International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

Statistical Analysis with Machine and Neural Learning-Based Model on Cardiovascular Diseases and Stroke Prediction

Lakkala Jayasree[1], Dr. D. Usha[2]
[1]Research Scholar, Department of C.S.E., Dr. M.G.R Educational & Research Institute, Chennai, India.
Email: jayasreemohan15@gmail.com
[2]Associate Professor, Department of C.S.E, Dr. M.G.R Educational & Research Institute, Chennai, India.
Email: ushahits@gmail.com

Abstract—Several risk factors, such as hypertension, hyperlipidemia, and an irregular heart rhythm, make an early diagnosis of cardiovascular disease challenging. Reducing cardiac risk calls for precise diagnosis and therapy. Clinical practice in the healthcare business is likely to evolve in tandem as a result of advancements in machine learning. Therefore, scientists and doctors need to acknowledge machine learning's significance. The fundamental purpose of this research is to a reliable analyzing Risk Factors for Cardiovascular Disease method that makes use of machine learning. Classifying well-known cardiovascular datasets But, on the other hand, is a job for state-of-the-art machine learning techniques and neural network algorithms. Several statistical and visualization indicators were used to assess the efficacy of the suggested approaches and to determine the optimal machine-learning and neural-network approach. Using these modeling methods acquired high and accurate accuracy on stroke and heart disease prediction.

Keywords-Cardiovascular Health Analytics; Statistical Analysis; Machine learning; Prediction

I. INTRODUCTION

Cancer Heart disease is now the primary reason for death worldwide. Predicting cardiovascular illness from large quantities of healthcare data is a major difficulty in the field of medical data analysis, but recent advances in machine learning have shown promising results [1]. Numerous risk factors, including hypertension, excessive cholesterol levels, and an irregular heart rate, make the cardiovascular disease difficult to diagnose. Due to the complexity of the disease, special care must be used when treating it. Heart problems or even death could result from not doing so. Medical science has benefited from technological progress through the use of computerized decision-support and prediction systems [2]. Machine learning methods have proven effective in the healthcare business, allowing for faster and more precise disease prediction. Lifesaving early diagnosis is especially important in the fight against cardiovascular disease [3]. Prevention measures against these illnesses are equally crucial. To aid in early diagnosis, many healthcare providers employ data analytics technologies. Worldwide, cardiovascular disease was a leading in 2015 Deaths attributable, responsible for the deaths of an estimated 17.7 million individuals. Optimal therapy and correct decision-making are necessary to reduce cardiac risk. Five machine-learning models were utilized in another Canadian investigation of Patients having a high risk of dying while in the hospital congestive cardiac arrest. Predictions for patients with myocardial infarction made within hospitals have been the subject of research in both China and South Korea. However, research has shown that heart disease is responsible for one in four deaths in America[4]. Roughly 92,1 million adult Americans have cardiovascular disease. The advancements made in machine learning have helped medical professionals in their fields. Consequently, a cardiovascular risk prediction system needs to be precise and reliable. Developments in machine learning can potentially revolutionize clinical practice in the healthcare sector. Therefore, scientists and medical professionals must understand the value of machine learning methods. Although algorithms for predicting risks do exist, they often only account for some of the potential hazards. It is still difficult to improve the precision of risk prediction systems when dealing with complicated interactions. When coronary heart disease is present, the heart is unable to pump enough blood to maintain normal bodily functions[5]. Symptoms include difficulty breathing, weakness, swelling feet, weariness, and exhaustion. As the healthcare industry adapts to new consumer expectations, it generates a large volume of data. Medical histories typically include details about cardiovascular disease, including symptoms and lifestyle factors. Checking your auscultation, BP, and lipids, electrocardiogram (ECG), and blood sugar are just some of the tests that may be performed before a diagnosis is made. The results of these examinations help decide whether or not a patient needs medicine. Human

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

expertise has its limits, therefore inaccurate diagnoses are possible in the healthcare system[6].

There is a higher chance of cardiac arrest in the currently suspended life scenario. People with chest pain who are afraid to go to the doctor for fear of contracting a communicable disease end up in worse shape. Accurate prognoses are fundamental for effective medical care. Successful decision assistance systems are the focus of ongoing research. There is still difficulty in making a diagnosis of cardiac disease. Classification methods are essential to the process of prediction. The fundamental objective of this research is to suggest a system for predicting cardiovascular illness using machine learning. To this end, the study classifies the most popular cardiovascular datasets using state-of-the-art machine learning algorithms like Naïve-Bayes, Random Forest, support vector machines, and neural networks. Accordingly, The right machine learning algorithm should be selected based on the performance of the chosen classification approach in situations of cardiovascular disease [3].

II. RESEARCH MOTIVATION AND LITERATURE

A. Research Objectives

(i) In this investigation, we compare efficiency of several machine learning methods for predicting cardiovascular risk.
(ii) Statistical Analysis carried out on the dataset
(iii) Most popular and Starlog cardiovascular illness datasets are used in minimal's attribute examination of machine learning classification methods (heart).
(iv) A novel aspect of this study is the comparison of the effectiveness of the latest Random Trees - machine learning algorithms in the context of cardiovascular disease prediction.
(v) Consequently, a reliable method of predicting the onset of cardiovascular disease is made available. In addition, we advise on the most appropriate machine learning method to use when developing advanced AI systems for predicting cardiovascular diseases.

B. Key Findings from Literature

The total accuracy of machine learning techniques for detecting cardiovascular disease was assessed. The plan was developed utilizing several March 2019-released databases. heart illness, coronary arrhythmias, and heart failure, and stroke were all disorders that could be predicted. Prediction analysis made use of the area under the curve measure[7]. It is still difficult to determine In the field of cardiovascular illness disease, the top machine-learning because of the wide variety of available options. Concentrations of heavy metals in bodily fluids were correlated with mortality from cardiovascular disease and cancer. Datasets collected as part of Use of data from the National Health and Nutrition Examination Survey for the analysis. Analyses of both single-metal and multi-metal exposure were conducted using Poisson's regression. The ages of the study's participants ranged from 25 to 85.

