Heart Stroke Prediction

Project report submitted

In partial fulfillment of the requirement for the degree of

Bachelor of Science

In

School of Engineering and Technology

By

Rohit Kumar

Nitin Kumar

Under the guidance of

Dr. Meenu Vijarania

K. R. Mangalam University, Gurugram - 122003

DECLARATION

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be for disciplinary action by the Institute and canal so evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed. We further declare that if any violation of the intellectual property right or copyright, my supervisor and university should not be held responsible for the same.

Student Name	(Roll No.)	(Signature)
Rohit Kumar	**2001840008**	
Nitin Kumar	**2001840003**	

Place: K.R. Mangalam University

ACKNOWLEDGEMENT

“Enthusiasm is the feet of all progress, with it there is accomplishment and Without it there are only slit alibis.”

Acknowledgment is not a ritual but is certainly an important thing for the successful completion of the project. At the time when we were made to know about the project, it was really tough to proceed further as we were to develop the same on a platform, which was new to us. More so, the coding part seemed so tricky that it seemed to be impossible for us to complete the work within the given duration.

We really feel indebted in acknowledging the organizational support and encouragement received from the university.

The task of developing this system would not have been possible without the constant help of our faculty members and friends. We take this opportunity to express our profound sense of gratitude and respect to those who helped us throughout the duration of this project.

We express our gratitude to our supervisors Dr. Meenu Vijarania for giving their valuable time and guidance to us.

Place: - K.R. Mangalam University

Name of Student
Rohit Kumar
Nitin Kumar

ABSTRACT

In today's generation many people face heart related diseases and lose their life due to heart attack, stroke, and cardiac arrest symptoms. Heart strokes have become a challenging symptom as patients start to face issues like speech difficulty, face drooping, arm weakness which shows the patient is suffering or going to suffer from heart related issues. In most of the cases patients face heart attacks involving the aches and pain around the center of chest areas, which results in uneven pressure, squeezing effect and aches which result in high pains. To stop and decrease the deaths due to heart strokes the new technologies could help to predict earlier about the symptoms, which are responsible for heart related diseases and reduce the number of deaths. By the use of prediction models involved in machine learning, data science and artificial intelligence the use of heart diseases can be easily resolved. The paper also shows the comparison of different algorithms and techniques to differentiate from old traditional methodologies and find the best methodology which gives the best result. The use of the technological advancement prediction model survival rate of heart disease patients can be also found and it helps to save many lives. In the case of prediction models there is also a need to care for parameters like Age, gender, bmi, medical history, drinking and smoking habits, level of glucose, type of work the person does etc. which will help to find the root cause for deaths in heart related diseases. By the use of prediction models, algorithms and methodologies techniques like logistic regression, random forest, naive Bayes, XG boost etc. helps to make comparative analysis and keep the record of accuracy models. The use of different models and the records maintained will help to establish the ideology of clinical setting which will be used for finding the probability of heart strokes. By the use of such techniques yearly records can be easily maintained and help to do risk analysis of the patients who have more symptoms and chances of death. By the use of modern technologies many lives can be saved, and also new technologies can be discovered to reduce heart diseases and useful in pre-diagnosis techniques.

Agreeing to American coronary heart affiliation, one man or woman passes on every 36 s from coronary heart contamination in the United states. The passing price of coronary heart stroke is nearly 18.1% out of every 3000 patients. The maximum goal is to advise a gadget studying-primarily based totally method to assume the coronary heart stroke of nice precision from evaluating an administered class of gadget studying calculations. We have furthermore computed the recipient hopeful bend and variety under the bends for every classifier. Furthermore, to evaluate and look at the execution of the exclusive gadget studying calculations from the given demonstration with an assessment report.

Keywords - Heart stroke, Pre-diagnosis, Prediction model, machine learning, risk analysis.

TABLE OF CONTENTS

Chapter 1: INTRODUCTION

According to a survey over 3 million people lose their life due to heart stroke and heart attack per year. Heart stroke is found to be a dominant reason that loses their life due to heart related diseases. To overcome the issues of death due to heart stroke, predictive modeling approach can be used to collect the data with multiple methods that find the root causes of heart stroke which will help to reduce the deaths due to heart stroke. By using the method many people can be saved and treated to avoid heart related diseases. Overall, the price of equipment which are used today for detection of heart related diseases are generally very expensive and not possible for everyone to buy. Also, high blood pressure is also found to be one more cause for heart stroke disease. By using the method of automation of new generation technology the problem can be reduced and help many people to cure from heart related issues. Machine learning is played as an important technology to solve the issues related to heart diseases. With the aid of technological development, people with heart illnesses can also be identified and their likelihood of survival predicted, potentially saving many lives. In the case of a prediction model, it is also necessary to take into consideration factors like age, gender, bmi, other illnesses if present, medical history, drinking and smoking habits, level of glucose, type of work the person performs, etc. that will aid in determining the underlying cause of deaths from heart-related diseases. With the aid of a prediction model, algorithms and approaches such as logistic regression, random forest, naive Bayes, XG boost, and others assist in making comparisons and maintaining records of model correctness. By utilizing several models and keeping records, it will be possible to build the therapeutic setting's guiding principles.

