Stroke Disease Identification System by using Machine Learning Algorithm

K.Veena Kumari [1], Mr.K. Siva Kumar [2],Dr. M.Sreelatha [3]

M.Tech Scholar-CSE Department, R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India. veenakumarikotapati@gmail.com

[2] Assistant Professor, CSE Department, R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India. kommerlasivakumar@gmail.com

[3] Professor &HOD,CSE Department, R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India. lathamoturicse@gmail.com

Abstract
A stroke is a medical disease where a blood vessel in the brain ruptures, causes damage to the brain. If the flow of blood and different nutrients to the brain is intermittent, symptoms may occur. Stroke is other reason for loss of life and widespread disorder. The prevalence of stroke is high in growing countries, with ischemic stroke being the high usual category. Many of the forewarning signs of stroke can be recognized the seriousness of a stroke can be reduced. Most of the earlier stroke detections and prediction models uses image examination tools like CT (Computed Tomography) scan or MRI (Magnetic Resonance Imaging) which are costly and difficult to use for actual-time recognition. Machine learning (ML) is a part of artificial intelligence (AI) that makes software applications to gain the exact accuracy to predict the end results not having to be directly involved to get the work done. In recent times ML algorithms have gained lot of attention due to their accurate results in medical fields. Hence in this work, Stroke disease identification system by using Machine Learning algorithm is presented. The ML algorithm used in this work is Artificial Neural Network (ANN). The result analysis of presented ML algorithm is compared with different ML algorithms. The performance of the presented approach is compared to find the better algorithm for stroke identification.

Keywords:
AI, Stroke predictions, random forest, Machine Learning.

1. Introduction

A stroke, also termed as a Cerebrovascular Accident (CVA) if some area of the brain discontinues its blood flow and the area of the body that regulates the blood cells that lose blood stops functioning. This interruption in blood supply can be ischemic due to absence of blood circulation or bleeding due to bleeding into brain tissue.

Stroke is a seriousness of health due to strokes can cause loss of life or long lasting disorder [1]. A stroke is a health situation in which low blood supply to brain leads to cell death. A stroke happens if blood supply to different parts of brain is interrupted or reduced, cells in that areas dying due to the lack of nutrients and oxygen they need. A stroke is a seriousness of health that needs instant medical support. Advance diagnosis and suitable administration are essential to avoid later injuries to the damaged parts of brain and different problems in remaining parts of the body. Of the four important cardiovascular disorders, stroke is highly harmful and deadly, and if the stroke is identified before, it can save a patient's life. [2].

Stroke disease is one of the normal acute diseases that lead to working disabilities in middle age people and the elderly, leading to significant problems with community or financial activities. Methods that can estimate actual-time health situations and disorders using different medical care centers are gaining interest [3]. Stroke is one of the main causes of loss of life around the globe, the most general subtype of Acute Ischemic Stroke (AIS) [4]. A great advancement of technology in medical field is feasible at present to predict at beginning of stroke by using Machine Learning methods [5]. Stroke can be detected examining the blood; brain imaging like Computed Tomography, Magnetic Resonance Imaging, and X-ray; ECG and EEG; and neurological physiological techniques like induced potential examinations. The methods, Computed Tomography and Magnetic Resonance Imaging are frequently used to detect stroke, and still carry dangers like being subjected to radiation or potential allergic reactions to different agents utilized [6].Many people fall victim to stroke and the numbers are increasing more in the developing countries. Several risk factors play a role in determining various types of stroke. Predictive algorithms establish a relationship between the risk factors and the types of strokes. Machine learning algorithms helps in early diagnosis and prevention of these stroke cases. It is very