Variables such as age, sex, education level, medical comorbidities, body mass index, and serum cotinine were analyzed. Metal ions in urine and blood were found to be associated with a higher risk of dying from cancer. However, the authors emphasize that the desire to learn more about cardiovascular disease prompted them to conduct this study. They paid special attention to the risk of heart disease during the COVID-19 epidemic. The government has been forced to enforce various sorts of lockdowns due to the statewide quarantine to stop the spread of COVID-19. Because of these rules, everyone stays at home and doesn't get any exercise. Although the World Health Organization (WHO) has established clear standards on the quantity of physical exercise required to maintain optimum health, rigorous confinement has been shown to raise cardiovascular mortality risk. There are harmful health impacts from being quarantined. That's why the authors advocated keeping up with regular exercise even while quarantined in order to lessen the likelihood of heart disease. The design of the present investigation was consequently affected. One of their suggestions was to use machine learning algorithms to detect cardiovascular disease based on the microbiome. Both cardiovascular and noncardiovascular patients' fecal ribosomal RNA (16S) was examined. Subject samples were collected as part of the American Gut Project. Decision trees, random forests, neural networks, elastic networks, and support vector machines were among the five types of machine learning algorithms that were learned[8]. Bacterial taxa of varying kinds were distinguished. The improved attributes curve that was generated using random forest was 0.70 [9]. The current study incorporated one of the machine learning methods, and the random forest because of its potential in predicting cardiovascular disease. The threat of rapid advancement of coronary atherosclerosis was evaluated using a variety of machine learning methods. Plaque characteristics on CT angiograms were analyzed for 983 patients, both qualitatively and quantitatively. Cardiovascular atherosclerosis risk was compared to the model's score. We compared the most critical clinical parameters. However, the authors stress that it remains difficult to evaluate hidden biases in the dataset using machine learning methods. They looked into how well machine learning methods have been able to forecast cardiovascular disease risks consistently. The authors examined 3.6 million patients in England who had been admitted to hospitals after receiving an emergency medical referral. Overall, the 19 prediction models were tested for their discriminating and calibration abilities. The accuracy of predictions made using various methods varied widely; for instance, prediction score variation using random forest was 2.8 to 9.1 percent, while the range of the Quantitative results from a neural network's forecasting was 2.3 to 7.1 percent. It was recommended that

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

while comparing models, logistic models not be used for predicting long-term risks, and that model levels be checked frequently. Many data science issues can be tackled with the help of machine learning. Machine learning relies on preexisting data to make predictions. The authors looked into ensemble classification, a potent machine-learning technique, to enhance the performance of several classifiers. Prediction classification is enhanced by the ensemble classification, but only by 7%. The Cleveland heart dataset was utilized for both training and testing purposes [10]. Random forest combined with Models yielded 85.48% accuracy in predicting coronary disease. Data mining is the practice of collecting relevant information from a wide variety of sources. The field of healthcare is the most popular place to apply data mining. The study used the random forest method to foresee cases of heart disease in patients[8]. We looked at all 303 samples available in the Kaggle dataset. Accuracy, sensitivity, and specificity were the measures employed in the evaluation process. The heart disease classification system attained a 93.3% accuracy rate.

rating overall S iterations were calculated. To prevent over-fitting, we employ S-fold cross-validation to independently perform a model selection for each classifier and produce new sets of data for the ensemble stage.

III. METHODOLOGY

The use of machine learning in cardiovascular care is on the rise. While many different machine learning methods exist, it is still difficult to find the one that is most suited to and practical for cardiovascular disease datasets [11]. The fundamental purpose of their suggested study is to suggest using machine learning to forecast cardiovascular illness with a high degree of accuracy. The suggested cardiovascular disease prediction system (statistical and learning methods) architecture is shown in Figure 1. The framework takes in patient medical records and uses modern machine-learning algorithms to classify common cardiovascular datasets to generate reliable predictions for use in expert consultations. These algorithms include Random Forest, Naive Bayes, Support Vector Machines, and Neural Networks [12]. This allows us to establish which machine learning approaches produce the best results algorithm for handling cases of cardiovascular illness based on the results of our chosen classification method [13].

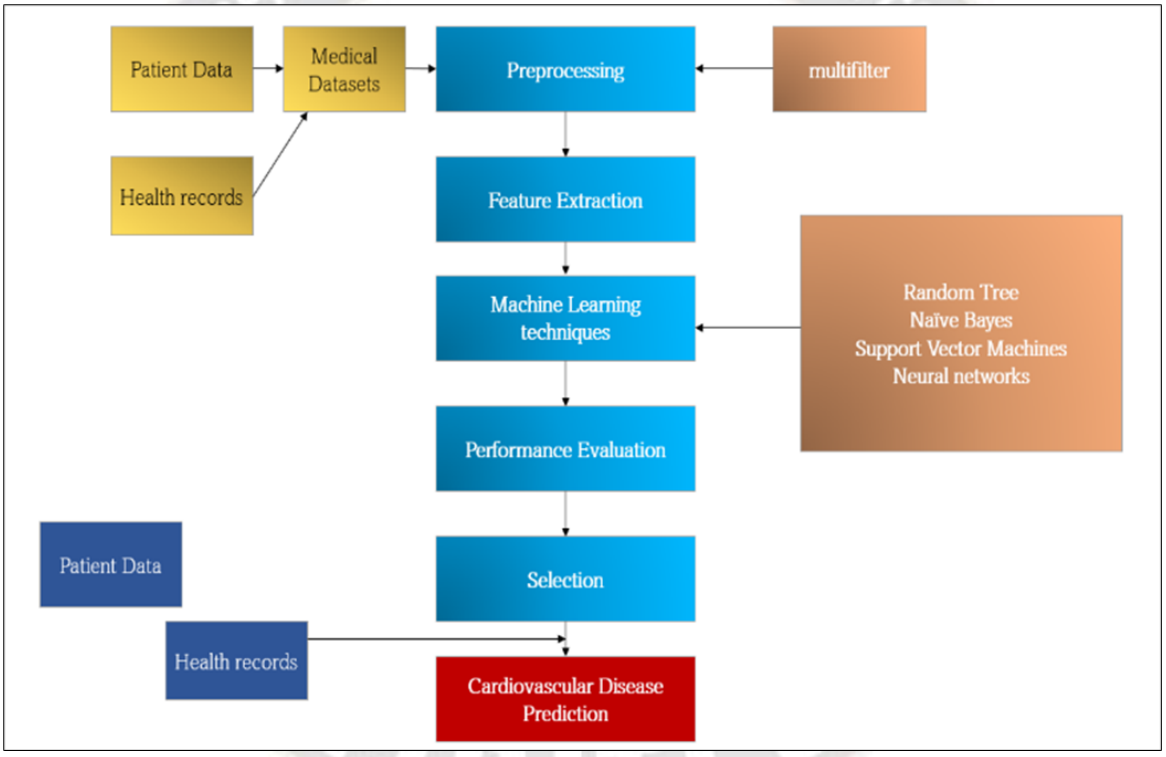


Figure 1: Process Flow Diagram

A. Data Visualization and Observations

The first step in data mining: missing and noisy values are common in real-world data. For this reason, the data is processed before use so that reliable forecasts may be made. We can't rely on the raw data because it's unreliable and incomplete. When there are missing values, it is possible to either exclude them or substitute a median value. Therefore, the collected data must be slightly adjusted using some filtering approach to conduct a successful analysis. In this case, we employ a multi-filtering strategy. From this data and visualization, Gender-based analysis is shown. Figure 2 shows that the Female category has higher heart diseases when compared to the Male.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

