Chapter 10

Book Chapter

Heart Disease Prediction using Machine Learning Algorithms

Gomathy Prathima E *[1], Hanamant R Jakaraddi †[2], and Pooja N M ‡[3]

[1]Associate Professor, Dept. of MCA , Atria Institute of Technology, Bengaluru
[2]Assistant Professor, Dept. of MCA , Acharya Institute of Technology, Bengaluru
[3]Dept. of MCA , Atria Institute of Technology, Bengaluru

Abstract

Cardiovascular diseases rank among the primary factors leading to mortality, early detection and excellent prediction are essential. The rapid level through increased diagnostic accuracy in forecasting the occurrence of cardiac disease. Using patient information such as gender, age, and hypertension to foretell the implement of ventricular ailments, cholesterol levels, and other clinical markers, this chapter examines Robotic intelligence methods that have undergone in-depth research developed such as neural networks as a field, logic- regret, Decision-trees, and sup-vet- mac trees, have been developed. The models are trained and validated using conventional performance metrics, such as Formula One, memory, quality, and sharpness score, by utilizing a dataset from a reputable medical repository. The outputs indicate that statistical learning models, especially outfit draws near and brain organizations zeal have an eagerness to reach high prediction performance in clinical settings.

Keywords: Machine Learning. Predicting Cardiovascular Attack. Feature Selection. Model Optimization.

*Email: prathima.g@atria.edu Corresponding Author
†Email: hanamant2504@acharya.ac.in
‡Email: pnm995930@gmail.com

1 Introduction

A cohort of algorithms is suggested for computers which have been declared to be "machine learning" capable of self-improvement without the need for explicit programming by a programmer. "Machine learning," a branch of artificial intelligence, has the ability to foresee an outcome by combining statistical analysis techniques with data, which can provide extremely insightful information. The idea that a computer could memorize information from data is the foundation of the innovation. For example, Reinforcement learning is closely related to Bayesian predictive analytics and data drilling modelling, and can generate accurate results on its own. After receiving input, the computer applies an algorithm to generate output. Making suggestions is a mere machine learning challenge. Using the input and output from the pre-trained data the computer will generate the rules. Fig 1 shows the machine learning working model.

Contrary to conventional programming, machine learning operates in a fundamentally distinct way. In traditional programming, every rule is explicitly written by a single programmer, often in collaboration with domain experts to develop enterprise software. Machine learning, however, focuses on algorithms that learn patterns from data rather than relying on predefined rules. Instead of manually coding every instruction, the system evolves through training and adapts autonomously based on experience. This shift enables machine learning models to handle complex, dynamic scenarios that traditional programming struggles to address efficiently. All the regulations must be followed are rational, and the machine will follow it and produce an output. As the system gets harder to maintain, more the rules shall be followed, needed, and it may soon become unachievable and unworkable. This problem is meant to be solved via computer learning; the machine creates a rule after inferring the correlation. The device will carry out the logic and generate an output as the system grows more complex. Each time a specific entity appears in new data, the algorithm must adjust based on fresh insights and experiences to enhance its effectiveness. This adaptive learning allows the system to evolve autonomously, minimizing the need for manual intervention from software engineers. As the algorithm continuously refines itself, it ensures higher accuracy and efficiency over time. This process promotes scalability and seamless integration with ever-changing data patterns, maintaining optimal performance.

Editors: Shilpa Sivashankar, S.Pandikumar, Pallavi M O
DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

