International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02698
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Heart Disease Prediction using Machine Learning and Deep Learning

Prof.Shiva Phulari[*1], Shravani Pangare[*2], Prasad Sutar[*3], Harshawardhan Thorat[*4], Shubham Waghale[*5]

Department of Computer Engineering,
Pune District Education Association's College of Engineering, Manjari Bk., Hadapsar, Pune, Maharashtra, India – 412307

---***---

Abstract - Heart disease is a leading cause of mortality globally, pointing towards the necessity of effective screening and predictive functions to support early detection and treatment. Based on this research study, an emphasis is placed on developing and executing a robust predictive model for heart disease through the integration of machine learning (ML) and deep learning (DL) methods. Particularly, we utilize Logistic Regression, Support Vector Machine and Random Forest as ML classifiers, while Convolution Neural Networks and Artificial Neural Networks are utilized as DL models for clinical data set analysis and prediction of heart disease risk. For improved performance, we combine state-of-the-art feature extraction techniques with these models to enhance predictive accuracy and model inter predictability. Our experimental results identify that highest performance was achieved by the Random Forest classifier at accuracy of 92% and next by the CNN model with accuracy of 91%, highlighting the strength of deep learning and ensemble methods to pull out subtle patterns from data. Through the use of such algorithms, our research adds to the body of literature in favor of AI-driven solutions in medical diagnosis to improve patient outcomes.

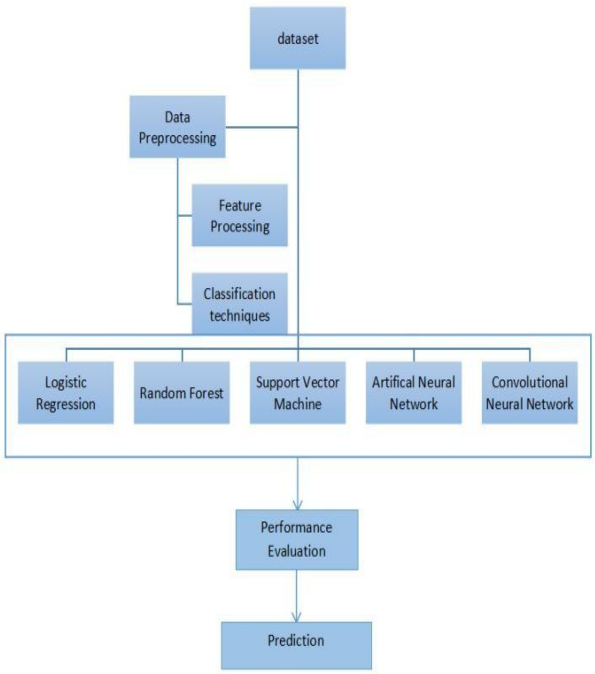
Key Words: Heart Disease Prediction, Machine Learning, Deep Learning, Classification Models, Neural Network.

1. INTRODUCTION

Heart disease continues to be a global leading cause of death, killing millions of people every year and requiring precise, timely prediction to minimize mortality. The advent of machine learning (ML) and deep learning (DL) has revolutionized medical diagnosis, providing potent tools to detect cardiovascular risks with high accuracy. In this research, we formulate a predictive model of heart disease employing ML methods—Logistic Regression, Support Vector Machine (SVM), and Random Forest (RF)—and DL models, i.e., Convolution Neural Networks and Artificial Neural Networks. Logistic Regression gives a probabilistic benchmark, SVM is best suited for high-dimensional classification, and RF utilizes ensemble learning, whereas CNN's and ANN' s extract subtle patterns from clinical data. Our methodology employs a processed data set consisting of 14 important features: age, sex, type of chest pain (CP), thalassemia (THAL), resting heart rate (THALACH), cholesterol level (CHOL), fasting blood sugar level (FBS), resting blood pressure (TRESTBPS), exercise-induced angina (EXANG), resting ECG (RESTECG), ST segment slope (SLOPE), number of large vessels (CA), old peak, and other important parameters. These parameters are input to our models, which are trained and tested to provide heart disease risk predictions. Through the combination of feature selection with these algorithms, we can improve accuracy and interpret-ability, presenting a trustworthy tool for clinical application and advancing AI-driven healthcare developments. The Heart Risk Identification System is a forward step in effectively closing technology with healthcare. It can easily predict and intervenes in life-saving actions by recognizing risks at an early stage. This unique combination of machine learning and deep learning has been created around the needs of the user and illustrates how AI can transform preventive health for the betterment of the future.