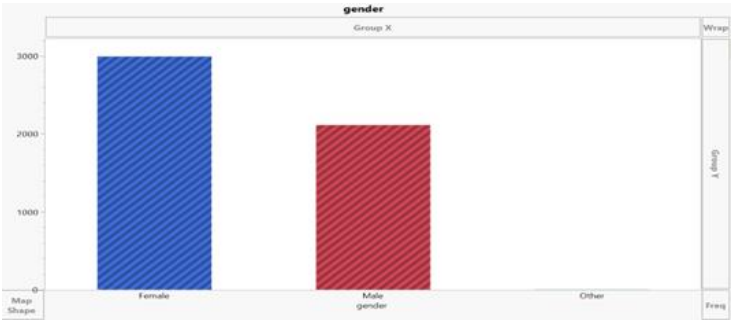


Figure 2: Data Visualized based on Gender Vs heart stoke relation

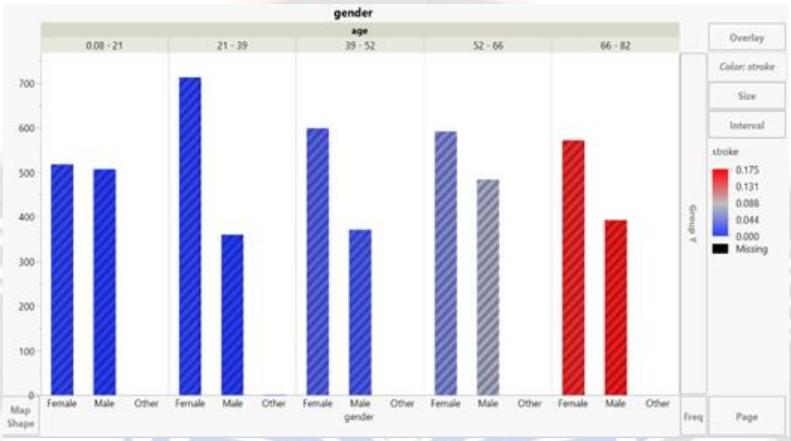


Figure 3: Age and Gender Vs Heart Stroke Relation

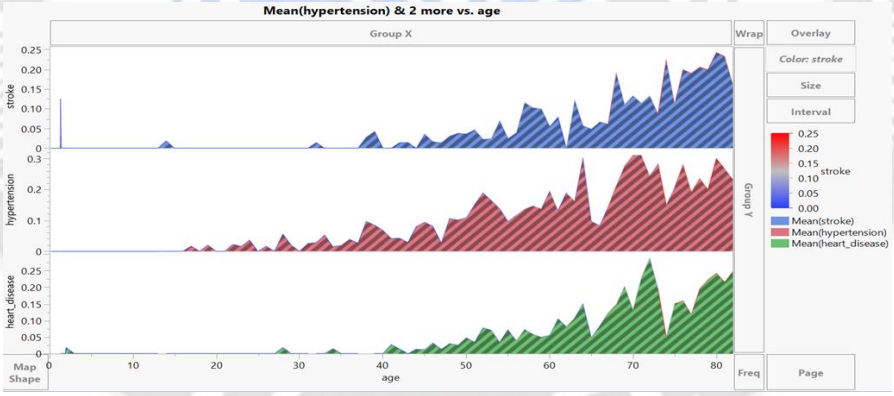


Figure 4: Factors influencing Heart Stroke depending on Age

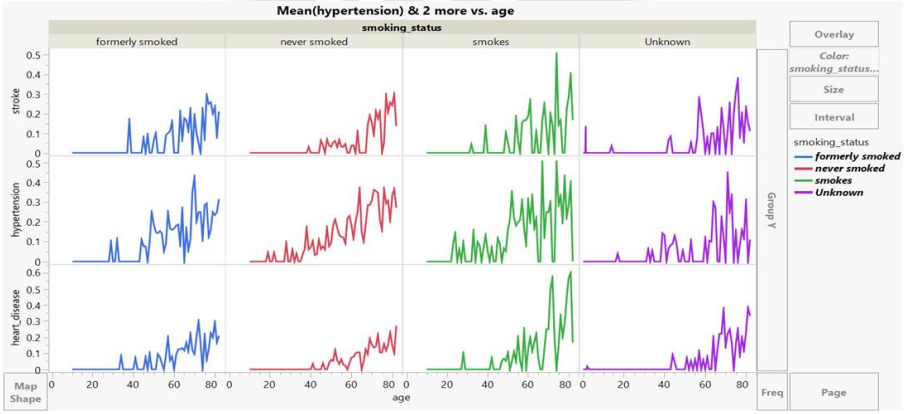


Figure 5: Factors influencing Heart Stroke depending on Age and Smoking Habits

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

Figure 3 depicts the Age factor and Gender-based analysis concerning heart stroke. Around 21-39 years of age group females are getting higher heart strokes.
Figures 4 and 5 show the factors that influence to get heart stroke concerning Age, Smoking habits, and other related factors. A gender-based analysis was also carried out.

B. Feature Extraction and Data Modeling

Cut down on input qualities before attempting data analysis. However, not all of them are created equal in terms of their predictive power. In addition to increasing complexity, a high number of attributes have negative performance. This necessitates painstaking feature extraction that doesn't compromise the efficiency of the system. Figure 6 shows the BMI Vs Gender as considered as an important feature.