Manuscript received November 5, 2023
Manuscript revised November 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.11.23

difficult to predict the stroke symptoms and outbreaks taking note on the risk factors, since stroke is a complicated medical condition. This has enhanced the interests of people in technology sector to apply machine learning techniques to diagnose the stroke effectively by routinely collecting the datasets and delivering the accurate results for diagnosis. The designs in Machine Learning are structured in accurate estimation and provide accurate examination. ML is an approach of artificial intelligence that can use several imaging aspects, including those not visible to mankind with constant precision [7]. Furthermore, many papers have been published frequently which explains machine learning techniques to address the issue. Classification of stroke disease using machine learning algorithms: This paper presents a prototype to classify stroke which uses machine learning algorithms. Machine learning can be portrayed as a significant tracker in areas like surveillance, medicine, data management with the aid of suitably trained machine learning algorithms. The agenda of this paper is to identify the better machine learning techniques used to identify stroke, which will also help to understand and resolve the problem in more effective ways.

The rest of the work is organized as follows: section II describes the related work. The section III demonstrates stroke disease identification system using ML algorithm. The section IV describes the result analysis of presented stroke disease identification system. Finally this work is concluded in section V.

2. Related Work:

Kunder Akash Mahesh, Shashank H N, Srikanth S, Thejas A M et. al., [8] presented prediction of stroke using ML. Numerous analysis and assessment models have shown acceptable accuracy in determining the DT, Naive Bayes and Neural Network, stroke patients. This project, therefore, supports assessing stroke danger using the prediction model and to give an individual caution and way of living correction message through a network application.

Minhaz Uddin Emon, Maria Sultana Keya, Tamara Islam Meghla, Mahfujur Rahman, M Shamim Al Mamun, and M Shamim Kaiser, et. al., [9] discussed the Performance Analysis of ML Approaches in Stroke Prediction. They provide initial detection of stroke disorder by using various ML methods to predict blood pressure, body mass index level, heart disorder, diabetes, smoking status, early stroke and period of occurrence. This design provides better precision for stroke prediction.

Vamsi Bandi, Debnath Bhattachrayya, Divya Midhunchakravarthy et. al,. [10] has demonstrated the Prediction of brain stroke using ML. They applied ML models to identify, classify and predict the stroke from the medical data. Stroke prediction algorithm is presented to overcome the earlier limitations. They achieved better accuracy compared to traditional stroke prediction methods.

Tasfia Ismail Shoily, Tajul Islam, Sumaiya Jannat, Sharmin Akter Tanna,Taslima Mostafa Alif, Romana Rahman Ema, et. al., [11] has discussed about the detection of stroke disease using ML Alogrithms. Four ML algorithms are utilized to identify the type of stroke that happen or happening in the form of a human physical condition and medical report data. Machine Learning algorithms support a better awareness of disorders and better medical care companion.

Thrombophilia testing in young patients with ischemic stroke [12]: The possible significance of thrombophilia in ischemic stroke remains controversial. We aimed to study inherited and acquired thrombophilias as risk factors for ischemic stroke, transient ischemic attack (TIA) and amaurosisfugax in young patients. We included patients aged 18 to 50 years with ischemic stroke, TIA or amaurosisfugax referred to thrombophilia investigation at Aarhus University Hospital, Denmark from 1 January 2004 to 31 December 2012 (N = 685). Clinical information was obtained from the Danish Stroke Registry and medical records. Thrombophilia investigation results were obtained from the laboratory information system. Absolute thrombophilia prevalences and associated odds ratios (OR) with 95% confidence intervals (95% CI) were reported for ischemic stroke (N = 377) and TIA or amaurosis fugax (N = 308). Thrombophilia prevalences for the general population were obtained from published data.

R. Jeena and S. Kumar et. al., [13] presents Stroke prediction using SVM. Early diagnosis of stroke is essential for timely prevention and treatment. Investigation shows that measures extracted from various risk parameters carry valuable information for the prediction of stroke. This research work investigates the various physiological parameters that are used as risk factors for the prediction of stroke. Data was collected from International Stroke Trial database and was successfully trained and tested using Support Vector Machine (SVM). In this work, we have implemented SVM with different kernel functions and found that linear kernel gave an accuracy of 90 %.