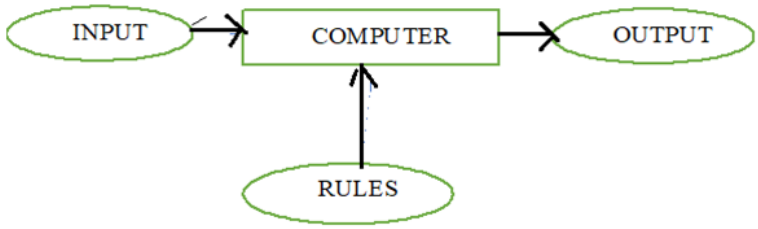

Figure 1. Machine Learning Model

2 Objectives

- Build a Predictive Model: Using input data, create a system of artificial intelligence that is capable of accurately estimating Potential of maladies for Using methods like decision tree classification and neural networks, this model should be trained on patient data and historical medical records.
- Boost Early Detection: By creating a Forecasting system that can spot subtle patterns and signs of cardiac problems before symptoms appear. By allowing for prompt intervention and treatment, early detection can potentially lower the pressure of complications and mortality.
- Improve Accuracy and Reliability: Make use of cutting-edge Learning through machines technique and algorithms to increase the precision and dependability of heart condition prognosis. This involves making the prediction model better.
- Simplify Diagnosis Process: Create an intuitive system that makes it easier for medical professionals to diagnose patients by automatically evaluating patient data and offering risk assessments and actionable insights about the possibility of cardiac. By prioritizing resources for patients who pose a greater risk Physicians can make more knowledgeable decisions.
- Validate and Evaluate Performance: To ascertain the effectiveness of the predictive model's performance and accuracy. Actual clinical settings, carry out extensive validation and evaluation studies.

3 Literature Review

Rani et al.'s (2022) entered a clinical deed-based ET-SVMRBF (vector device framework radial work with extra tree support) method for diagnosing CAD. The suggested hybrid algorithms reduce feature dimensionality while increasing classification accuracy. The effectiveness of GARFE was evaluated utilizing the SVM classifier. Four classifiers—DT,

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

KNN, XGBoost, and AdaBoost—are employed to assess HPCBE. High-dimensional data presents challenges for machine learning in real-world applications like e-healthcare. There could be additional and redundant components within all the details. The potency of prediction-based categorization systems is impacted by these superfluous features. Uçar et al.'s (2020) have described that algorithm accuracy is among the metrics used to evaluate algorithm performance. The training and testing datasets of deep learning algorithms impact their precision. Evaluation of the algorithms with the help of dataset whose properties are shown and the confusion sequence indicates that the KNN algorithm is the best one. The Anaconda (Jupiter) notebook is the ideal tool used to implementing Python programming; it contains an assortment of libraries and header files that improve accuracy and precision of the task the best, she discovered, is KNN of them all with 87% accuracy while table review the comparison. Nagavelli, Samanta, and Chakraborty's (2022) attempted to estimate the probability of developing cardiovascular disease by retrieving the patient's medical history from a dataset containing the following a fatal heart illness the patient's medical history, including conditions like blood pressure, sugar levels, and chest pain. If a patient has received a diagnosis already with heart disease, this cardiac condition. The detection system offers assistance by utilizing the clinical data of the person. In contrast to the Prior classifiers with good accuracy included logistic regression and KNN, plus to naive Bayes and others. Making the diagnosis is a challenging process that must be completed accurately and promptly. Pin-pointing those people who have a higher risk of developing sick disease given a range of medical characteristics.

Researchers had taken into account 14 essential qualities, naïve bayes, K-nearest neighbour and random forest are the best-performing algorithms. he found that K-nearest neighbours (k = 7) had the highest degree of accuracy after utilizing four different approaches. The performance be enhanced by incorporating additional data mining methods, such as Artificial intelligence, or machines for support vectors, clustering and association rules, time series, and genetic algorithms. Owing to the restrictions of the research, more advanced models mixing many models are required to enhance the early detection of heart disease. Each author reviewed and approved the finished text in sum to making an equal contribution to the research accuracy score is achieved with K-nearest neighbour. Stonier, Gorantla, and Manoj's (2024) have shown in their comparative analysis, that extreme gradient boosting classifier has the highest accuracy (81%) among those seven. To help with the system's evaluation, 14 pertinent raits that have been regarded as gathered from the dataset's 76 variables. The recently referenced unite those that will presumably end up fortifying the heart illness in people incorporating every feature result in a system that is less efficient for the inventor. The focus of selecting attributes is to increase output. Selecting n characteristics determining which prototype holds the top spot exact is essential. Many of the collection's features are eliminate that their correlations are nearly