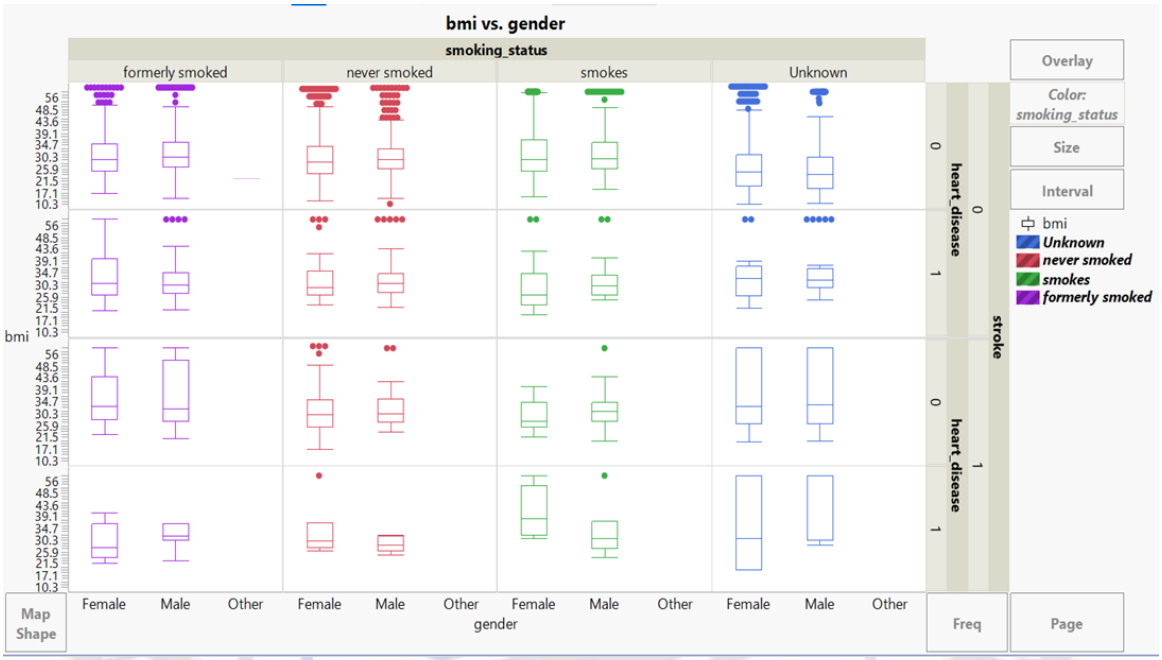


Figure 6. Factors influencing Heart stroke and Heart Diseases with BMI and Gender

C. Learning Methods – Machine Learning models and Neural Networks

Random trees Tree makes use of the logic of regression trees by producing numerous trees at various stages of the iteration process. As a Random Treesresentation of the trees formed, it picks the one with the most favorable characteristics. Cut down on the tree-like knowledge by calculating the mean square error. Random trees speed up the learning process by creating decision trees from the data collected. Therefore, Random trees offer a more straightforward categorization tree, or when working with massive volumes of data. Random Tree: The Random Tree model tree is implemented calculational forecasting. Class values of instances are predicted at each layer and stored in a Machine learning model [14]. Dividing the T portion data used for training yields the optimal characteristic. To get to a certain node, the dividing criterion is applied. Predicting importance of the values of quantitative An example of a response variable is the job of the RF model tree, a decision tree generated in two stages. The standard deviation figures are used first as the splitting criteria. The resulting quality is diminished by the error function of each value. The Machine learning model's parameter space informs the branching of the model tree. When evaluating a node's performance, the amount of mistakes introduced is employed as a metric.

Table 1: Difference: heart_disease-avg_glucose_level

heart_disease	0.05401	t-Ratio	-167.612
avg_glucose_level	106.148	DF	5109
Mean Difference	-106.09	Prob > |t|	<.0001*
Std Error	0.63297	Prob > t	1.0000
Upper 95%	-104.85	Prob < t	<.0001*
Lower 95%	-107.33		
N	5110		
Correlation	0.16186		

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

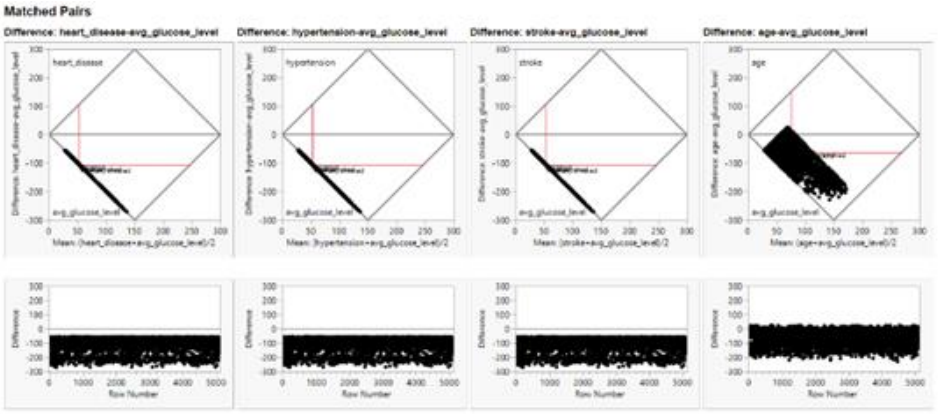


Figure 7. Attributes Paired information and Analysis

Table 2: Difference: heart_disease-avg_glucose_level - Analysis

smoking_status	Count	Mean Difference	Mean Mean
formerly smoked	885	-112.8	56.487
never smoked	1892	-107.5	53.803
smokes	789	-107.9	54.047
Unknown	1544	-99.57	49.816

Test Across Groups	F Ratio	Prob>F		
Mean Difference	18.4184	<.0001*	Within Pairs	Y Axis
Mean Mean	18.6538	<.0001*	Among Pairs	X Axis

Table 3: Difference: hypertension-avg_glucose_level

hypertension	0.09746	t-Ratio	-167.598
avg_glucose_level	106.148	DF	5109
Mean Difference	-106.05	Prob > |t|	<.0001*
Std Error	0.63277	Prob > t	1.0000
Upper 95%	-104.81	Prob < t	<.0001*
Lower 95%	-107.29		
N	5110		
Correlation	0.17447		

Table 4: Difference: hypertension-avg_glucose_level - Analysis

smoking_status	Count	Mean Difference	Mean Mean
formerly smoked	885	-112.8	56.511
never smoked	1892	-107.4	53.84
smokes	789	-107.9	54.068
Unknown	1544	-99.57	49.818

Test Across Groups	F Ratio	Prob>F		
Mean Difference	18.2568	<.0001*	Within Pairs	Y Axis
Mean Mean	18.8161	<.0001*	Among Pairs	X Axis

Table 5: Difference: stroke-avg_glucose_level

stroke	0.04873	t-Ratio	-167.59
avg_glucose_level	106.148	DF	5109
Mean Difference	-106.1	Prob > |t|	<.0001*
Std Error	0.63309	Prob > t	1.0000
Upper 95%	-104.86	Prob < t	<.0001*
Lower 95%	-107.34		
N	5110		
Correlation	0.13195		

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

Table 6: Difference: stroke-avg_glucose_level - Analysis

smoking_status	Count	Mean Difference	Mean Mean
formerly smoked	885	-112.8	56.483
never smoked	1892	-107.5	53.803
smokes	789	-108	54.035
Unknown	1544	-99.57	49.816

Test Across Groups	F Ratio	Prob>F		
Mean Difference	18.4369	<.0001*	Within Pairs	Y Axis
Mean Mean	18.6352	<.0001*	Among Pairs	X Axis