P. A. Sandercock, M. Niewada, and A. Członkowska et. al., [14] discussed The international stroke trial database.The International Stroke Trial (IST) is one of the largest randomized trials ever conducted on individual patients in acute stroke. The IST dataset includes data on 19 435 patients with acute stroke, with 99% complete follow-up. Over 26.4% patients were aged over 80 years at study entry. Background stroke care was limited and none of the

patients received thrombolytic therapy. This clinical trial was conducted between 1991 and 1996 and a pilot phase between 1991 to and 1993. This study is a large, prospective, randomized controlled trial, with 100% complete baseline data and over 99% complete follow-up data. For each randomized patient, data were extracted on the variables assessed at randomization; the early outcome point was 14-days after randomization or prior discharge, and at 6-months and provided as an analyzable database. The aim of the trial was to establish whether early administration of aspirin, heparin, both or neither influenced the clinical course of an acute is chaemic stroke.

A. Sudha, P. Gayathri, and N. Jaisankar et. al., [15] presents Effective analysis and predictive model of stroke disease using classification methods. In today world data mining plays a vital role for prediction of diseases in medical industry. Stroke is a life threatening disease that has been ranked third leading cause of death in states and in developing countries. The stroke is a leading cause of serious, long term disability in US. The time taken to recover from stroke disease depends on patients‟ severity. Number of work has been carried out for predicting various diseases by comparing the performance of predictive data mining.

Earlier System: A stroke is a medical condition in which poor blood flow to the brain results in cell death. It is now a day a leading cause of death all over the world. Several risk factors believe to be related to the cause of stroke has been found by inspecting the affected individuals. Using these risk factors, a number of works have been carried out for predicting and classifying stroke diseases. Most of the models are based on data mining and machine learning algorithms. In this work, we have used four machine learning algorithms to detect the type of stroke that can possibly occur or occurred form a person’s physical state and medical report data. We have collected a good number of entries from the hospitals and use them to solve our problem. Naive Bayes Algorithm, J48 Algorithm, KNN Algorithm, And Random Forest Algorithms are used in this approach. The classification result shows that the result is satisfactory and can be used in real time medical report.

3. Methodology and implementation

.In this work, Stroke disease identification system by using Machine Learning algorithm is presented. According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. The Fig. 1 shows the workflow of presented stroke disease identification system.

Stroke disease data set is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relavant information about the patient. The attributes are as follows:

id:unique identifier; ii) gender: "Male", "Female" or "Other"; iii) age: age of the patient; iv) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension; v) ever_married: "No" or "Yes";vi) work_type: "children", "Govt_jov", "Never_worked", "Private" or "Self-employed"; vii) Residence_type: "Rural" or "Urban"; viii) avg_glucose_level: average glucose; level in blood ix) bmi: body mass index; x) smoking_status: "formerly smoked", "never smoked", "smokes" or "Unknown"*; xi) stroke: 1 if the patient had a stroke or 0 if not.

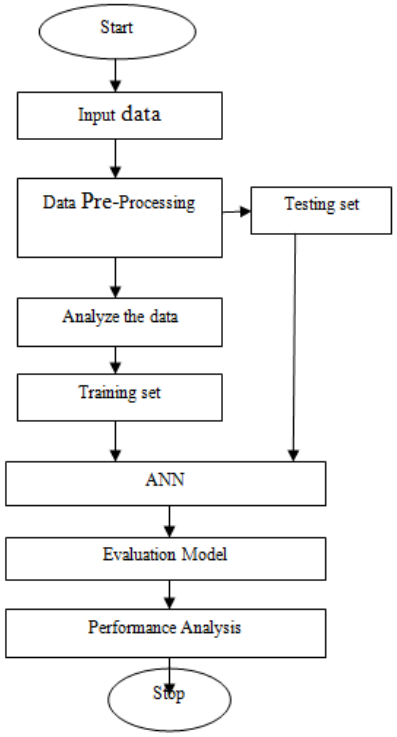


Figure 1: The workflow of stroke disease identification system

Using Dataset Preprocessing module, we will clean dataset by replacing missing values with 0 and then apply label encoding algorithm to convert non-numeric values to numeric values and then select features from dataset. The dataset is divided into train and test where application used 80% data for training and 20% for testing.

