KeAi
CHINESE ROOTS
GLOBAL IMPACT

International Journal of Innovation Studies

HEART DISEASE PREDICTION USING RNN WITH ADVANCED PREPROCESSING AND FEATURE EXTRACTION TECHNIQUES

A.R.Sangeetha
Research Scholar, Department of Computer Science, Dr. M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India, E-Mail: sangeethaar711@gmail.com[1]

Dr. S. Ismail Kalilulah
Professor, Department of Computer Science Engineering, Dr. M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India, E-Mail: ismailkalilulah125@gmail.com[2]

Abstract:- *One of the biggest causes of death worldwide is still heart disease, which emphasizes the need for reliable predictive models to support early detection and treatment. Developing precise and effective prediction systems is essential since conventional approaches for diagnosing cardiac disease sometimes rely on difficult and time-consuming processes. By utilizing huge datasets and complex algorithms, ML(Machine Learning) techniques have become increasingly effective tools for improving diagnostic procedures in recent years. In this work, we used a Recurrent Neural Network (RNN) in conjunction with extensive preprocessing and feature extraction techniques to construct an advanced system for heart disease prediction. Min-max scaling, IQR (Interquartile Range), and z-score normalization were used in the preprocessing stage to standardize the data and guarantee consistent input for the RNN model. Chi-square tests to extract features, substantially reduced the number of dimensions and improved the relevance of the features. The heart disease dataset from the UCI ML Repository was used to assess the suggested system. The results showed that model performance was greatly enhanced across accuracy, precision, and recall measures when IQR and Chi-square tests were combined with RNN.*

Keywords:- *Heart Disease Prediction, Recurrent Neural Network (RNN), Min-Max Scaling, Z-Score Normalization, Interquartile Range (IQR), Chi-Square Test, Feature Extraction, Machine Learning, Data Preprocessing, Classification*

1. INTRODUCTION

Heart disease continues to rank among the top causes of death globally, emphasizing the importance of early detection and effective intervention to alleviate its worldwide impact. Traditional diagnostic methods, although useful, frequently require complex and time-intensive processes that may hinder prompt diagnosis and treatment. With the increasing size and complexity of healthcare datasets, machine learning (ML) techniques have demonstrated significant potential to improve diagnostic accuracy, optimize workflows, and offer actionable insights for medical professionals. By utilizing sophisticated algorithms, ML has become a powerful tool for heart disease prediction, providing more accurate and efficient diagnoses than conventional methods.

Numerous risk factors, including a poor diet, inactivity, and heavy alcohol and tobacco use, contribute to heart disease (Tao R et al., 2019). Adopting a healthy lifestyle, which includes cutting back on salt in the diet, eating more fruits and vegetables, getting regular exercise, and giving up alcohol and tobacco use, lowers the risk of heart disease (Spencer R et al., 2020). The amount of data generated by the healthcare industry about patients, illnesses, and diagnoses is enormous, but it is not being properly examined, which prevents it from offering the value it should. The leading cause of death is heart disease (Ch. Anwar ul Hassan, et al., 2022). Figure 1 shows the difference between Normal Artery and Blocked Altery.

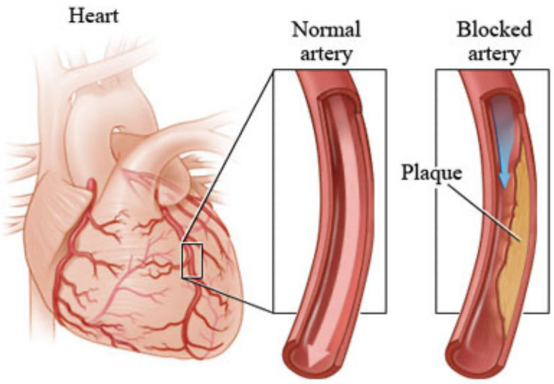

Fig 1 Normal Artery Vs Blocked Artery (www.inrechopen.com)

Traditional diagnostic methods for heart disease prediction face several challenges, mainly due to their reliance on manual interpretation and limited use of advanced technology. Techniques like ECGs, stress tests, and physical exams often depend on the expertise of the clinician, which can result in inconsistencies in diagnosis. Moreover, these methods tend to detect heart disease only in its later stages when symptoms are more noticeable, reducing the chances for early intervention and effective treatment.