- Since a few years ago, the prevalence of cardiovascular disorders has been rising quickly over the globe. Even though these diseases have been identified as the leading cause of death, they have also been identified as the most controllable and preventable diseases. Heart stroke is mostly brought on by artery obstruction. It happens when the heart cannot effectively pump blood throughout the body.

- One of the major contributing factors to developing heart disease is having high blood pressure. According to a report, 35% of people worldwide had hypertension between 2011 and 2014, which is a risk factor for heart disease. Like this, there are other factors that can lead to heart disease, including obesity, inadequate nutrition, elevated cholesterol, and a lack of exercise. Understanding cardiac disorders is crucial for prevention. Nearly 47% of fatalities occur outside of hospitals, which demonstrates that people frequently ignore warning indications.

- Heart disorders today shorten a person's life expectancy. To achieve its goals for preventing non-communicable diseases (NCDs), the World Health Organization (WHO) adopted targets in 2013. By 2025, at least 50% of patients with cardiovascular illnesses are anticipated to

have access to the right medications and medical advice. In 2016, 17.9 million fatalities worldwide—or 31% of all deaths—were directly related to cardiovascular disease.

- The detection of heart problems is a significant obstacle. It is challenging to determine whether a person has a heart condition or not. Although there are technologies that can predict heart disease, they are either expensive or ineffective at estimating the likelihood of heart disease in humans. There has been extensive research done in this field because, according to a World Health Organization (WHO) report, only 67% of cardiac disease can be predicted by medical professionals. In India, there is a severe lack of access to high-quality medical care and hospitals in rural areas. According to a 2016 WHO report, just 58% of doctors in metropolitan regions and 19% in rural areas hold medical degrees.

- Every 40 seconds, someone in the USA suffers a heart attack, which results in more than one death each year. Turkmenistan had the highest death rate until 2012 at 712 per 100,000 inhabitants. On the other hand, Kazakhstan has the second-highest rate of heart disease-related deaths. In this series, India is ranked 56th. Based on the study, there were 1.3 million cardiovascular fatalities among people aged 30-69, of which 0.9 million (68.4%) were due to coronary heart diseases and 0.4 million (28.0%) to stroke. The prediction of any heart illness in humans is a serious difficulty for medical science. Machine learning may be a promising option. Neural networks, decision trees, KNNs, and other methods can be used to forecast heart disorders.

Chapter 2: LITERATURE REVIEW

Various facets of heart stroke prediction have been inspected in the literature which identify heart related disease using machine learning and data mining techniques. In [1] authors have investigated various risk factors for the stroke prediction. In this article SVM technique is used to identify the relation between risk factor and its impact on stroke.

In [2] authors used machine learning algorithms to classify, identify and predict the stroke based on available medical information. In this article random forest algorithms have been used for predicting stroke and analyzed various risk levels with the stroke and achieved better accuracy than the other existing techniques.

In order to detect the stroke possibility authors in [3] have used various data mining classification techniques like Multi-layer perceptron, c4.5 and Jrip. Authors claimed to achieve 95% accuracy but time taken during training and prediction is a bit more due to the use of complex algorithms.

In [4] authors suggested decision trees, Naïve Bayes and Neural networks algorithms to predict the stroke. Maximum accuracy achieved using decision tree techniques was 75% in comparison with other two algorithms. However, this algorithm was not well suited for real world problems due to the confusion matrix values.

In [5] authors identified and analyzed heart disease, age, hypertension and glucose level as most significant factors in prediction of stroke in patients using principal component analysis and statistical techniques. Authors claimed that the perceptron neural network achieved maximum accuracy using these above-mentioned factors and minimum miss rate in comparison with existing benchmarking methods.

In [6, 7] authors proposed a model using logistic regression techniques of machine learning to improve the accuracy of predicting heart disease in patients. Its key drawback is that it cannot perform cardiac disease prediction with high accuracy.

In [8], a heart stroke prediction model is proposed based on various parameters viz. gender, age, smoking status, body mass index, glucose level, residence and work type using a variety of ML algorithms. Authors implemented Logical regression, SVM, KNN, Naïve Bayes, Random Forest and decision tree techniques and found out that the decision tree technique is most efficient with 100% accuracy.

In [9] authors analysed and proposed stroke prediction using various ML methods with a variety of parameters. The analysis has shown important and non-important features that can be used for prediction. Authors claimed that gradient Boosting techniques provide better results with an ROC area score 0.090. After fine tuning a few parameters like tree depth, learning rate optimization, minimum sample split and number of trees the ROC score is improvised to 0.94.