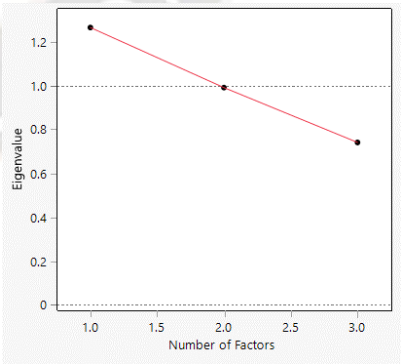
Tables 1-6 signify the matched pairs' information and analysis reports. When it comes to mathematical analysis and modeling understanding the errors in the standard notation plays an important role. where Ti is the branching point used to construct the model related to the objective. Standard deviation at the node is used to measure the error decrease as the splitting process is randomly Trees reacted recursively. Standard deviation reduction, sd, is used to quantify the attributes that contribute most to a system's ability to reduce errors (1). The quality of a forecast is measured by its accuracy metric. To a group of feature spaces Zi, where the model is stored, [z →∈ z1,. . .,zn] stretches from lower bound z →i ∈ inf[z → ∈ Zi] to upper bound z →i ∈ max[z → ∈ Zi]. After that is assembled. It makes use of the n-column matrix that contains Zj characteristics, in addition to y. An alternative notation for using the logarithmic form of the expression is B. According to the split technique insights, the data included in the offspring nodes is smaller than the standard deviation of that found in the parent node. After exploring all possible outcomes, NB and SVM make their selection based on the attribute with the largest influence, which often leads to an overfitting tree-like structure [15]. Overfitting can be fixed by cutting back on the tree's growth. Machine learning is a method for making predictions about the values of label attributes given a set of input attributes. Upon prediction of the model, tests have been carried out and shown in Figures 7 and 8.

Where is the observation of the desired property, and X is the predictive function? A 1 is assigned if the value is larger than the threshold, and a 0 otherwise. The Naive Bayes classifier is a straightforward method for applying Bayes' theorem to classification problems[16]. It takes for granted that qualities are uncorrelated with one another. Probability can be calculated with the use of the Bayes theorem, a mathematical notion. There is no interdependence or correlation between the predictors. The probability of maximization is a function of all of the qualities separately, as shown below in the equation. Using Naive Bayes methods, but not Bayesian ones, it can be useful. [17]. There are a variety of complicated real-world applications for Naive Bayes classifiers, in Table 9 and Figure 8 model fit shown.

The posterior probability, denoted as P(X/Y), is the product of the class prior probability, P(X), and the predictor prior probability, P(Y). It is possible to conduct classification and prediction using a Random Tree by averaging the results of multiple separate base models. The algorithm they created—the random forest—was rebranded as Random Trees to avoid trademark infringement. Therefore, it is a reliable technique for predicting missing data and keeping precision even when as much as 80% of the data is absent. Errors in datasets with an uneven class population can be reduced using the strategy.. A rule learner is built into this class. Rule learning in this approach is performed in a bottom-up fashion using the Random Treeseated Incremental Pruning to Reduce Errors (RIPPER) technique. It provides numerous alternatives for making a decision tree, both with and without pruning. To guarantee that the data was categorised as precisely as feasible on the training data, the fundamental method classifies iteratively until each leaf is pure.

Table 9: Naïve Bayes Fit model details

Measure	Training	Definition
Entropy RSquare	-0.179	1-Loglike(model)/Loglike(0)
Generalized RSquare	-0.371	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.8017	∑ -Log(ρ[j])/n
RASE	0.5085	√∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4598	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.4127	∑ (ρ[j]≠ρMax)/n
N	5110	n



Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	1.2665	42.217		42.217
2	0.9920	33.068		75.285
3	0.7415	24.715		100.000

Figure 9. EigenValues on Factors Analysis – Variables.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

D. Evaluation Metrics

Accuracy was evaluated alongside other metrics including MAE and RMSE. The reliability of continuous variables can be determined using both the mean absolute error and root mean squared error. The mean absolute error (MAE) measures how consistently off a group of predictions is. Root-mean-squared error (RMSE) quantifies the typical size of errors. It can be calculated as the following equation's square root. In the context of predicting future outcomes, RAE is a straightforward metric that simply averages the absolute difference between the actual value and the predicted value. For each of the factors under consideration, the calculation of the answer variable by means of a prediction equation, with Pij standing in for the predictor for the model I with j records. For a given set of j records, Tj Random trees represent the desired value and T as described.

IV. RESULTS AND DISCUSSIONS

Examples of the cardiac disease include coronary artery disease, arrhythmias, and various congenital heart problems. Heart attack, angina, Heart disease, angina, and stroke are all indicators of cardiovascular disease. characterized by the narrowing of blood arteries as a result of plaque buildup. Since cardiovascular disease is now one of the leading causes of death, its prediction is a major focus of clinical data analysis. Timeseries analysis to apply machine learning methods to aid professionals in making judgments and forecasts.

This essay draws from two widely-used databases—the and Heart datasets. There are 294 records in the Hungary database, which was compiled by the Institute of Cardiology in Budapest. The heart dataset has 304 unique samples. There are 76 attributes in this database, however, only 28 have been used in any of the existing publications. The many features of cardiovascular disease. There are two types of assessments in this study. Machine learning classification strategies like Machine learning and Neural Networks were first applied to the Heart dataset. Methods of machine learning classification similar to Random Trees, Nave Bayes, Support Vector Machines, and Neural networks were used in the dataset as well[18]. Accuracy was evaluated alongside other metrics including MAE and RMSE. The Random Trees Tree and the Random Tree were also compared in depth. The results of an examination of machine learning strategies applied to the database. The MAE-based performance of machine learning models on the database. The mean absolute error (MAE) for the Random Forest and machine learning, and Random Trees are 0.418, 0.3763, 0.3978, and 0.3838, respectively. The mean absolute error (MAE) is the most reliable measure of a model's ability to make accurate predictions. With an MAE of 0.3763, ranks lowest among the methods tested. Lower MAEs are preferred for optimum cardiovascular disease prediction because of the positive correlation between accuracy and MAE. This frees up medical professionals to focus on how to best apply the suggested machine learning model to the examination of clinical data related to cardiovascular disease. It is essential to realize that Decisions and forecasts made by the system using Random Tree will be accurate achieving a score of 0.3838 and Random Tree achieving a value of 0.3838. Too much emphasis on the mean will lead to a mistake. To take into account extreme outliers, one must compute the root mean square error (RMSE). With the root-mean-square error (RMSE) as a metric, machine learning models predict in the database. Variational mean squared error (RMSE) values are 0.5415 for the Random Trees Tree, 0.4769 for the, 0.471 for the Machine learning, and 0.6328 for the Random Tree. The root-mean-squared error (RMSE) is the most accurate measure of a model's predictive power, hence minimizing it is the objective here. Results show it has the best RMSE performance at 0.4769. Relative standard error (RMSE) values closer to zero are preferred for predicting cardiovascular disease. When compared to algorithms, however, the other models show that they are just as capable of producing accurate forecasts and judgments within the context of the proposed system. The accuracy with which database-based prediction models generated predictions using machine learning is shown with accuracy. An accuracy of 89.44% was found using Random Forest, and 91% using Neural Networks. The target is a more precise prognosis of cardiovascular disease. Given these findings, a tuned neural network is the method of choice for best cardiovascular disease prediction. Its accuracy is 99.91%. This frees up medical professionals to focus on how to best apply the suggested machine learning model to the examination of clinical data related to cardiovascular disease. Random Forest and machine learning, and Random Tree each take 0.05 seconds, 0.53 seconds, 0.02 seconds, and 0.03 seconds to make their predictions, respectively. Here, we hope to improve the speed and precision with which cardiovascular disease may be predicted. Results showed that Machine learning and Random Trees used 0.02 seconds and 0.03 seconds, respectively, to make predictions. Thus, these two models are strongly suggested for the most accurate cardiovascular disease forecasting.