Another significant challenge is the lack of precision in traditional diagnostics. False positives and false negatives are common, leading to either unnecessary treatments or missed diagnoses. Furthermore, these methods typically do not account for individual risk factors such as genetic predisposition, lifestyle choices, or other health conditions, which are crucial in predicting heart disease. The lack of integration with modern tools like advanced imaging or genetic testing limits the effectiveness of these traditional approaches, highlighting the need for more innovative and data-driven solutions in heart disease prediction.

By utilizing huge datasets and complex algorithms, ML techniques have become increasingly effective tools for improving diagnostic procedures in recent years. The goal of this work is to enhance the prediction of heart disease by utilizing an RNN in conjunction with sophisticated feature extraction and preprocessing techniques.

In this study, we present an innovative heart disease prediction system that leverages the power of a RNN alongside thorough data preprocessing and feature extraction methods. The preprocessing phase utilized techniques such as min-max scaling, IQR, and z-score normalization to standardize the dataset and make it suitable for training the model. To further enhance model performance, Chi-square tests were employed for feature extraction, effectively reducing dimensionality and improving feature relevance. By utilizing the heart disease dataset from the UCI ML Repository, the proposed system achieved significant improvements in accuracy, precision, and recall when IQR-based preprocessing and Chi-square feature selection were integrated with the RNN. These results highlight the effectiveness of combining advanced preprocessing, feature selection, and neural network architectures for accurate heart disease prediction.

2. LITERATURE SURVEY

Cardiovascular disease is a generic term for any serious heart-related illness. Because heart disorders can be fatal, scientists are concentrating on developing intelligent systems that employ ML algorithms to precisely diagnose heart ailments based on electronic health data. Using patient data on key health indicators, C. Boukhatem et al., 2022 provides several ML techniques for predicting cardiac disorders. To construct the prediction models, the study presented four classification techniques. Before creating the models, phases for feature selection and data pretreatment were completed. The models were assessed using the F1-score, accuracy, precision, and recall metrics. With 91.67% accuracy, the SVM model yielded the best results.

Identifying and forecasting cardiovascular disease are vital medical duties that help cardiologists appropriately diagnose and treat their patients. Since the purpose of machine learning algorithms is to find comparable patterns in patient data. Using ML Patterns, clinicians can reduce misdiagnosis in the diagnosis of CV disorders. to lower the mortality rate from heart-related illnesses. The model utilized in this paper's design can predict the conditions of CVDs with accuracy. A novel initial k-mode clustering technique is proposed by S. Kale et al., 2024, which may improve classification accuracy. The parameters of an implemented machine learning model can be hyper-tuned using a variety of models. The proposed model is applied to a real-world dataset consisting of more than 300 cases from Kaggle.

Heart disease, sometimes called cardiovascular disease, is the leading cause of death globally over the last several decades. It encompasses a variety of disorders that affect the heart. It links several heart disease risk factors and emphasizes the urgency of finding timely, accurate, and reasonable ways to diagnose the condition early on and begin treating it. One popular method for handling massive amounts of data in the healthcare industry is data mining. Researchers analyze vast amounts of intricate medical data using a variety of data mining and ML approaches, assisting medical personnel in the prediction of cardiac disease.

Various attributes associated with heart disease are presented by Devansh Shah et al. 2020, along with a model based on supervised learning methods. It makes use of the current dataset from the UCI heart disease patient repository's Cleveland database. There are 76 attributes and 303 instances in the collection. Just 14 of these 76 attributes are taken into account during testing, which is crucial to proving the effectiveness of various algorithms. The purpose of this research work is to estimate the patient's risk of acquiring heart disease. The findings show that the K-nearest neighbor yields the highest accuracy score.