In[10], authors compared machine learning models with Cox proportional hazards model for prediction of heart stroke. This article proposed conservative mean algorithm for selecting robust features. Using SVM and a proposed feature selection algorithm the model achieved more area under the ROC curve as compared to other existing models. This algorithm can be used in scenarios where risk factors are not clearly known and missing data is present.

In [11] authors used multilayer perceptron (MLP) to predict stroke patient mortality rate. This article utilized a dataset of stroke patient considering the risk factors like hypertension, diabetes mellitus, embolism, sex, atrial fibrillation etc. using various MLP techniques like backpropagation (BP), quick propagation (QP), delta bar delta (DBD), conjugate gradient descent etc. MLP trained QP attained the maximum accuracy, specificity, AUC and sensitivity values.

In [12] authors proposed an algorithm to calculate stroke severity index using data mining methods and develop a prediction model using multiple linear regression (MLR) and KNN. In this KNN algorithm showed better results than MLR on seven predictive features.

In [13] authors analysed the cardiovascular health study dataset for stroke prediction using multiple ML algorithms. For feature selection a decision tree with C4.5 techniques is used, for reducing dimension PCA is used and SVM and ANN is used for classification. The authors claimed the optimal result is achieved using a combined approach.

In [14] authors proposed ANN model with back propagation algorithm for predicting heart stroke and aid in existing diagnostic methods. This article claims that ANN provides higher consistency with accuracy in the prediction of the stroke.

In [15] the author suggested a Neural Network, decision tree and Naïve Bayes algorithm for prediction of heart disease using significant risk factors. It uses PCA technique for dimension reduction and predicts whether a patient is suffering from stroke or not.

In [16] authors developed a model for stroke pre-diagnosis with potential risk factors. Logistic regression technique has been applied on the Korea National Health Insurance Service database.

Chapter 3: METHODOLOGY OF THE PROJECT

In the proposed version, coronary heart stroke prediction is achieved on a dataset amassed from Kaggle. The version predicts the probabilities someone may have stroke primarily based totally on signs like age, gender, common glucose level, smoking status, frame mass index, paintings kind and house kind.

The under parent suggests that they go with the drift diagram of the proposed method. Initially the given dataset is given to the pre-processing methods. The information is pre-system and characteristic scaling through the use of standardization methods. Then choose the functions from the pre-processed information, which make the classifier as higher prediction and type purpose. Then the extraordinary device mastering classifier version to expect the coronary heart stroke significantly.

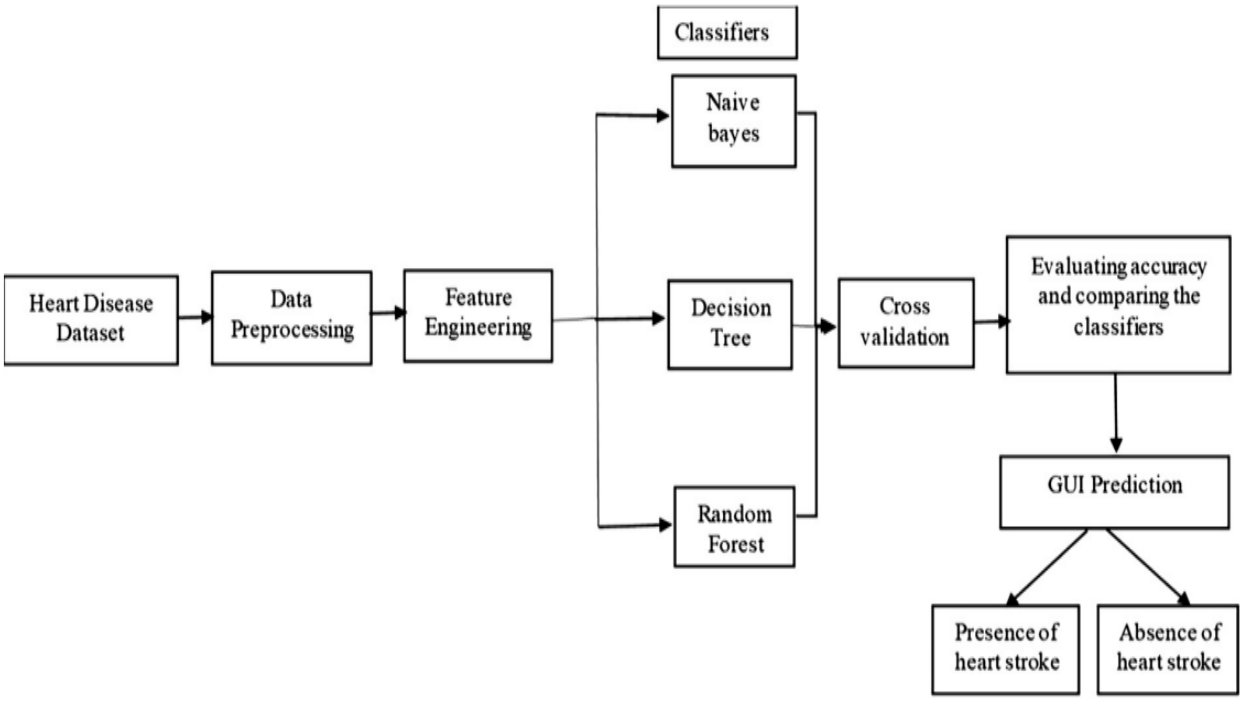

Figure - 3 Proposed system framework predicting heart stroke.