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

identical equal. The effectiveness drops off quickly when every attribute in the collection is considered. The ensemble classifier they created can conduct hybrid classification by combining the best features of both strong and weak classifiers. It can do this by utilizing many training and validation examples. Scholars studied that the biggest difficulty Heart attack and stroke is linked to with identifying it. While there exist gadgets that can forecast the chance of acquiring heart issues in humans, they are either extremely expensive or ineffective in doing so. In medical data, the hidden patterns liable to diagnose illnesses.

By resolving the feature selection, or backward exclusion and RFECV, behind the models, this effort successfully predicted heart disease with 85% accuracy. The model used was logistic regression learn Instructions on utilizing the logistic regression model using the information acquired. To safeguard that whether patients will experience cardiac illness in ten years, we have also developed a model in this project that classifies patients based on different features (i.e., possible terrible for heart disease) by using regression using logistics. The creation of appropriate computer-based systems and decision assistance that can helps at the outset. Identification of cardiac problems is the driving force behind the field as a whole. Crtics have suggested that techniques in tandem with automated learning possess greater accuracy rates of over 95%, which establish them as important models for disease prediction and detection in the biological sciences to determine which model performs optimally for the databases under investigation, further analysis rules are employed. Compared to Random Forest and Simple Logistic models, SVM offers superior F-score, clarity, particularity, empathy, and exceptional consistency. SVM furthermore has the lowest miss rate. The number of features also affects the way the framework is sorted and make predictions. Unpublished the findings of the present study suggest that. SVM worked best when feature quantities were fewer in addition to the model performance metrics. During the same analysis, the Python platform was utilized, and SVM also identified the templates that show ideal execution while utilizing the Kernel Function of Radial Basis. Given SVM, It is intended to serve as the most opposing different algorithms for data mining, it was found that when the same types and quantities of features were used, none of the algorithms could predict heart disease with an accuracy level of more than 90%. In addition, this study combined two datasets that had the same quantity and kind of attributes. The chosen model will therefore be more trustworthy than those discovered in other studies.

Researchers have described their potential for greatness is immense. AI to enhance the avoiding and overseeing of cardiovascular disease. Large datasets can be analysed to find risk factors, predict outcomes, and develop tailored interventions. These algorithms can also seek in real-time decision support and remote monitoring to ensure patients receive timely and personalized care. But however, evidently operates a quantity of issues that must be resolved, like the availability and calibre of data, the interpretability of

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

models, and the ethical consequences of applying machine learning to the medical field, Furthermore, machine learning might be utilized as it relates to the new treatments and therapies for CVD. Significant improvements in cardiovascular disease steering clear and leading are possible with machine learning. ML technique can find novel biomarkers and possible therapeutic targets by examining enormous patient data sets. This information can then be utilized to create therapies that are more successful. The prevention and oversight of CVD could be revolutionized by machine learning, but to guarantee sufficient fully reap the rewards of this technology, a huge obstacle must be overcome. Areas that demand focus later on include raising the standardization and calibre of data, making models easier to understand, addressing ethical issues, and creating more individualized and adaptable interventions. We can use machine learning and lessen of CVD on society by striving to overcome these obstacles.

Kee et al.'s (2023) demonstrated that on the diagnosis and prognosis of numerous diseases have zeroed in on prediction models since the turn of the century. Machine learning (ML) has developed into a popular tool for creating prediction models due to advances in the computational technology. They had reviewed the current state of machine learning-based prediction models for cardiovascular illness (CVD) in Folks who have obesity type 2. (T2DM) is examined. To locate relevant articles, obtain on the research question, a thorough search of Scopus and Web of Science (WoS) was undertaken. Based The chance of bias is stated inside the projected price model's Risk of prejudice Inspection Software (PROBAST) statement for each article was evaluated. Neural network with 88.06% sensitivity, 76.6% precision, and a region under the area next to the curve (AUC) within the region most significant. Reliable algorithm to generate a model to conjecture the gamble cardiovascular disease poses a significant at of Diabetes type 2 have a coefficient of 0.91. Adhering to the PROBAST and TRIPOD assessment is strongly advised for future model advancement to reduce bias and ensure that its practicality in clinical settings.