The identification and prognosis of cardiovascular illness are critical medical responsibilities that assist cardiologists in correctly classifying patients and treating them accordingly. Because ML algorithms can identify patterns in data, their applications in the medical field have grown. Diagnosticians can decrease misdiagnosis by classifying cardiovascular disease occurrence using ML. In 2023, Chintan M. Bhatt et al., create a model that accurately predicts cardiovascular disorders to lower the mortality rate associated with these conditions. This research suggests a Huang beginning k-modes clustering approach that can increase classification accuracy. Several models are employed, To maximize the outcome, GridSearchCV was utilized to hyper-tune the employed model's parameters. On a real-world dataset of 70,000 cases from Kaggle, the suggested model is used. Accuracy was attained by training models using 80:20 split data. This underlying research leads to the conclusion that MLP with cross-validation has performed more accurately than any other method. At 87.28%, it had the highest accuracy.

The term artificial intelligence (AI) is used to describe a broad range of automated systems that require "intelligence" to perform particular tasks. Across the past ten years, AI approaches have become more and more common across a wide range of scientific sectors, including cardiovascular medicine. The prevalence of heart disease has increased as a result of the increased awareness of risk factors and the improved prognosis of patients going through cardiovascular events. This has led to the need for accurate patient identification at high risk of disease growth and progression. Some of the drawbacks that impede the effectiveness of traditional regression models might be solved by AI-based predictive models. However, to ensure the safe and efficient use of AI techniques in routine clinical practice, understanding the potential hazards of these approaches is necessary for the successful application of AI in this field. The purpose of Mauro Chiarito et al., 2022 is to provide an overview of the benefits and drawbacks of various artificial intelligence techniques and their possible use in the field of heart medicine, with an emphasis on the creation of risk assessment tools and models for prediction.

Currently, heart failure disease is a complex clinical condition that affects a larger proportion of the global population. Hospitals and cardiac centers mostly rely on ECG to assess and diagnose heart failure in its early stages. One may think of the ECG as a standard tool. Early identification of heart disease is a major priority for healthcare. The various ML approaches based on a quick analysis of heart disease diagnosis are presented by Umarani Nagavelli et al. in 2022. First, heart disease is predicted using Naïve Bayes and a weighted approach. The second method is automatic and analyzes the location and diagnosis of coronary heart disease based on information theory, time field, and frequency domain aspects. In this method, the best performing two classifiers, such as support vector machine (SVM) and XGBoost, are chosen for classification. The third one uses an enhanced SVM based on the duality optimization strategy that was also examined to identify heart failure automatically. In the medical field, ML can be used for illness detection, diagnosis, and prediction. This paper's main goal is to provide physicians with a tool to aid in the early diagnosis of cardiac issues. It will be simpler to treat patients successfully and stay away from major consequences as an outcome.

3. PROPOSED SYSTEM

Globally, heart disease continues to be the primary cause of morbidity and mortality. Predicting heart illness accurately and on time can greatly improve early detection and intervention, possibly saving lives and improving patient outcomes. Conventional techniques for predicting heart disease frequently depend on static characteristics like age, blood pressure, and cholesterol levels. These techniques might not, however, be able to capture the temporal dynamics of a patient's health, which might be vital for forecasting the course of a disease. RNN is an interesting and innovative ML application in the medical field for heart disease prediction. Figure 2 shows the structure of the suggested heart disease prediction model.