The technology of Machine learning to improve the process of Automation -

The process by which computers learn to recognize patterns, or the capacity to continuously learn from and make predictions based on data, then adjust without being specifically programmed to do so, is known as machine learning (ML), a subcategory of artificial

intelligence. The operation of machine learning differs according to the task at hand and the algorithm employed to do it. However, at its foundation, a machine learning model is a computer that analyzes data to spot patterns before using those realizations to better fulfill the work that has been given to it. Machine learning can automate any operation that depends on a set of data points or rules, even more difficult tasks like taking customer service calls and looking over resumes. Machine learning algorithms work with more or less human interaction or reinforcement depending on the circumstance. supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four main machine learning models. In supervised learning, a computer is given a set of labeled data so it may learn how to perform a human skill. Given that it aims to mimic human learning, this model is the simplest. When unsupervised learning is used, the computer uses unlabelled data to discover patterns and insights that were previously undiscovered. In semi-supervised learning, the computer is given a collection of partially labeled data and is given the task of understanding the parameters for interpreting the unlabelled data using the labeled data. Through observation of its surroundings, the computer employs reinforcement learning to choose the best course of action that will minimize risk and/or maximize reward. This method is iterative and calls for a reinforcement signal of some form to aid the computer in choosing the optimum course of action.

Chapter 4: MACHINE LEARNING TECHNIQUE

The algorithms that transform a data set into a model are known as machine learning algorithms, and they are the heart of machine learning.

4.1 Decision tree classifier –

A decision tree is used in decision tree learning as a predictive model to link observations about an item to judgments about its intended value. It is one of the predictive modeling strategies used in data mining, machine learning, and statistics. Classification trees are tree models where the goal variable can take a finite set of values. In these tree-like structures, the leaves stand in for class labels, while the branches represent the features that combine to form those class labels. Regression trees are decision trees when the target variable can take continuous values (usually real numbers). A decision tree can be used in decision analysis to reflect decisions and decision-making formally and visually. In data mining, a decision tree describes data but not decisions; the resulting classification tree can be an input for decision-making. The non-parametric supervised learning approach used for classification and regression applications is the decision tree. It is organized hierarchically and has a root node, branches, internal nodes, and leaf nodes. A decision tree is a particularly particular sort of probability tree that enables you to choose a course of action. For instance, you might have to decide between producing item A or item B, or between investing in option 1, option 2, or option 3.

The decision tree algorithm belongs to the supervised learning subcategory. They can be applied to both classification and regression issues. A decision tree employs a tree representation to answer an issue, with each leaf node denoting a class label and the inside nodes of the tree displaying attributes. Using the decision tree, any Boolean function on discrete qualities can be represented. The most effective and well-liked technique for categorization and prediction is the decision tree. A decision tree is a type of tree structure that resembles a flowchart, where each internal node represents a test on an attribute, each branch a test result, and each leaf node (terminal node) a class label.

4.2 K- nearest neighbour

The k-nearest neighbor's algorithm, sometimes referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to produce classifications or predictions about the grouping of a single data point. Although it can be applied to classification or regression issues, it is commonly employed as a classification algorithm because it relies on the idea that comparable points can be discovered close to one another. One of the simplest machine learning

techniques based on supervised learning is K-Nearest Neighbour. The K-NN algorithm assumes that the new case and the existing cases are comparable, and it places the new instance in the category that is most like the existing categories. The K-NN algorithm saves all the information that is accessible and categorizes fresh input based on similarity. This means that using the K-NN method, fresh data can be quickly and accurately sorted into a suitable category. The K-NN approach can be used for both classification and regression problems, but it is more frequently used for classification issues. Because K-NN is a non-parametric method, it makes no assumptions about the underlying data. It is also known as a lazy learner algorithm since it keeps the dataset rather than learning from the training set right away. Instead, it uses the dataset to act when classifying data. The KNN algorithm simply saves the dataset during the training phase and classifies fresh data into a category that is quite similar to the training data. The k-nearest neighbors (KNN) technique calculates the likelihood that a data point will belong to one group, or another based on which group the data points closest to it do. An example of a supervised machine learning technique used to resolve classification and regression issues is the k-nearest neighbor algorithm. However, classification issues are its primary application. KNN is a lazy learning and non-parametric algorithm. Because it doesn't perform any training when you provide the training data, it is known as a lazy learning algorithm or lazy learner. It doesn't make any calculations during the training period, just keeps the data. Before a query is run on the dataset, a model is not built. KNN is hence perfect for data mining.