A. Results With ML and NN model on Heart Stroke based Databases

Table 3 displays the results of an examination of machine learning strategies applied to the Heart database. Table 3 shows the estimation results from machine learning algorithms on the Heart database, as measured by mean absolute error, root means squared error, accuracy, and time. The MAE values for Naive Bayes, Random Tree, and JRIP are 0.0012, 0.0012, 0.0021, and 0.0015, respectively. The RMSE values for Naive Bayes, Random Tree, and JRIP are 0.0232, 0.0232, 0.0232, and 0.0328, respectively. Naive Bayes and random trees both have an

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

accuracy of%. With the models, 99.91% accuracy was recorded [19]. However, a Random Tree yields optimal results in the quickest possible time. Prediction accuracy as measured by mean absolute error (MAE) for machine learning models in the Heart database is shown in Figure 8. When compared to Naive Bayes, Random Tree, and JRIP, the MAE values achieved by these methods are 0.0012, 0.0012, 0.0012, and 0.0015. The mean absolute error (MAE) is the most effective metric for judging a model's predictive abilities, and minimizing the prediction error is the goal here. The results showed that the MAE was lowest for the Naive Bayes, and Random Tree techniques, at 0.0012. A lower MAE is preferred for optimal cardiovascular disease prediction since it indicates more accuracy. This frees up medical professionals to focus on implementing the proposed machine learning models to enhance clinical data analysis for cardiovascular disease. Too much emphasis on the mean will lead to a mistake. To take into account extreme outliers, one must compute the root mean square error (RMSE). RMSE-based the accuracy with which machine learning models make predictions using the Heart database, Obtaining RMSE values of 0.0232, 0.0232, 0.0232, and 0.0328 for Naive Bayes, Random Tree, and JRIP models, respectively. Minimizing the root-mean-squared error (RMSE) of the predictions made by the model is the primary goal. The results showed that the RMSE was 0.0232% at its lowest with the Naive Bayes and Random Tree methods. Relative standard error (RMSE) values closer to zero are preferred for predicting cardiovascular disease. The predictive efficacy of machine learning models on the Heart database is depicted in Figure 10. Accuracy values of %, NB (88.13%), 91% (Random Tree), and SVM (87.99%) were obtained. The major goal is to enhance the precision with which cardiovascular disease may be predicted. Results show that Naive Bayes and Random Tree have the highest accuracy (91%) for predicting cardiovascular illness.

B. Results With Timeseries Analysis and Modeling

Prediction time as a metric for evaluating machine learning model performance in the Heart database frees medical experts to focus on ways in which the proposed machine learning model can be used to enhance the processing of clinical data from cardiovascular studies illness [20]. Naive Bayes takes 0.02 seconds, SVM 0.16 seconds, Random Tree takes 0.02 seconds, and NB takes 3.26 seconds to make a prediction[21]. Improved and more efficient methods of Analyzing Risk Factors for Cardiovascular Disease are the focus of this research. According to the data, both the Naive Bayes and Random Tree prediction algorithms required a time investment of 0.01 or one-hundredth of a second. Thus, these two models are strongly suggested for the most accurate cardiovascular disease forecasting. Contrasting Random Trees Tree with Random Tree Predictions. Random Trees Tree and Random Tree, respectively, are generated from the Heart database. Decision tree outputs are often calculated with a completely arbitrary set of features. While Random Forest uses a combination of decision tree outputs to produce a conclusion, Random Trees, Tree constructs a decision tree for a specific dataset [22]. It took 0.03 seconds to construct the 21-node Random Trees Tree. It just took 0.02 seconds to construct a Random Tree with a depth of 141. That's why, when it comes to predicting complicated diseases like cardiovascular disease, the Random Tree is superior to the Random Trees Tree in terms of both depth analysis speed and accuracy. As shown in the Random Tree's performance was tested against the Heart and databases. When it comes to predicting cardiovascular disease, Random Tree excels, with a prediction accuracy of 100%, a mean absolute error of just 0.0012, a root-mean-squared error of just 0.0232, and a forecast time of just 0.01 seconds (secs). Thus, it is strongly suggested that a Random Tree be used for the most accurate prediction of cardiovascular illness. In addition, doctors can focus on perfecting their analysis of patient records related to cardiovascular disease using the proposed machine learning model.

V. CONCLUSION

Since cardiovascular disease is one of the leading causes of death, its prevalence in medical Considerable weight is placed on the results of thorough data analysis. Machine learning has the potential to improve clinicians' insights, especially in the prediction of cardiac disease, which may then be used to better adjust diagnosis and treatment to individual patients.. Several machine learning methods are examined for their practicability and potential use. The proposed statistical and learning methods use machine learning to aid specialists in generating sound judgments and forecasts. This work provides two datasets heart diseases and —that may be used with several machine learning classification algorithms, including Random tree, Random Trees Machine learning, Naive Bayes, and neural networks Different criteria were used to assess the proposed Statistical and learning methods' performance and determine the most appropriate machine-learning model for it. The Random Tree model had the best results for predicting individuals with cardiovascular illness, with an accuracy of 100%, an MAE of 0.0012, an RMSE of 0.0232, and a prediction time of 0.02. (secs). Improved performance in the classification of various forms of medical data could be the outcome of further work on the existing statistical and learning methods model, making it a more cost-effective and time-saving choice for patients and clinicians alike. Experiments can also be done to assess high-dimensional data for use in future investigations.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

REFERENCES

[1]. J. E. Naschitz, G. Slobodin, R. J. Lewis, E. Zuckerman, and D. Yeshurun, "Heart diseases affecting the liver and liver diseases affecting the heart," *Am. Heart J.*, vol. 140, no. 1, pp. 111–120, 2000.