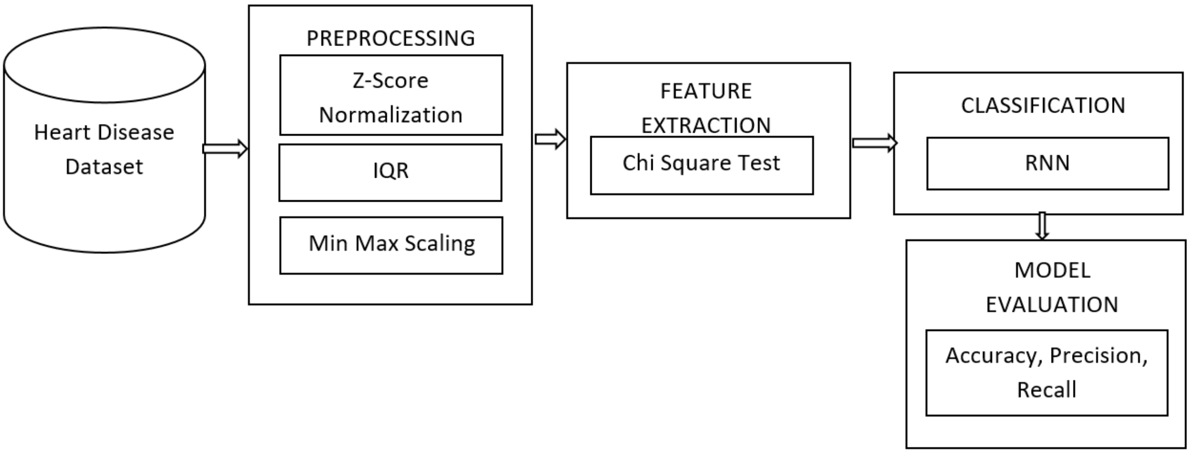

Fig 2 Framework of Proposed Heart Prediction Model

3.1 Data Collection

Gathering the dataset on heart illness from the UCI ML Repository is the initial stage in the procedure. This dataset contains several heart health-related characteristics, including age, blood pressure, cholesterol, and more. Obtaining a thorough and representative dataset that includes the essential characteristics for precise cardiac disease prediction is the aim.

3.2 Importance of Data Preprocessing

Data preprocessing is a crucial step in heart disease prediction models, as it directly impacts the accuracy, efficiency, and reliability of the predictions. To guarantee that the dataset is in a format appropriate for analysis, data preparation is essential. The data is normalized by applying normalization algorithms.

Managing Missing Data

Medical datasets often have gaps due to incomplete records or missed tests, which can lead to inaccurate or biased predictions. For instance, a dataset with missing blood pressure or cholesterol values might fail to capture the complete health profile of patients. Data preprocessing helps by applying techniques such as imputing missing values with statistical measures or advanced algorithms, or by removing incomplete records, ensuring that the dataset remains consistent and reliable for analysis.

Identifying and Addressing Outliers

Outliers, or data points that differ significantly from other observations, are common in healthcare datasets. For example, extreme cholesterol or blood pressure values can distort statistical analyses and machine learning models. These anomalies can lead to poor model performance or erroneous predictions. Preprocessing methods such as statistical analysis, visualization tools, and interquartile range techniques help in identifying and treating outliers, either by removing them or capping their values within a reasonable range.

Scaling Data for Consistency

Medical data often consists of features with varying units and ranges, such as heart rate, age, and cholesterol levels. If these differences are not addressed, machine learning models may disproportionately weigh larger values, affecting their ability to learn effectively. Normalizing or standardizing data ensures that all features contribute equally to the model by transforming them into a common scale, enhancing training and prediction accuracy.

Removing Redundancy and Noise

Healthcare datasets frequently include irrelevant or redundant features, such as patient identifiers or unrelated administrative data, which add noise and complexity. Including these features can hinder model performance and slow down processing. Preprocessing removes unnecessary information and emphasizes relevant attributes, often using techniques like feature selection or dimensionality reduction. This improves both the efficiency and accuracy of predictive models.

Ensuring Class Balance

Class imbalance, where one category (e.g., patients without heart disease) dominates over others, is a significant challenge in medical data. Such imbalance can cause prediction models to favor the majority class, reducing their ability to detect the minority class, which is often critical in disease prediction. Techniques like oversampling minority classes, undersampling majority classes, or synthetic data generation help balance the dataset, enabling models to perform equitably across all categories.

Encoding Categorical Information

Medical records often contain categorical variables such as gender, smoking status, or family history, which must be converted into numerical formats for most predictive algorithms. Encoding techniques such as one-hot encoding or label encoding transform these categories into formats that machines can process while preserving their interpretive value. This step ensures that categorical features are effectively utilized in the prediction process.