4.3 Support vector machine -

Cortes and Vapnik created the support vector machine (SVM), a relatively new classification or prediction technique, as a result of their work with the statistical and machine learning research communities in the 1990s. SVM seeks to categorize situations by locating a dividing line known as a hyperplane. The fundamental benefit of the SVM is its relative simplicity in dealing with "the high dimensionality problem," or the issue that occurs when there are many input variables in comparison to the number of available observations. Additionally, the SVM approach may have the significant discriminative potential for classification because it is data-driven and feasible without a theoretical foundation, particularly in situations where sample numbers are limited. Recently, this technology has been applied to enhance strategies for spotting disorders in clinical settings. Additionally, SVM has proven to be quite effective at resolving classification issues in bioinformatics. In this study, we employ SVM to forecast the medication adherence of HF patients using common characteristics that may be reasonably collected. A deep learning system known as a support vector machine (SVM) uses supervised learning to classify or predict the behavior of groupings of data. Supervised learning systems in AI and machine learning give input and intended output data that are labeled for classification. A learning foundation for future data processing is provided by classification. Two data groups are sorted using support vector machines according to similar categorisation. To divide the groups into different patterns, algorithms create lines (or hyperplanes).Classification and regression issues are resolved using Support Vector Machine, or SVM, one of the most used supervised learning techniques. It is mostly used, nevertheless, in Machine Learning Classification problems. In order to swiftly

categorize new data points in the future, the SVM algorithm aims to determine the optimum line or decision boundary that can divide n-dimensional space into classes. The name for this ideal decision boundary is a hyperplane. The extreme vectors and points that help create the hyperplane are chosen via SVM. The algorithm known as a Support Vector Machine is named after these extreme examples, which are referred to as support vectors.

4.4 Random forest

Regression and classification both employ Random Forest methods. It organizes the data into a tree and bases its predictions on it. Even when significant portions of record values are missing, the Random Forest algorithm can still produce the same results when applied to huge datasets. The decision tree's generated samples can be preserved and used for different sets of data. In a random forest, there are two stages: first, generate a random forest, and then, using a classifier produced in the first stage, make a prediction. A "forest" is created by growing and combining various decision trees using the supervised machine learning method Random Forest. Both R and Python support their use for classification and regression issues.

These bushes are generated at the time of schooling, and the outputs are acquired from every choice tree. For the very last prediction from this set of rules, a way called "voting" takes place. This approach means that every choice tree votes for an output elegance (in this case, the 2 instructions are:`stroke` and 'no stroke'). The random wooded area chooses the elegance with the most range of votes because of the very last prediction. The accuracy acquired with the aid of schooling the version the use of this specific set of rules is 73%. The F1 Score acquired with this set of rules is 72.7%.

Chapter 5: RESULTS AND DISCUSSION

By making use of one of a kind system mastering algorithms after which the usage of deep mastering to look at what distinction comes while its miles are implemented to the records, 3 techniques have been used. In the primary method, everyday dataset that is obtained is without delay used for classification, and withinside the 2d method, the records with function choice are looked after and there's no outliers detection. The effects which might be executed are pretty promising after which withinside the 1/3 method the dataset changed into normalized looking after the outliers and function choice; the effects executed are a great deal higher than the preceding techniques, and while as compared with different studies accuracies, our effects are pretty promising.

- **Data Processing**

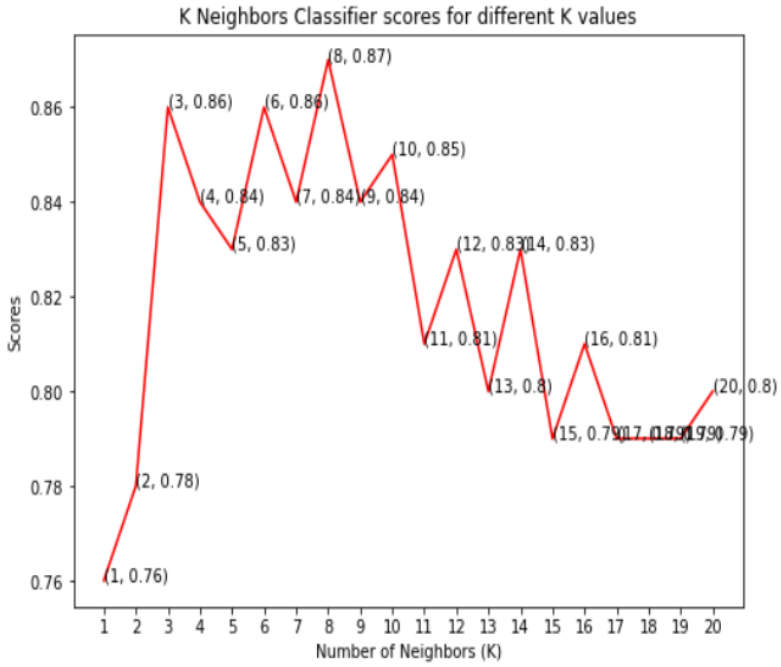

Figure -4 score achieving using K-neighbor

In the figure mentioned above, it can be seen clearly that the maximum score achieved using the K Neighbor Classifier is 0.87 for the 8 neighbors Which is one of the highest scores achieved as compared to other methods mentioned previously.