Researchers had developed the model has made use of (MLA) like Random Forest, (SVM), Naive Bayes, and Decision Tree. They looked for correlations between the various features included in the dataset using conventional Strategies for AI, which they have effectively applied to the prediction of the odds of heart disease. The outcome demonstrates that Random Forest produces predictions with higher accuracy in less time than other machine learning techniques. The medical professionals at their clinic may find this model useful as a decision support system. They have tried to dig deeper the various machine learning approaches and predict whether a specific person, given various personal characteristics and indicators, will acquire coronary artery illness or not. They had examined the accuracy and the elements that play a part to the variations among various algorithms. They have divided the 1025-item Cleveland dataset for heart diseases into collections for experimentation and instruction using a percent split method. To drag the accuracy, he

Editors: Shilpa Sivashankar, S.Pandikumar, Pallavi M O
DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

made use of four different learnings and taken into account 14 attributes. They have focused the Random Forest is providing highest possible fidelity level—99 percent, while Decision Tree is performing the lowest—85 percent. Scholars demonstrated that an inaccurate prediction of cardiovascular disease can be fatal, an accurate prediction can also avert life-threatening situations. Heart disease, sometimes called cardiovascular disease, is among the complicated illnesses that people worldwide. They demonstrated a method for estimating coronary artery using an interface based on electrocardiogram Evaluation and testing of several methods for learning from machines. An enormous Many Folks have suffered from this condition. To create the interface, Django and Bootstrap are used. You can use it to find out if you have a heart condition. Upon uploading the ECG image to the website, strategies for neural networks like Naïve Bayes, Random Forest, and Decision Tree are utilized to estimate an opportunity of heart disease. Considering In laboratories, heart diseases are identified through the laborious process of continuously monitoring electrocardiogram (ECG) signals. They introduce an automated technique to identify cardiac conditions.

Srivenkatesh's (2020) explained that ML is modernized learning that requires virtually no human involvement. It includes programming personal computers to benefit from public data sources. Researching and producing estimates that have lessons to learn from the past data and make predictions based on new data is the guiding principle behind artificial intelligence. Preparing material and speaking to understanding are contributions to learning calculations, and any proficiency resulting from these inputs is the yield, which usually manifests as another calculation that can complete an assignment. Numerical, literary, auditory, visual, or sight and sound data can all be input into a machine learning framework. The framework's corresponding yield information can be symbolized by a gliding point number. A region's dataset is accustomed to contrast the precision of applying rules to the person's results of RF, SVM, along with naive Bayesian reasoning classifier, and logistic regression to accomplish the task, present an accurate model of heart problem prediction. Individuals with heart conditions may could be predicted the computations for predictive modelling with an accuracy ranging from 58.71% to 77.06% under investigation. It was demonstrated that comparing logistic regression to other machine learning models, the accuracy is higher (77.06%). In summary, information-digging systems are a great tool for prospective analysis in the domain of wellbeing because they allow us to anticipate illnesses and, as consequently, by holding out hope for a cure, save lives. There being ingested a persistent cardiovascular disappointment infection in patients and those without it was predicted research using learning computations, including RF, K- The closest neighbour, support vector machine, and Logistic Regression. The re-enactment showed in the Logistic Regression classifier was the most accurate and fastest to execute when it came to making predictions.