Improving Data Quality by Reducing Errors

Raw medical data may contain errors due to equipment faults, transcription mistakes, or duplicate records. These errors degrade the quality of the data and can mislead the model. Preprocessing focuses on error detection and correction through cleaning processes, removing duplicates, and validating entries to enhance the dataset's reliability. High-quality data ensures that the predictive model is built on a strong and accurate foundation.

3.2.1 Min-max scaling

For consistent scaling across features, min-max scaling converts feature values into a [0, 1] range. Ensuring that every feature contributes equally to the model and enhancing the convergence of optimization algorithms during training necessitate this. Min-max scaling uses the following formula to convert each value x in a feature with a minimum value (x_{min}) and a maximum value (x_{max}) to a scaled value (x′) in the range [0, 1]:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} ---------(1)$$

Where:

- x is the original value of the feature.
- x_{min} is the minimum value of the feature in the dataset.
- x_{max} is the maximum value of the feature in the dataset.
- x' is the scaled value.

When features have varying units or ranges, min-max scaling is especially helpful since it unifies them into a single scale, improving model performance and training.

3.2.2 Z-score normalization

Z-score normalization helps stabilize the RNN's training process and enhances model performance by standardizing features to have a mean of 0 and an SD(Standard Deviation) of 1. Feature values are transformed to have a mean of 0 and an SD of 1 using Z-score normalization, sometimes referred to as standardization. This method frequently enhances the performance and convergence of ML models and aids in bringing the data similar across features with various sizes or units.Z-score normalization uses the following formula, given a feature x with values, to convert each value x to a normalized value x':

$$x' = \frac{x - \mu}{\sigma} ---------(2)$$

where:

- x is the original value of the feature.
- μ is the mean of the feature values.
- σ is the standard deviation of the feature values.
- x' is the standardized value.

3.2.3 IQR

Outliers can be removed by computing the range between each feature's first quartile (Q1) and third quartile (Q3) using the IQR method. This helps the model focus on the most important information. An indicator of statistical dispersion, the IQR shows the range that the central 50% of data points fall inside. By concentrating on and removing extreme values from the middle 50% of the data, it is frequently used for choosing features and outlier detection. This lessens the impact of noise and outliers on the dataset.

Calculate the Q1 and Q3:

- **Q1:** The value below which 25% of the data points fall.
- **Q3:** The value below which 75% of the data points fall.

The sorted data can be used to compute these quartiles. Sorting the data and locating the values at positions $\frac{n+1}{4}$ and $\frac{3(n+1)}{4}$, respectively, will yield the quartiles for a dataset containing n data points.

Calculate the IQR:

$IQR = Q3 - Q1 - - - - - - - - (3)$

The middle 50% of the data's spread is represented by this value.

Identify Outliers:

The IQR can be used to identify outliers. Generally speaking, values that are above Q3+1.5×IQR or below Q1−1.5×IQR are regarded as outliers.

3.3 Feature Extraction using the Chi-square test

To determine if category features are independent of the target variable, the chi-square test is utilized. By concentrating on key characteristics, features with substantial chi-square values are chosen for inclusion, improving the model's capacity for precise prediction.

Make a contingency table for each category feature against the target variable (heart disease presence, for example) before using the Chi-square test for feature selection in a heart disease dataset. Utilizing the Chi-square formula$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, where O_i stands for the observed frequency and E_i for the predicted frequency, determine the expected frequency for each cell in the table based on the overall proportions. To assess significance, utilize the p-value or compare the Chi-square statistic to a crucial value from the Chi-square distribution table. A low p-value (usually less than 0.05) suggests that the feature is relevant to your model since it shows a meaningful relationship with the target variable.

3.4 Model Training

After the data is ready, the preprocessed and feature-extracted dataset is used to train the RNN. Because of the RNN architecture's ability to handle sequential input, training entails utilizing optimization methods to change the network's weights and biases to reduce the loss function. The goal of this step is to develop a model that, using the features that have been provided, can reliably predict cardiac disease.