- **Support Vector Classifier -**

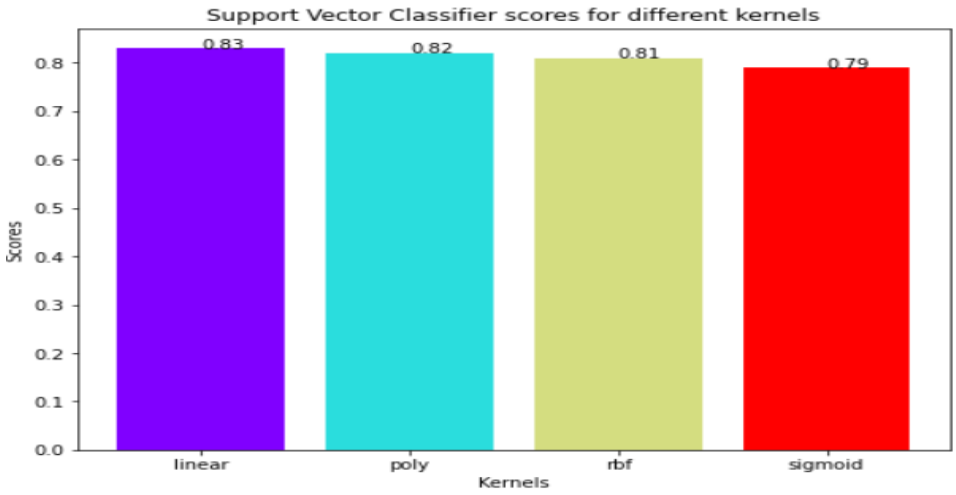

Figure -5 Support vector classifier for different kernels

The figure mentioned above, can be seen clearly as it shows the scores for the different kernels with the help of the Support Vector Classifier as the method. Here it can be clearly seen that the linear kernel performed the best, among the others being slightly better than the rbf Kernel.

- **Decision Tree Classifier -**

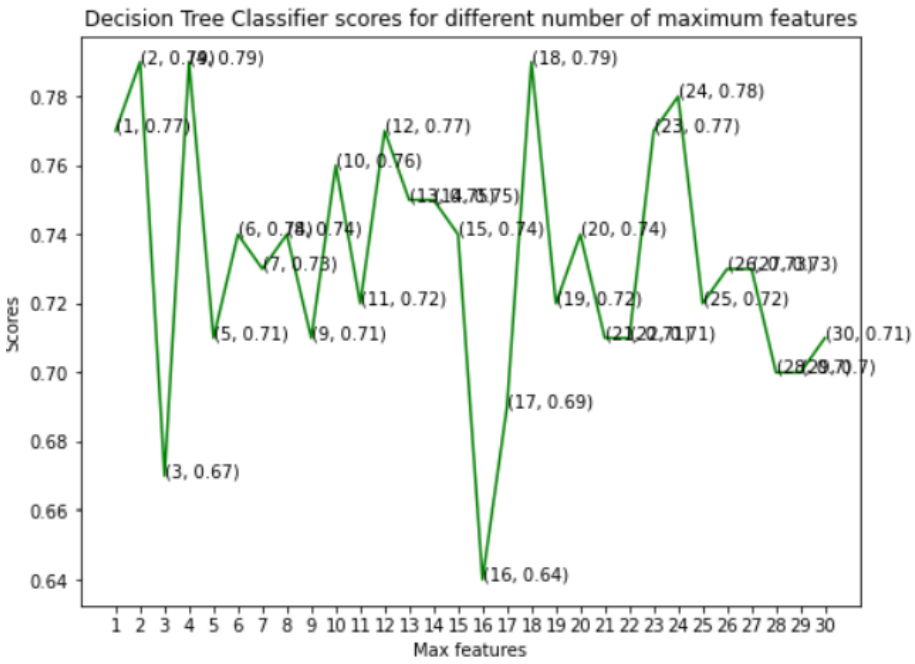

Figure -6 Decision Tree Classifier for score finding

In the figure mentioned above, it can be seen that here Decision Tree Classifier has been used for scoring the different numbers of the maximum features as shown in the above-mentioned figure. The model showed that it has achieved the best accuracy at the three values of the maximum features, 2, 4 and 18.