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

Ogunpola et al.'s (2024) had conducted a study for the sake of argument finished in contrast evaluation for heart disease prediction, showing promising outcomes. This investigation show that ML approaches perform better. XGBoost performed better in the ML technique for the 13 features in the dataset when data pre-processing was applied. With scores of 91% and 89% through the training and test, respectively, the XGBoost achieved the highest results. XGBoost produced comparable results, with 92% accuracy and an AUC score of 0.94. Shah and Patel's (2022) described that Kaggle is the origin of the database. Random Forest, XG-Boost, K-Nearest Neighbours (KNN), Logistic Regression, and Support Vector Machines (SVM) among the aforementioned are formulas for machine learning are the ones being developed employed in this instance in this particular situation to foresee the detection of cardiac problems. Python programming and Google Collaboration suggested to implement every one of them algorithms. The parameters performance evaluation is Fi-score, accuracy, precision, and recall. Testing and training programs are applied for various ratios resembles the XG-Boost algorithm yields the best prediction for heart disease. Doctors they possess the capability to use this type of heart disease prediction as a quick and efficient secondary diagnostic tool. This can enhance the early identification of cardiac disease, improving the patient's chances of survival is also important to observe that additional examination of it along with a greater comprehension of the underlying patterns and relationships can be achieved through the unsupervised learning algorithms. Foremost accurate and quick heart disease prediction, Physicians possess a knack for to this type of prediction as a supplementary diagnostic tool. As a result, it may raise saving patient's life.

Devi et al.'s (2023) expressed that putting the Warning signs of cardiac events application into practice is the project's goal. The information supplied by the user's device or cautious such as Android. The application determines the designation of the viral infection as an output. The intelligent system of the proposed system uses KNN, a machine learning technology. The user's data is compared with a few existing standard datasets to determine likelihood. Using KNN, the probability was discovered. The system will be evaluated, and errors will be identified and fixed exactly. Heart disease has emerged as many leading causes of death worldwide, thus early detection is essential. The project's objective is to develop a smartphone program to forecast heart disease using the approach known as KNN. The concept that ailment arises during the use relating to specific files and information entered by users. The patient receives content via a messaging app, which also provides specifics regarding.

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

4 Methodology

A Data-set description and pre-processing:

Heart disease continues to rank among the world's leading causes of death, highlighting the significance of swift detection and prevention. Utilizing machine learning to estimate how likely a cardiac disease can significantly enhance results and clinical decision-making. This project uses the Kaggle Heart Disease dataset, it encompasses an assortment of medical attributes, such as age, sex, kind of chest discomfort, blood pressure at rest, cholesterol, blood sugar levels, peak beat reached, results of resting electrocardiography, exercise-induced angina, ST depression caused by exercise in comparison to rest (old peak), the count of major boats and the climb of the peak exercise ST section coloured by fluoroscopy, and thalassemia. The target variable shows if cardiac disease is present or not.

Three Models of neural networks were applied for this objective of predicting cardiac disease: K-Nearest Neighbours (KNN), Decision Tree Algorithm (DTA), and Convolutional Neural Network (CNN). Convolutional layers of CNNs, commonly utilized for picture data, can also be modified to capture complex patterns in tabular data. The convolution and pooling layers of the CNN model are followed by dense layers that generate the final prediction. By splitting the information into branches and making predictions says the best important attributes, the DT Algorithm provides a straightforward but efficient technique. Last but not least, the K-Nearest Neighbours principles is a simple and natural model for classification problems because it classifies data points according to how close they are to other points.

- Data Cleaning: Managing absent elements and detecting and handling outliers.
- Data Transformation: Encoding categorical variables and normalizing or standardizing numerical features.
- Data Splitting: Train-test split

The testing set functioned as evaluate each model after trained on the training set. To contrast the models, performance metrics including F1-score, recall, accuracy, and precision were computed. This thorough process guarantees the robustness and dependability of the selected Simulation for forecasting stroke. The project's findings can help medical personnel detect high-risk individuals early, it will facilitate timely interventions and eventually improve patient outcomes and care. Using the 80% to train the data and 20% to test the data.