RNN is a family of neural networks that are especially helpful for tasks involving temporal or sequential data. While RNNs are effective for sequence-based tasks, other classifiers or conventional neural networks may be a better fit for classification tasks involving static features, such as usual cardiac disease. The RNN modifies its hidden state h_t at each time step t by taking into account the previous hidden state h_{t-1} and the current inputx_t. The equation is:

$h_t = \tanh(W_h \,.\, h_{t-1} + W_x . x_t + b) - - - - - - - - (4)$

where:

- W_h is the weight matrix for the hidden state
- W_x is the weight matrix for the input.
- b is the bias term.
- $\tanh$ is the activation function that introduces non-linearity.

The RNN produces an output y_t for each time step after processing the full sequence, usually by using:

$$y_t = W_y . h_t + b_y - - - - - - - - (5)$$

where:

- W_y is the weight matrix for the output.

b_y is the output bias.

RNNs offer significant advantages in heart disease prediction, particularly due to their ability to process and analyze time-series data, which is common in healthcare. RNNs excel at capturing the temporal dependencies within sequential patient data, such as ECG signals or patient history over time. By learning patterns in these dynamic sequences, RNNs can detect early signs of heart disease, even before they become clinically apparent. This ability to recognize complex patterns from continuous data allows for more accurate, real-time prediction, improving early diagnosis and enabling timely intervention.

Additionally, RNNs can integrate diverse sources of patient information, such as clinical records, genetic data, and lifestyle factors, providing a more holistic approach to heart disease prediction. This capability enables the model to consider a wide range of variables simultaneously, improving its predictive power. RNNs also offer the potential for personalized predictions, tailoring risk assessments based on individual patient profiles. As the model continuously learns from new data, it can adapt to changing patterns in patient health, making it a valuable tool for ongoing monitoring and management of heart disease risk.

Model Evaluation

After training, metrics like accuracy, precision, and recall are used to assess the model's performance. These measures shed light on the model's ability to discriminate between people who have and don't have heart disease. The test subset is used for validation to evaluate the model's applicability and generalizability in real-world situations.

4. RESULTS AND DISCUSSION

4.1 Dataset Description

The dataset is a multivariate collection designed for classification tasks within the domain of health and medicine. It comprises 303 instances and features a diverse set of attribute types, including categorical, integer, and real values. This varied feature set allows for a comprehensive analysis and modeling of health-related conditions or outcomes. The multivariate nature of the dataset provides a rich ground for exploring relationships between different health indicators and predicting classifications based on a combination of these attributes.

4.2 Performance Metrics

Accuracy, Precision, and Recall are crucial measures for assessing your model's effectiveness in predicting heart disease. The following components of a confusion matrix can be used to calculate these metrics:

True Positives (TP): The number of cases when heart illness was accurately predicted (i.e., the prediction and the actual state are both positive).

True Negatives (TN): Total number of cases where the absence of heart disease was accurately anticipated.
False Positives (FP): The number of cases when the prognosis for heart disease was erroneous (predicted positive but real negative).
False Negatives (FN): The number of cases when the prediction for the absence of heart disease was off (predicted negative but real positive).

Accuracy:
The accuracy of the model is measured by how well it predicts both heart disease and non-heart disease.

$$Accuracy - \frac{TP + TN}{TP + TN + FP + FN} - - - - - - - - (6)$$

Precision (also known as Positive Predictive Value):

Precision quantifies the proportion of cases correctly identified as heart disease.

$$Precision = \frac{TP}{TP + FP} - - - - - - - - -(7)$$

Recall (also known as Sensitivity or True Positive Rate):

Recall quantifies how many real-world instances of heart disease the model accurately recognized.

$$Recall = \frac{TP}{TP + FN} - - - - - - - - -(8)$$

Table 1 Machine Learning Models Vs Metrics

ML Model/Metrics	Accuracy	Precision	Recall
IQR+ CST+RNN	95.50	0.93	0.91
Z-Score+ CST +RNN	90.25	0.89	0.87
Min Max+ CST+RNN	89.42	0.87	0.85

Figure 3 shows the **accuracy** of ML models that utilize various data preprocessing techniques in combination with CST and RNN.