- **Random Forest Classifier -**

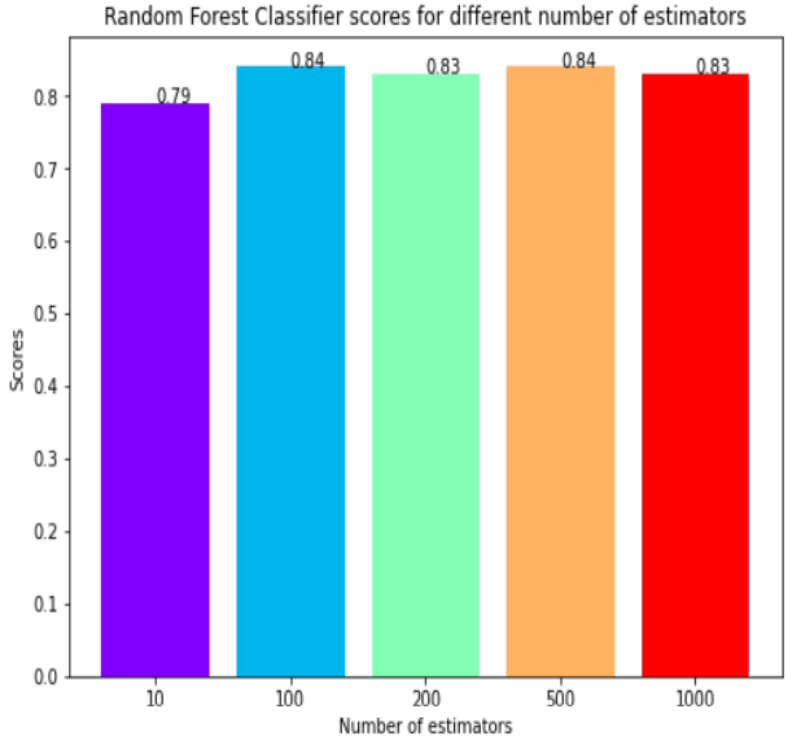

Figure- 4 Random Forest classification

In the figure mentioned above, it can be seen that here Random Forest Classifier has been used as a method for scoring the different number of estimators as follows. Here it can be seen, and a result can be achieved that the maximum score has been achieved here only when the total estimators are 100 or 500 which cannot be said for the same for others also. The same does not happen for others as it happens in the case when Random Forest Classifier is used.

Chapter 6: CONCLUSION

Machine Learning is used to predict whether a person is suffering from heart disease or not. After importing the data, it is analyzed with the help of using the plots. Then, the dummy variables are generated for categorical features and scaled for other features. Then the four Machine Learning algorithms are applied which are named as follows, K Neighbours Classifier, Support Vector Classifier, Decision Tree Classifier and Random Forest Classifier. I varied parameters across each model to improve their scores. In the end, K Neighbours Classifier achieved the highest score of 87% with 8 nearest neighbors. So accordingly, it is seen that K

Neighbours Classifier is the best method to be used for the project as compared to the other methods as K Neighbour Classifier has the highest score among all the others and it also has 8 nearest neighbors with it also.

In this paper, we proposed 3 techniques wherein comparative evaluation changed into completed and promising consequences have been done. The end which we determined is that gadget studying algorithms accomplished higher on this evaluation. Many researchers have formerly counseled that we need to use ML in which the dataset isn't always that large, that's proven in this paper. The techniques which can be used for assessment are confusion matrix, precision, specificity, sensitivity, and F1 score. For the thirteen capabilities which have been withinside the dataset, KNeighbors classifier accomplished higher withinside the ML method while records preprocessing is applied.

The computational time changed additionally decreased that's beneficial while deploying a version. It changed into additionally determined out that the dataset need to be normalized; otherwise, the education version receives overfitted on occasion and the accuracy done isn't always enough while a version is evaluated for real-international records troubles that can range extensively to the dataset on which the version changed into trained.

References –

[1] S. M. Reza, M. M. Rahman, M. H. Parvez, M. S. Kaiser, and S. Al Mamun, “Innovative approach in web application effort & cost estimation using functional measurement type,” in 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). IEEE, 2015, pp. 1–7.

[2] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, “Classification of heart stroke disease using machine learning algorithms,” Neural Computing and Applications, vol. 32, no. 3, pp. 817–828, Feb. 2020.

[3] C.V. Krishna Veni, T. R. Shoba, On the classification of imbalanced Datasets, International Journal of Computer Science & Technology 2011; 2:145-148

[4]. Turkmen HI, Karsligil ME. Advanced computing solutions for analysis of laryngeal disorders. Medical & biological engineering & computing. 2019 Nov;57(11):2535-52.