- Demographic: Age, sex
- Medical History: History of hypertension, diabetes, smoking status
- Clinical Measurements: BP Level, fat levels, blood sugar, electrocardiographic results

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

- Symptoms: Chest pain type, resting ECG results, exercise-induced angina
- Target Variable: on or off of cardiac problem (binary classification)

B Feature Selection

A crucial phase in the device's training procedure is feature selection. It involves choosing the most relevant attributes via the collection. Greatly enhance the model's ability to forecast the future. Selecting features wisely can decrease over-fitting, increase model performance, and expedite training. This procedure aids in identifying the critical health markers that are most indicative of cardiovascular disorders within the framework of projecting the core cancer.

- Improved Model Performance: By eliminating irrelevant or redundant features, models can perform better with increased accuracy and precision.
- Reduced Over-fitting: Models are less likely to learn noise from the practice data, resulting in better generalization to unseen data.
- Simplified Models: With fewer features, models become simpler, easier to interpret, and faster to train.
- Insight into Data: Identifying the most significant features can provide valuable insights into the factors contributing to heart disease.

C Machine Learning Algorithm and tools used

Algorithms and statistical models work in artificial intelligence (ML), a branch of artificial intelligence (AI), to help computers learn from and make predictions or choices based on data. There are several types, each suited for different types of tasks and data structures. Here's an overview of bunch of most common categories and algorithms:

(a) Supervised Learning: Supervised learning involves training a pattern based on a labelled dataset, meaning that each training example is paired with an output label. The goal is for studying of a mapping from inputs to outputs that can anticipate labels for new, unseen data.
   - Decision Trees: Models that use a tree-like structure to make decisions based on input features.
   - K-Nearest Neighbours (KNN): Classifies new samples depends on the majority label of their k-nearest neighbours during the lecture set.
   - Neural Networks: Composed of layers of interconnected nodes, capable of learning complex patterns in data.

(b) Unsupervised Learning: The mission of being devoid of supervision, which works with unlabelled data, is to deduce the inherent structure that exists inside a collection of data points.

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

(c) Partially Supervised Education: In semi-supervised learning, a small quantity of tagged data is utilized for training with large volumes of unlabelled data. This method can greatly increase learning accuracy in situations where getting tagged data is costly or time-consuming. To retrain itself, a model is first trained on labelled data and then iteratively applies labels to unlabelled data. Co-training is an itinerary of training two classifiers on two distinct feature sets and then utilizing both to label new data for the supplemental detector.

(d) Learning via Reinforcement: In reinforcement learning, an agent is trained by paying it to describe an uproar of decisions, rewarding it for wise choices and penalizing it for poor ones. When an agent interacts with an environment, whether in gaming or robotic control, this kind of learning is frequently employed.

5 Experimental Setup

K-Nearest Neighbours (KNN), Convolutional Neural Networks (CNN), and Decision Tree Algorithm (DTA) are distinct machine learning models with unique approaches. KNN is a lazy learning algorithm that stores training data and performs computations during the prediction phase, classifying new points based on the 'K' nearest neighbors. CNNs, on the other hand, use convolutional, activation, pooling, and fully connected layers, with key hyperparameters such as epochs, batch size, optimizer, and learning rate fine-tuned for tasks like image recognition. In contrast, DTA builds models by recursively splitting the dataset using selected criteria, with hyperparameters like maximum depth and minimum samples per leaf optimized through cross-validation to ensure the model generalizes well on unseen data.

The performance analysis involves evaluating each model's ability, identifying which performed the best, and exploring the possible reasons for its success. Error analysis is essential to pinpoint common mistakes across models and uncover any patterns in those errors. For the Decision Tree Algorithm (DTA), feature importance should focus on the most significant criteria influencing the decision-making process. A comparative analysis of KNN, CNN, and DTA highlights the trade-offs between aspects such as profitability and model complexity. To measure performance across classes, confusion matrices should be provided for each model, while CNN's learning curves can illustrate training and validation progress over epochs. Understanding the strengths, limitations, and impacts of these models after training and evaluation requires detailed interpretation based on the experimental setup and results.

