and G. P. Reddy, “Heart disease prediction using xgboost,” in *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*. IEEE, 2022, pp. 1317–1320.

[22] “Kaggle,” https://www.kaggle.com/ronitf/heart-disease-uci.