[5]. Lapham AC, Bartlett RM. The use of artificial intelligence in the analysis of sports performance: A review of applications in human gait analysis and future directions for sports biomechanics. Journal of Sports Sciences. 1995 Jun 1;13(3):229-37.

[6]. Orjuela-Cañón AD, Ruíz-Olaya AF, Forero L. Deep neural network for EMG signal classification of wrist position: Preliminary results. In2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI) 2017 Nov 8 (pp. 1-5). IEEE.

[7]. Reaz MB, Hussain MS, Mohd-Yasin F. Techniques of EMG signal analysis: detection, processing, classification and applications. Biological procedures online. 2006 Dec;8(1):11-35.

[8]. Buongiorno D, Cascarano GD, Brunetti A, De Feudis I, Bevilacqua V. A survey on deep learning in electromyographic signal analysis. InInternational Conference on Intelligent Computing 2019 Aug 3 (pp. 751-761). Springer, Cham.

[9]. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, ‘‘Can machine-learning improve cardiovascular risk prediction using routine clinical data?’’ PLoS ONE, vol. 12, no. 4, Apr. 2017, Art. no. e0174944.

[10]. J. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, and J. H. Heo, ‘‘Machine learning-based model for prediction of outcomes in acute stroke,’’ Stroke, vol. 50, no. 5, pp. 1263–1265, May 2019.

[11]. R. Garg, E. Oh, A. Naidech, K. Kording, and S. Prabhakaran, ‘‘Automating ischemic stroke subtype classification using machine learning and natural language processing,’’ J. Stroke Cerebrovascular Diseases, vol. 28, no. 7, pp. 2045–2051, Jul. 2019.

[12]. P. Xu, G. Zhao, Z. Kou, G. Fang, and W. Liu, ‘‘Classification of cancers based on a comprehensive pathway activity inferred by genes and their interactions,’’ IEEE Access, vol. 8, pp. 30515– 30521, 2020.

[10]. J. Bamford, P. Sandercock, M. Dennis, C. Warlow, and J. Burn, ‘‘Classification and natural history of clinically identifiable subtypes of cerebral infarction,’’ Lancet, vol. 337, no. 8756, pp. 1521– 1526, Jun. 1991

[11]. R. I. Lindley, C. P. Warlow, J. M. Wardlaw, M. S. Dennis, J. Slattery, and P. A. Sandercock, ‘‘Interobserver reliability of a clinical classification of acute cerebral infarction.,’’ Stroke, vol. 24, no. 12, pp. 1801–1804, Dec. 1993.

[12]. P. Amarenco, J. Bogousslavsky, L. R. Caplan, G. A. Donnan, and M. G. Hennerici, ‘‘Classification of stroke subtypes,’’ Cerebrovascular Diseases, vol. 27, no. 5, pp. 493–501, 2009. [13]. S. Ricci, S. Lewis, and P. Sandercock, ‘‘Previous use of aspirin and baseline stroke

severity: An analysis of 17 850 patients in the international stroke trial,'' Stroke, vol. 37, no. 7, pp. 1737–1740, Jul. 2006.

[13] L. Amini, R. Azarpazhouh, M. T. Farzadfar, S. A. Mousavi, F. Jazaieri, F. Khorvash, R. Norouzi, and N. Toghianfar, "Prediction and control of heart stroke by data mining," International Journal of Preventive Medicine, vol. 4, no. Suppl 2, pp. S245–249, May 2013.

[14] S.-F. Sung, C.-Y. Hsieh, Y.-H. Kao Yang, H.-J. Lin, C.-H. Chen, Y.- W. Chen, and Y.-H. Hu, "Developing a heart stroke severity index based on administrative data was feasible using data mining techniques," Journal of Clinical Epidemiology, vol. 68, no. 11, pp. 1292–1300, Nov. 2015.

[15] M. C. Paul, S. Sarkar, M. M. Rahman, S. M. Reza, and M. S. Kaiser, "Low cost and portable patient monitoring system for ehealth services in bangladesh," in 2016 International Conference on Computer Communication and Informatics (ICCCI), 2016, pp. 1–4.

[16] Jeena R.S., Kumar S. Stroke prediction using SVM**,** Proc. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) (2016), pp. 600-602, 10.1109/ICCICCT.2016.7988020

[17] Vamsi Bandi, Debnath Bhattacharyya, Divya Midhunchakkravarthy: Prediction of Brain Stroke Severity Using Machine Learning. In: International Information and Engineering Technology Association (2020).

[18] Ohoud Almadani, Riyad Alshammari: Prediction of Stroke using Data Mining Classification Techniques. In: International Journal of Advanced Computer Science and Applications (IJACSA) (2018).

[19] Kansadub, T., Thammaboosadee, S., Kiattisin, S., Jalayondeja, C.: Stroke risk prediction model based on demographic data. In: 8th Biomedical Engineering International Conference (BMEiCON) IEEE (2015).

[20] Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. Healthcare Analytics, 2, 100032.