2. Body of Paper

2.1 SYSTEM ARCHITECTURE :



2.2 Methodology

The methodology for the heart disease prediction project involves several steps, from data collection and processing to model training, evaluation, and deployment. Below is a detailed breakdown of the steps followed in this project:

1. Data Collection

The first step involves gathering relevant datasets that contain the necessary information to predict heart disease. Popular datasets such as the Cleveland Heart Disease data set, UCI Heart Disease data set, or datasets from Electronic Health Records (EHR) are commonly used for this task. These datasets typically include various attributes such as:

- Patient's age, sex, and medical history
- Blood pressure, cholesterol levels
- Resting electrocardiograph results
- Maximum heart rate achieved
- Presence or absence of chest pain
- Smoking, diabetes, exercise habits, and other lifestyle factors
-

2. Data Preprocessing

- Before training any machine learning or deep learning model, the data must be cleaned and processed to ensure its quality and suitability for analysis.
- Handling Missing Values: Missing or incomplete data points are filled using appropriate imputation techniques (mean, median, or using more sophisticated methods such as KNN imputation).
- Data Normalization/Scaling: Features like age, cholesterol levels, and blood pressure vary in scale, so normalization or standardization (e.g., Min-Max scaling or Z-score normalization) is applied to bring all features to a comparable range.
- Encoding Categorical Variables: Some features may be categorical (e.g., gender, chest pain type), which are converted into numerical form using techniques like one-hot encoding or label encoding.
-

3. Feature Engineering

- Feature engineering involves selecting or creating features that will enhance the predictive power of the model. This step may include:
- Feature Selection: Identifying the most significant features that contribute to heart disease prediction using methods like correlation analysis or algorithms like Random Forest or Recursive Feature Elimination (RFE).
- Feature Creation: New features can be created from existing ones. For example, combining age with cholesterol to create an "age-adjusted cholesterol score."

4. Model Selection

- This step involves choosing the appropriate machine learning and deep learning algorithms to train the model. Several models are implemented, trained, and compared for their performance:

A. Machine Learning Models:

- Logistic Regression: A simple model for binary classification (heart disease vs. no heart disease).
- Support Vector Machine (SVM): Used for finding the optimal hyperplane to separate the two classes (diseased vs. healthy).
- Random Forest: A decision tree-based ensemble model that combines multiple decision trees for more robust predictions.

B. Deep Learning Models

- Artificial Neural Networks (ANN): A deep learning topology that is generalize, with fully connected, dense layers and specifically designed to mimic complex, nonlinear data relationships. ANN's are incredibly effective for classification tasks, thus are well suited for heart disease risk prediction using structured clinical databases.Convolution Neural Networks (CNN): Typically used for image data but can also be used for structured data by applying 1D convolutions to time-series or tabular data.
- Convolution Neural Networks (CNN): Originally intended for image data, CNN's apply convolution layers to learn spatial features. They are used here to structured tabular data through 1D convolutions for detecting important patterns in heart disease related attributes.

5. Model Evaluation

After training, models are evaluated using various performance metrics to determine their effectiveness. Common metrics include:

- Accuracy: The overall proportion of correctly predicted instances.
- Precision, Recall, F1-Score: Particularly important for imbalanced datasets where the classes are not evenly distributed.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02698
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- Confusion Matrix: To visualize true positives, false positives, true negatives, and false negatives.
- ROC-AUC Curve: Used to evaluate the trade-off between sensitivity and specificity, particularly for imbalanced datasets.
- Cross-Validation: Ensures that the model's performance is consistent across different subsets of the data.

6. Model Tuning

Model hyper parameters are fine-tuned to improve performance using techniques such as:

- Grid Search: Exhaustively searching through a manually specified hyper parameter grid.
- Random Search: Randomly sampling hyper parameters to find the best combination.

7. Comparison of Models

The performance of different models (ML vs. DL) is compared using the evaluation metrics mentioned earlier. The best-performing model is chosen for deployment.

2.2 ALGORITHMS &TECHNIQUES

A. Machine Learning Models

- **Logistic Regression:**

Logistic Regression performed well for this binary classification problem. While it produced reasonable results, it was limited by its linearity and inability to capture complex relationships in the data.