[2]. K. V. R. Kumar and S. Elias, "Real-Time Tracking of Human Neck Postures and Movements," in *Healthcare*, 2021, vol. 9, no. 12, p. 1755.

[3]. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020.

[4]. V. T. Nkomo, J. M. Gardin, T. N. Skelton, J. S. Gottdiener, C. G. Scott, and M. Enriquez-Sarano, "Burden of valvular heart diseases: a population-based study," *Lancet*, vol. 368, no. 9540, pp. 1005–1011, 2006.

[5]. K. V. R. Kumar and S. Elias, "Smart Neck-Band for Rehabilitation of Musculoskeletal Disorders," 2020.

[6]. V. Chaurasia and S. Pal, "Data mining approach to detect heart diseases," *Int. J. Adv. Comput. Sci. Inf. Technol. Vol*, vol. 2, pp. 56–66, 2014.

[7]. K. V. R. Kumar and S. Elias, "Predictive Analysis for Detection of Human Neck Postures using a robust integration of kinetics and kinematics," arxiv, 2020.

[8]. K. U. Rani, "Analysis of heart diseases dataset using neural network approach," *arXiv Prepr. arXiv1110.2626*, 2011.

[9]. K. V. R. Kumar, B. R. Devi, M. Sudhakara, G. Keerthi, and K. R. Madhavi, "AI-Based Mental Fatigue Recognition and Responsive Recommendation System," in *Intelligent Computing and Applications*, Springer, 2023, pp. 303–314.

[10]. T. L. V Ulbricht and D. A. T. Southgate, "Coronary heart disease: seven dietary factors," *Lancet*, vol. 338, no. 8773, pp. 985–992, 1991.

[11]. P. M. Kumar and U. D. Gandhi, "A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases," *Comput. \& Electr. Eng.*, vol. 65, pp. 222–235, 2018.

[12]. R. Das, I. Turkoglu, and A. Sengur, "Diagnosis of valvular heart disease through neural networks ensembles," *Comput. Methods Programs Biomed.*, vol. 93, no. 2, pp. 185–191, 2009.

[13]. P. Ghosh, S. Azam, A. Karim, M. Jonkman, and M. D. Z. Hasan, "Use of efficient machine learning techniques in the identification of patients with heart diseases," in *2021 the 5th International Conference on Information System and Data Mining*, 2021, pp. 14–20.

[14]. R. Kannan and V. Vasanthi, "Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease," in *Soft computing and medical bioinformatics*, Springer, 2019, pp. 63–72.

[15]. M. Gudadhe, K. Wankhade, and S. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network," in *2010 International Conference on Computer and Communication Technology (ICCCT)*, 2010, pp. 741–745.

[16]. A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, "Design and implementing heart disease prediction using naives Bayesian," in *2019 3rd International conference on trends in electronics and informatics (ICOEI)*, 2019, pp. 292–297.

[17]. E. M. K. Reddy, A. Gurrala, V. B. Hasitha, and K. V. R. Kumar, "Introduction to Naive Bayes and a Review on Its Subtypes with Applications," *Bayesian Reason. Gaussian Process. Mach. Learn. Appl.*, pp. 1–14, 2022.

[18]. N.-S. Tomov and S. Tomov, "On deep neural networks for detecting heart disease," *arXiv Prepr. arXiv1808.07168*, 2018.

[19]. N. Priyanka and P. R. Kumar, "Usage of data mining techniques in predicting the heart diseases—Na{\"\i}ve Bayes \& decision tree," in *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, 2017, pp. 1–7.

[20]. A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, pp. 180235–180243, 2019.

[21]. C. B. Gokulnath and S. P. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease," *Cluster Comput.*, vol. 22, no. 6, pp. 14777–14787, 2019.

[22]. Y. K. Singh, N. Sinha, and S. K. Singh, "Heart disease prediction system using random forest," in *International Conference on Advances in Computing and Data Sciences*, 2016, pp. 613–623.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

Table 7: Wilcoxon Signed Rank

	heart_disease-avg_glucose_level	hypertension-avg_glucose_level	hypertension-heart_disease	stroke-avg_glucose_level	stroke-heart_disease	stroke-hypertension	age-avg_glucose_level	age-heart_disease	age-hypertension	age-stroke
Test Statistics	-6.50E+06	-6.50E+06	531413	-6.50E+06	-66083	-597974	-6.50E+06	6529303	6529303	6529303
Prob>|S|	<.0001*	<.0001*	<.0001*	<.0001*	0.1934	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*
Prob>S	1	1	<.0001*	1	0.9033	1	1	<.0001*	<.0001*	<.0001*
Prob<S	<.0001*	<.0001*	1	<.0001*	0.0967	<.0001*	<.0001*	1	1	1

Table 8: Sign Test

	heart_disease-avg_glucose_level	hypertension-avg_glucose_level	hypertension-heart_disease	stroke-avg_glucose_level	stroke-heart_disease	stroke-hypertension	age-avg_glucose_level	age-heart_disease	age-hypertension	age-stroke
Test Statistic M	-2555	-2555	111	-2555	-13.5	-124.5	-2327	2555	2555	2555
Prob ≥ |M|	<.0001*	<.0001*	<.0001*	<.0001*	0.2104	<.0001*	.	<.0001*	<.0001*	<.0001*
Prob ≥ M	1	1	<.0001*	1	0.9113	1	.	<.0001*	<.0001*	<.0001*
Prob ≤ M	<.0001*	<.0001*	1	<.0001*	0.1052	<.0001*	.	1	1	1