Data analysis is a process of inspecting, cleaning, transforming and modelling the data with the objective of discovering useful information. In this work Artificial neural network is used to identify the stroke. Artificial neural networks, usually simply called neural networks or neural nets, are computing systems inspired by the biological neural networks that constitute animal brains. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain.

Artificial neural networks provide a powerful tool to help doctors analyze, model, and make sense of complex clinical data across a broad range of medical applications. Most applications of artificial neural networks in medicine are classification and identification problems. To implement this project the following modules are designed:

The training set is applied as input to ANN algorithm to train the ANN model to identify the stroke disease based on the input data and dataset. The ANN classifier identifies the stroke disease as ischemic or normal. The performance of the ANN model is evaluated in terms of accuracy. In addition, the performance of presented ANN classifier is compared with other ML classifiers.

4. Result analysis:

In this section, result analysis of presented stroke disease identification using Machine Learning algorithm is discussed. Artificial neural Network is used in this approach to identify the stroke. The performance of the presented approach is evaluated using confusion matrix parameters namely True Positive (TP), True Negative (TN), False Positive, False Negative (FN) which are defined as follows:

True Positive (TP): If an instance is classified correctly as positive and actually it is positive.

True Negative (TN): If an instance is classified correctly negative and actually it is negative.

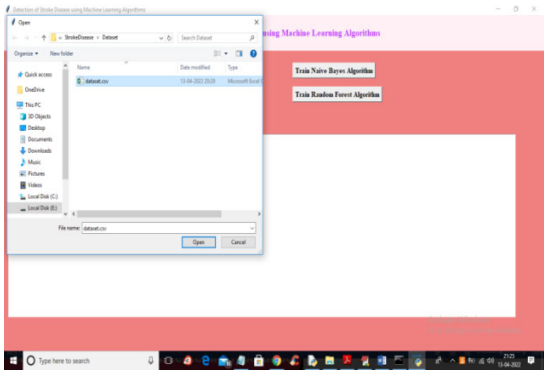
False Positive (FP): if an instance is classified incorrectly as positive but actually it is negative.

False Negative (FN): if an instance is classified incorrectly as negative but actually positive. Accuracy: It is defined as the ratio of correctly detected instances to the total number of instances and is given as

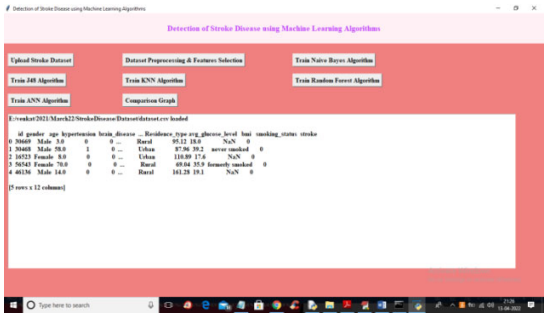
$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \times 100 \quad (1)$$

In this work different types of Machine learning algorithms namely Random Forest (RF) K-Nearest Neighbour (KNN), J48 and Decision Tree (DT) algorithms are compared with the presented ANN algorithm to determine the better one for stroke disease identification.

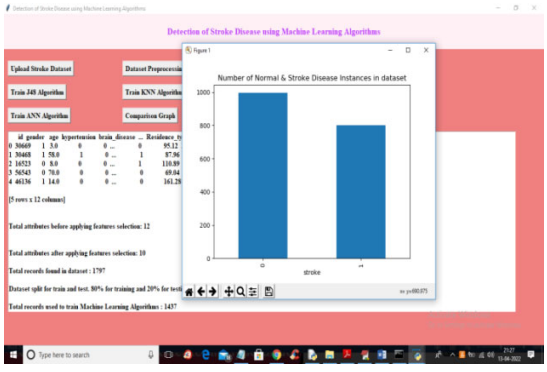
Comparison Graph: Using this module we will plot accuracy comparison graph between all algorithms.