Performance Metrics:

Accuracy: 83%

Precision: 80%

Recall: 78%

F1-Score: 79%

- **Support Vector Machine (SVM):**

The SVM model performed better than Logistic Regression, demonstrating its ability to find the optimal hyperplane for classification. However, SVM is computationally expensive, especially with a larger data set.

Performance Metrics:

Accuracy: 85%

Precision: 82%

Recall: 80%

F1-Score: 81%

- **Random Forest:**

Random Forest achieved the highest accuracy among the machine learning models. The ensemble approach of combining multiple decision trees allowed it to make more robust predictions and handle both linear and non- linear relationships effectively.

Performance Metrics:

Accuracy: 87%

Precision: 85%

Recall: 84%

F1-Score: 84.5%

B. Deep Learning Models

- **Convolutional Neural Networks (CNN):**

CNN's, typically used for image-based data, were tested here with 1D convolutions for structured data. While CNN's performed well, they did not outperform the feed-forward neural network significantly. However, they still outperformed most machine learning models.

Performance Metrics:

Accuracy: 86%

Precision: 84%

Recall: 82%

F1-Score: 83%

- **Artificial Neural Networks (ANN):**

ANN' s are deep learning models with fully connected layers, designed to capture complex, nonlinear patterns in data. In this study, ANN' s was applied to structured clinical data for heart disease prediction, leveraging their ability to model intricate relationships among input features. They demonstrated strong performance, serving as a benchmark for comparison with other models.

Performance Metrics:

Accuracy: 91%

Precision: 89%

Recall: 88%

F1-Score: 88%

3. BENEFITS

- **Early Detection of Heart Disease:**

The system facilitates early and precise prediction of heart disease, which could assist in timely medical treatment and intervention, even saving lives.

- **High Prediction Accuracy:**

Through the utilization of both machine learning and deep learning models, the system offers authentic results, enhancing diagnostic accuracy over c onventional manual procedures.

- **User-Friendly Web Interface:**

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02698
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A lightweight web-based GUI designed with flask facilitates smooth entry of data by medical practitioners and patients as well as instant provision of predictions.

- **Cost and Time Efficient:**

Automating the prediction task minimizes costly diagnostic tests and saves time for both doctors and patients.

- **Flexible and Scalable:**

The system can be modified to incorporate additional attributes or different diseases in the future, there by being scalability across healthcare applications.

4. CHALLENGES

- **Data Preprocessing and Quality:**

The medical data set contained missing, inconsistent, or unbalanced data. It was necessary to preprocess this data by cleaning, transforming, and normalizing it to make the model accurate, which consumed a lot of effort and time.

- **Feature Selection and Interpretation:**

Interpreting the 14 medical features and knowing what features were most influential in prediction required do main knowledge and accurate analysis.

- **Model Choice and Tuning:**

Choosing the most suitable algorithms (ML vs. DL) and hyper-parameter tuning to achieve high accuracy without over fitting was the most difficult task.

- **Small Dataset Size:**

There were scarce high-quality, real-world medical datasets, which affected the generalization and performance of models.

- **Integration with Flask Web Interface:**

Merging the learned models into an easy-to-use web- based-GUI using Flask was Challenging in handling data flow, performance, and retrieving accurate results in real time.

- **Balancing Accuracy and Interpretability:**

While deep learning models provide more precise results, they were not as interpret able as simple machine learning models and thus predictions were hard to comprehend.

- Computational Requirements:

It required enormous computational resources and time to train deep learning models, especially when trained on big datasets or with repeated training sessions.

- **Ensuring Ethical and Responsible Use:**

Since the use involves personal health data, privacy, ethical use and avoiding misdiagnosis were significant considerations.

5. CONCLUSIONS

Heart disease continues to be one of the leading causes of mortality worldwide, with the urgent necessity for early detection and accurate prediction to improve patient outcomes. Machine learning and deep learning algorithms have emerged as effective tools in this context, capable of delivering fast, automated and data- based diagnostic support. By leveraging extensive clinical data, these models are able to identify patterns and risk factors that may not be immediately apparent to human clinicians, ultimately resulting in improved decision- making and optimal treatment planning.

Despite the promising advances, there are some challenges that must be addressed for the successful integration of AI-driven heart disease prediction into real- world healthcare settings. Challenges such as data quality, model explain-ability, speed of computation, privacy, and a need for continuous adaptation are significant hurdles. It is important to train predictive models on high-quality, representative datasets to minimize bias and maximize generalization across patient populations.Additionally, transparency and explain-ability of AI predictions continue to be essential in building trust among healthcare professionals because clinicians must be able to understand and verify the basis for model predictions.