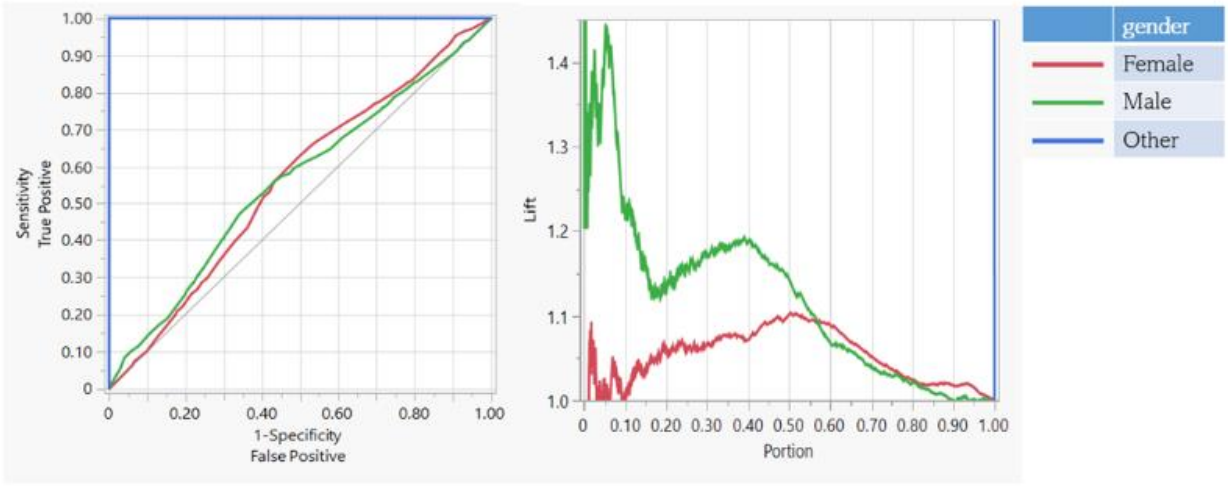


Figure 8. ROC and Lift curve details about Gender on Heart stroke

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5919
***Article Received**: 15 October 2022 **Revised**: 18 November 2022 **Accepted**: 24 December 2022*

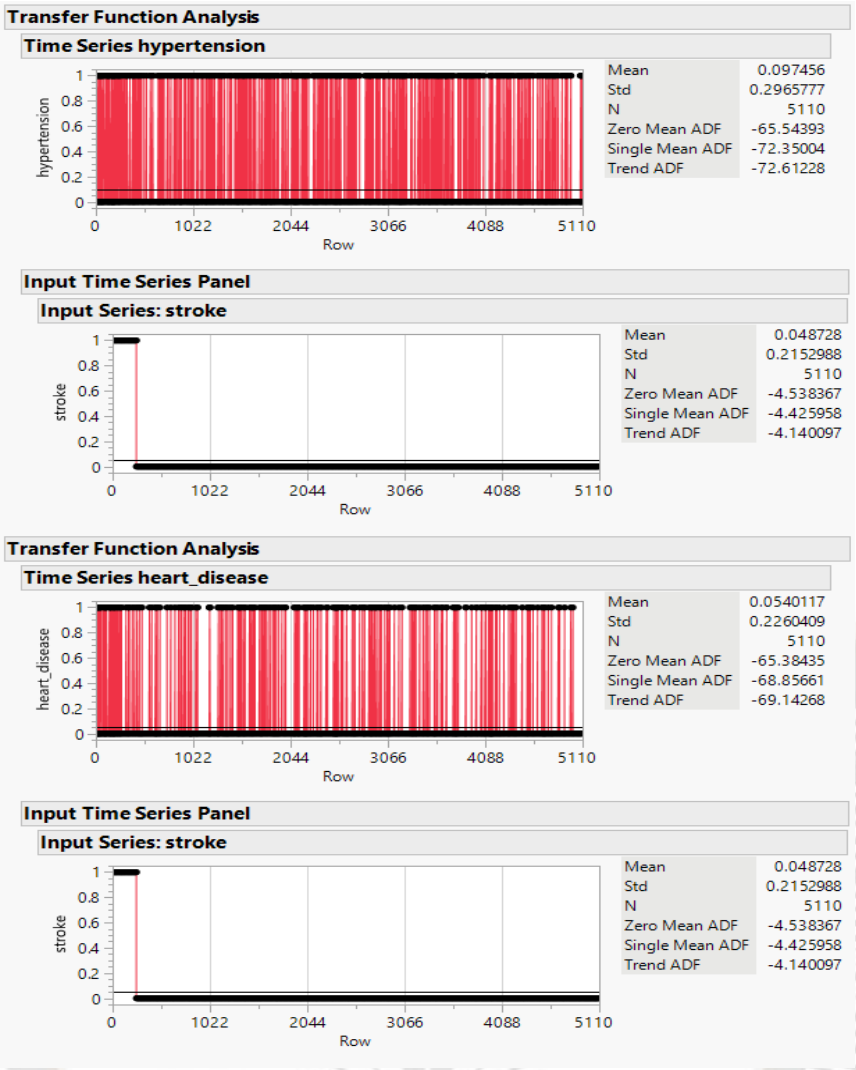
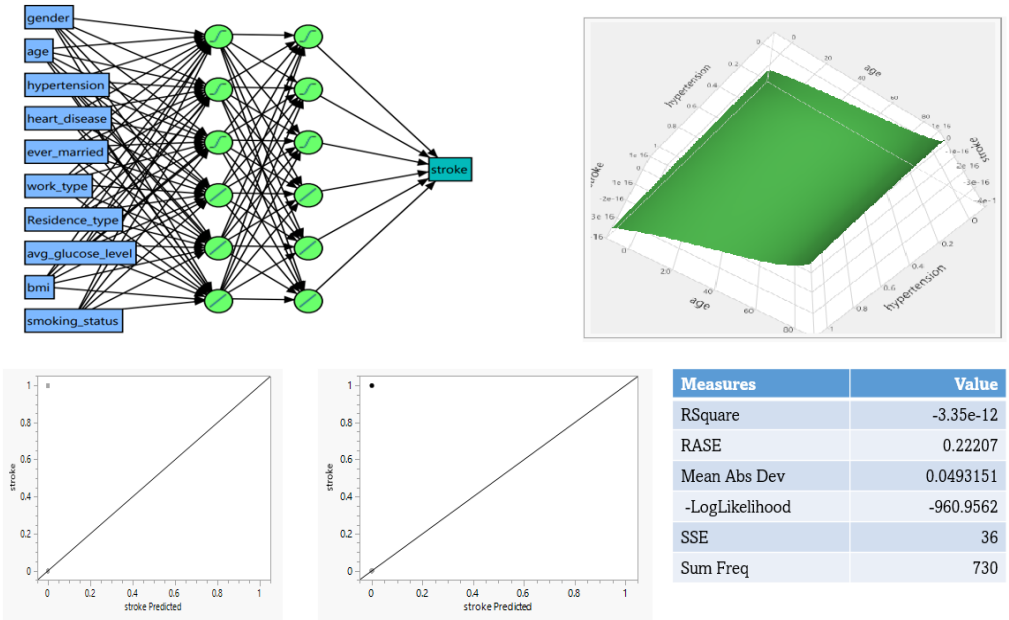


Figure 10. Timeseries Modeling.



Measures	Value
RSquare	-3.35e-12
RASE	0.22207
Mean Abs Dev	0.0493151
-LogLikelihood	-960.9562
SSE	36
Sum Freq	730

Figure 11. Neural Networks – Analysis.