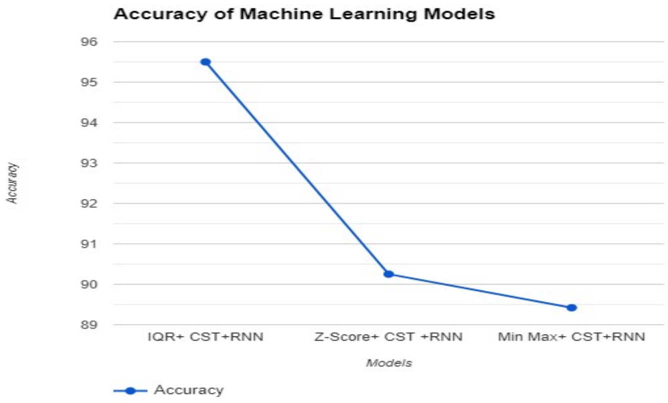

Fig 3 Accuracy Vs ML Models

Figure 4 shows the precision and recall of ML models that utilize various data preprocessing techniques in combination with CST and RNN.

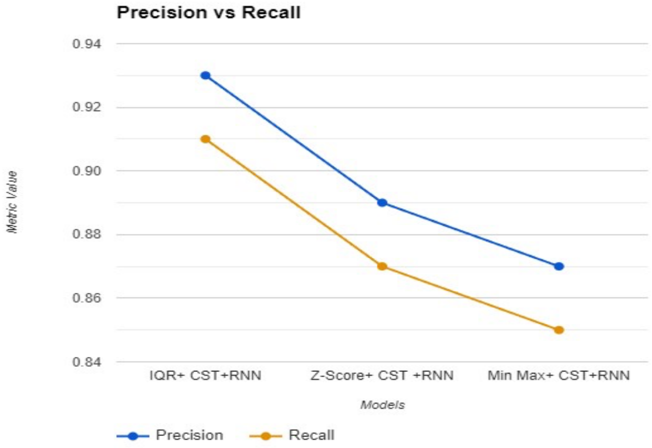

Fig 4 Precision and Recall Vs ML Models

With an outstanding accuracy of 95.50%, precision of 0.93, and recall of 0.91, the IQR + CST + RNN model performs better than the other approaches. This model has the highest overall accuracy by accurately detecting both positive and negative cases. With a precision of 0.93, this model minimizes false positives by being accurate 93% of the time when it predicts a case to be positive. Its high recall of 0.91 indicates that it can accurately identify 91% of real positive cases, which makes it very useful for heart disease detection and ensuring that the majority of true instances are not overlooked.

IQR + CST + RNN performs better than the other model. Although it performs well, the Z-Score + CST + RNN model displays somewhat lower metrics when compared to the IQR method. While not as strong as the IQR model, this model is still fairly good with 90.25% accuracy, 0.89 precision, and 0.87 recall. Its reduced accuracy indicates that a lesser percentage of cases are accurately classified. With a recall of 0.87, it is apparent that the model misses a few more positive examples than the IQR model, and the precision of 0.89 suggests a marginally higher proportion of false positives. It is nevertheless a powerful model with excellent performance in classifying cardiac diseases despite these small drawbacks.

Out of the three models, the Min Max + CST + RNN model has the lowest accuracy (89.42%, precision of 0.87, recall of 0.85). With its lowest accuracy, this model is thought to accurately classify the fewest cases overall. With a precision of 0.87, it predicts a positive case 87% of the time, which means that it produces more false positives than the other models. Its recall of 0.85, which is the lowest of the models and suggests that it misses more positive examples, also demonstrates that it recognizes 85% of the real positive cases. As such, compared to the IQR and Z-Score models, the Min Max model is less useful for accurately detecting cardiac disease, even if it is still operational.

The optimal option for classifying heart illness is the IQR + CST + RNN model since it offers the best balance of recall, accuracy, and precision. While the Min Max + CST + RNN model is effective, it performs comparatively worse across all metrics than the Z-Score + CST + RNN model, which performs well but with slightly lower metrics.