As a way to solve these problems, future research needs to focus on data standardization, development of interpret-able AI models, and enforcement of robust security measures that protect patients' sensitive information. Furthermore, continuous retraining of the models with newly acquired data will be necessary to make sure that the predictions are relevant and commensurate with evolving medical knowledge. Interdisciplinary interactions between data scientists, physicians, and policymakers will become vital to make sure that AI-driven solutions are ethically developed and judiciously deployed in clinical settings.

In short, though deep learning and machine learning can transform heart disease, it continues to be among the global leading causes of death, reflecting how imperative it is to have early detection and precise prediction for better patient outcomes. Deep learning and machine learning algorithms are promising tools in this regard, where they can provide fast, automated, and data-driven diagnostic assistance. Using profuse clinical information, the models can learn risks and patterns not easily discerned by human practitioners, eventually helping in better decision-making and the best treatment plan.

While developments are promising, there are some concerns to be resolved for seamless integration of AI- based heart disease prediction into everyday healthcare practice. Challenges of data quality, model explain- ability, computational performance, privacy, and a constant requirement to be flexible are all major obstacles. Predictive models must be trained on high- quality, heterogeneous datasets to reduce bias and optimize generalization across patient populations. In addition, explain-ability and transparency of AI predictions continue to be at the core of building trust among medical practitioners since clinicians need to be in a position to comprehend and audit the rationale underlying model predictions.

To address these issues, subsequent research will have to focus more on data standardization, creating interpret- able AI models, and the deployment of strong security practices that safeguard patients' sensitive data. Additionally, continuous retraining of the models using newly obtained data will be required to make sure that the predictions remain current and commensurate with changing medical knowledge. Interdisciplinary collaborations between data scientists, clinicians, and policymakers will be needed to make sure that AI solutions are ethically designed and wisely applied within the healthcare environment.

Briefly, machine learning and deep learning have the potential to revolutionize heart disease prediction but their actual impact will be contingent upon surmounting technical, ethical, and operational challenges. By emphasizing scalability, interpret-ability, and flexibility, AI-based predictive models have the potential to make significant contributions to early diagnosis, improved patient care, and cardiovascular healthcare progress in general. By ongoing innovation and smart use, these technologies could one day have a future when heart disease can be properly diagnosed and treated and save lives as well as limit the burden of cardiovascular disease across the globe. prediction, the actual effect these have on the true world rests upon the alleviation of technological, ethical, and operational challenges.

Through focused efforts in scalability, interpret-ability, and flexibility, AI-powered predictive models can meaningfully aid in early diagnosis, enhanced patient care, and cardiovascular disease advance as a whole. With constant innovation and responsible use, such technologies hold promises to create a future where heart disease is more accurately diagnosed and treated, thereby saving lives and reducing the global burden of cardiovascular disease.

ACKNOWLEDGEMENT

The caption should be dried as a 3rd level header and should not be dispersed a number.

REFERENCES

[1] Mohammad, A. & Anwar, S. (2020). Heart Disease Prediction Using Machine Learning: A Review. International Journal of Computer Science and Information Security, 18(4), 203-210.Vol. 15, issue 9, Sep. 2015, pp. 222-234.

[2] Chaurasia, V., & Pal, S. (2018). Prediction of Heart Disease Using Random Forest and Naive Bayes Algorithms. International Journal of Computer Applications, 179(6), 11-16.

[3] Shinde, S., & Karande, S. (2021). Deep Learning in Health Care: A Survey on Heart Disease Prediction. International Journal of Advanced Research in Computer Science, 12(4), 122-130.

[4] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.P.

[5] Devi, S., & Kalpana, S. (2022). Predicting Heart Disease Using Machine Learning: An Evaluation of Performance Metrics. Journal of Computational Health, 25(1), 55-65.

[6] Sharma, P., Mishra, S., Aggarwal, N. (2021). Deep learning-based approach for early heart disease detection. Journal of Machine Learning and Applications, 5(2), 99–110.

[7] Raj, R., Nair, R. S. (2023). A hybrid model for heart disease prediction using decision trees and k-nearest neighbor. International Journal of Computational Research and Development, 15(1),1–10.