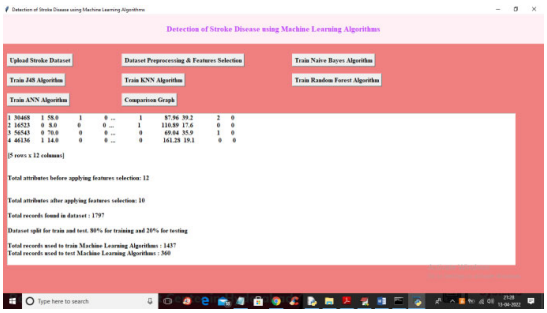
In above screen selecting and uploading dataset.csv file and then click on 'Open' button to load dataset and to get below output.



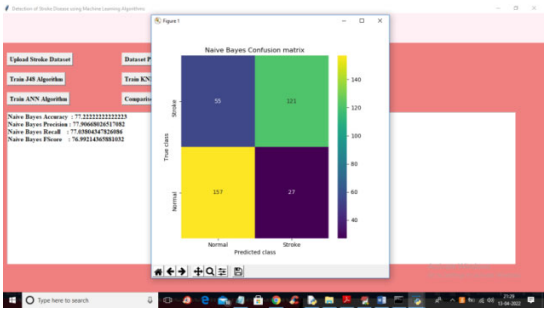
In above screen we can see dataset loaded and dataset contains so many missing and non-numeric data so click on 'Dataset Pre-processing & Features Selection' button to process dataset and to get below output.



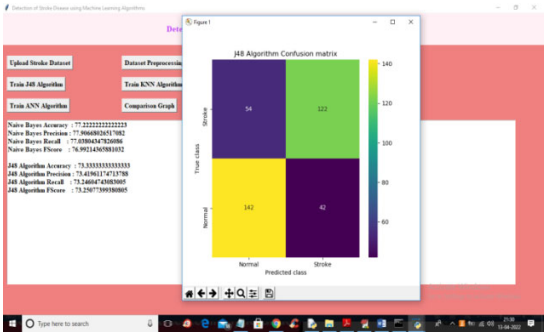
above graph x-axis represents 0 (normal) and 1 (stroke) and y-axis represents number of instances available in those categories in dataset and now close above graph and see below screen.

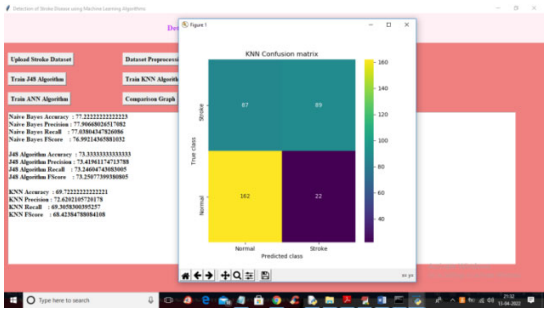


In above screen we can see all dataset converted to numeric format and then split dataset into train and test and now click on 'Train Naïve Bayes Algorithm' button to train Naïve Bayes on above dataset and get below output.



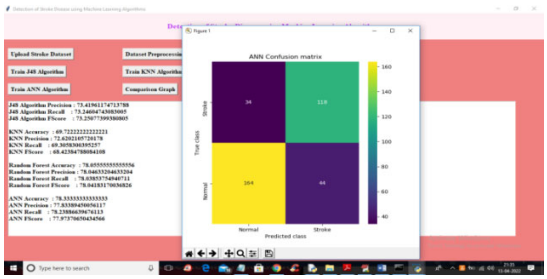
In above screen with Naïve Bayes we got 77% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by Naïve Bayes. Now click on 'Train J48 Algorithm' button to get below output.In above screen with J48 we got 73% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by J48.Now close above Graph and then click on 'Run KNN Algorithm' button to get below output.