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

6 Results

1. Performance of KNN : The surgical on the test set, the KNN model attained an integrity of 80%. KNN performs best in datasets when nearly a distinct demarcation between classes; nevertheless, noisy data or big, high-dimensional arrays can cause it to perform poorly. The distance metric and the magnitude of neighbours (k) that are selected have a big harm on the outcomes.

2. Performance of Convolutional Neural Networks (CNN): The exactness of the test set attained by the CNN model was 20%. CNNs work well with image data and other grid-like structures are excellent at capturing spatial hierarchies in the data. From unprocessed input data, they can automatically learn and extract features. The high accuracy of the CNN model shows that it is good at identifying intricate patterns and characteristics in the information on hand. However, the model's performance affected by materials architecture chosen, the amount of training data, and the data augmentation techniques applied. CNNs are particularly suitable for image classification tasks, where their ability to learn hierarchical representations to raw leads to superior performance compared to traditional methods.

3. Decision Tree Algorithm (DTA) Performance: The DTA model achieved an accuracy of (%) on the simulated set. Understanding feature importance and the decision-making process's structure is made easier with the help of the intelligent choice tree model, which offers an understandable and transparent decision-making process. Nevertheless, over fitting while participating in dislocation difficulty might tighten the model's performance; these an abundance of by employing ensemble approaches (e.g., Random Forest) and pruning.

7 Conclusion

While this project provides a solid foundation about this topic obtaining machine learning, several for impending tasks may be explored to enrich this model's accuracy and robustness. Advanced Feature Engineering: Investigate and incorporate additional relevant features, such as lifestyle factors (e.g., diet, exercise), genetic markers, and detailed medical history. Genomic Data Integration: Utilize genetic information to identify hereditary risk factors for heart disease. Combining genomic data with traditional medical records can lead to more personalized and accurate predictions. AutoML: Implement automated machine learning (AutoML) tools to automate the method by which model selection, hyper parameter tuning, and feature engineering, ensuring that the best possible model is chosen efficiently.

DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1

References

Devi, R. R., Dharshini, P., Hemala, R., & Swetha, D. (2023). Heart Disease Prediction using Random Forest Classifier. 2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2023. https://doi.org/10.1109/ICDSAAI59313.2023.10452459

Kee, O. T., Harun, H., Mustafa, N., Abdul Murad, N. A., Chin, S. F., Jaafar, R., & Abdullah, N. (2023). Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review. Cardiovascular Diabetology, 22(1). https://doi.org/10.1186/s12933-023-01741-7

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine Learning Technology-Based Heart Disease Detection Models. Journal of Healthcare Engineering, 2022. https://doi.org/10.1155/2022/7351061

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). https://doi.org/10.3390/diagnostics14020144

Rani, G. E., Murugeswari, R., Siengchin, S., Rajini, N., & Kumar, M. A. (2022). Quantitative assessment of particle dispersion in polymeric composites and its effect on mechanical properties. Journal of Materials Research and Technology, 19, 1836–1845. https://doi.org/10.1016/j.jmrt.2022.05.147

Shah, A., & Patel, R. (2022). Heart Disease Prediction Based on Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 10(8), 1027–1036. https://doi.org/10.22214/ijraset.2022.46341

Srivenkatesh, M. (2020). Prediction of Cardiovascular Disease using Machine Learning Algorithms. International Journal of Engineering and Advanced Technology, 9(3), 2404–2414. https://doi.org/10.35940/ijeat.b3986.029320

Stonier, A. A., Gorantla, R. K., & Manoj, K. (2024). Cardiac disease risk prediction using machine learning algorithms. Healthcare Technology Letters, 11(4), 213–217. https://doi.org/10.1049/htl2.12053

Uçar, M. K., Nour, M., Sindi, H., & Polat, K. (2020). The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets. Mathematical Problems in Engineering, 2020. https://doi.org/10.1155/2020/2836236

Editors: Shilpa Sivashankar, S.Pandikumar, Pallavi M O
DOI:10.48001/978-81-966500-0-1-10 | ISBN: 978-81-966500-0-1