5. CONCLUSION AND FUTURE WORK

This work shows that heart disease prediction can be greatly improved by combining RNN with sophisticated preprocessing methods. We standardized the data by using z-score normalization and min-max scaling, which enhanced the RNN's training procedure. Better model performance resulted from the identification of the most pertinent characteristics through feature extraction using IQR and Chi-square testing. Improvements in accuracy, precision, and recall are demonstrated by the results, which are assessed using the UCI heart disease dataset. This suggests that the integrated strategy provides a strong remedy for early heart disease diagnosis. This improved performance demonstrates how well-sophisticated preprocessing and feature extraction methods work in conjunction with RNNs to provide medical predictive modeling with a reliable method for diagnosing heart disease.

Future studies could look into several ways to improve forecast accuracy. Deeper insights and increased model accuracy may result from incorporating additional feature extraction approaches, such as Principal Component Analysis (PCA) or sophisticated statistical methods. Increasing the dataset's diversity and sample size could aid in the model's ability to be applied to a variety of demographics. Further advancements might also be possible by looking into hybrid models and further ML methods. Lastly, combining these models with real-time data processing and implementing them in clinical settings may provide useful applications and proactive healthcare management.

REFERENCES

[1] C. Boukhatem, H. Y. Youssef & A. B. Nassif, "Heart Disease Prediction Using Machine Learning" *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ASET53988.2022.9734880.

[2] S. Kale, A. Shimpukade, A. M. Yadav &S. Dhavane, "Predictive Analytics for Heart Diseases Using Machine Learning," *2024 5th International Conference for Emerging Technology (INCET)*, Belgaum, India, 2024, pp. 1-5, doi: 10.1109/INCET61516.2024.10593243.

[3] Devansh Shah, Samir Patel & Santosh Kumar Bharti (2020), “Heart Disease Prediction using Machine Learning Techniques”, SN COMPUT. SCI. 1, 345 (2020). https://doi.org/10.1007/s42979-020-00365-y

[4] Chintan M. Bhatt, Parth Patel, TarangGhetia& Pier Luigi Mazzeo(2023), “Effective Heart Disease Prediction Using Machine Learning Techniques“, Algorithms, Vol. 16, No. 210.3390/a16020088

[5] Mauro Chiarito, Luca Luceri, Angelo Oliva, GiulioStefanini& GianluigiCondorelli(2022), “Artificial Intelligence and Cardiovascular Risk Prediction: All That Glitters is not Gold”, EurCardiol. 2022 Feb; 17: e29, doi: 10.15420/ecr.2022.11

[6] UmaraniNagavelli, DebabrataSamanta& Partha Chakraborty(2022), “Machine Learning Technology-Based Heart Disease Detection Models“, J HealthcEng. 7351061, doi: 10.1155/2022/7351061

[7] Tao R., Zhang S & Huang X(2019), “Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods. *IEEE Transactions on Biomedical Engineering.* 2019;66(6):1658–1667. doi: 10.1109/tbme.2018.2877649.

[8] Spencer R., Thabtah F & Abdelhamid N(2020) Thompson M. Exploring feature selection and classification methods for predicting heart disease. *Digital Health.* 2020;6 doi: 10.1177/2055207620914777.2055207620914777

[9] Ch. Anwar ul Hassan, Jawaid Iqbal, Rizwana Irfan, Saddam Hussain, Abeer D. Algarni, Syed SabirHussain Bukhari, Nazik Alturki& Syed SajidUllah(2022), “Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers”, Sensors (Basel). 2022 Oct; 22(19): 7227. doi: 10.3390/s22197227

[10] B Bavani, S NirmalaSugirthaRajini, MS Josephine&V Prasannakumari(2019), “Heart disease prediction system based on decision tree classifier”, Jour of Adv Research in Dynamical & Control Systems”, Vol. 11, No. 10, pp. 1232-1237

[11] https://archive.ics.uci.edu/dataset/45/heart+disease