In above screen with KNN we got 69% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by KNN. Now close above Graph and then click on 'Run Random Forest Algorithm' button to get below output.

In above screen with Random Forest we got 78% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by Random Forest. Now close above Graph and then click on 'Run ANN Algorithm' button to get below output.



In above screen with ANN we got 78.33% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by ANN and in all algorithm ANN got high accuracy. Now close above Graph and then click on 'Comparison Graph' button to get below graph.

In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics like precision, recall etc. different colour bar represents different metrics and in all algorithms ANN got high accuracy.

5. Conclusion:

In this work, stroke disease identification system by using Machine Learning algorithm is presented. In this approach ANN algorithm is used for the stroke disease identification. The performance of ANN algorithm is compared with different types of ML algorithms. Different ML algorithms include Random Forest, KNN, ANN, NB and J48. The performance of different ML algorithms is evaluated and compared. Among these algorithms ANN has better accuracy for identifying the stroke from the CT scan image; thereby the neurologist provides appropriate diagnosis to the patient. As a result patient life will be saved. In this work, the ANN provides 78.33% of accuracy for testing N images of each type of stroke. The results are depending on the number of images that are being used in training process. More images used in training process, the higher the accuracy. In the results, different colour bar represents different metrics and in all algorithms ANN got high accuracy. In future work, it is possible to extend the research by using different classification techniques. Moreover, the prediction of stroke will be done by adding some non-stroke data with the existing dataset.

References

[1] Mrs. Neha Saxena, Mr. Deep Singh Bhamra, Mr. Arvind Choudhary, Mr. Preet Maru, "BrainOK: Brain Stroke Prediction using Machine Learning", 2022 JETIR April 2022, Volume 9, Issue 4, www.jetir.org (ISSN-2349-5162)

[2] Mandeep Kaur , Sachin R. Sakhare , Kirti Wanjale , and Farzana Akter, "Early Stroke Prediction Methods for Prevention of Strokes", Hindawi Behavioural Neurology Volume 2022, Article ID 7725597, 9 pages, doi.org/10.1155/2022/7725597

[3] Yoon-A Choi, Sejin Park, Jong-Arm Jun, Chee Meng Benjamin Ho, Cheol-Sig Pyo, Hansung Lee, and Jaehak Yu, "Machine-Learning-Based Elderly Stroke Monitoring System Using Electroencephalography Vital Signals", Appl. Sci. 2021, 11, 1761, doi.org/10.3390/app11041761

[4] Haichen Zhu, Liang Jiang, Hong Zhang, Limin Luo, Yang Chen, Yuchen Chen, " An automatic machine learning approach for ischemic stroke onset time identification based on DWI and FLAIR imaging", NeuroImage: Clinical 31 (2021) 102744, doi: 10.1016/j.nicl.2021.102744

[5] Tahia Tazin , Md Nur Alam, Nahian Nakiba Dola, Mohammad Sajibul Bari, Sami Bourouis, and Mohammad Monirujjaman Khan, "Stroke Disease Detection and Prediction Using Robust Learning

[6] Yoon-A Choi, Se-Jin Park, Jong-Arm Jun 3, Cheol-Sig Pyo, Kang-Hee Cho, Han-Sung Lee, and Jae-Hak Yu, "Deep Learning-Based Stroke Disease Prediction System Using Real-Time Bio Signals", Sensors 2021, 21, 4269, doi.org/10.3390/s21134269

[7] Hyunna Lee, Eun-Jae Lee, Sungwon Ham, Han-Bin Lee, Ji Sung Lee, Sun U. Kwon, Jong S. Kim, Namkug Kim, Dong-Wha Kang, "Machine Learning Approach to Identify Stroke Within 4.5 Hours", 2020 American Heart Association, Inc, DOI: 10.1161/STROKEAHA.119.027611

[8] Kunder Akash Mahesh, Shashank H N, Srikanth S, Thejas A M, "Prediction of Stroke Using Machine Learning", Conference Paper · June 2020, w ww.researchgate.net/publication/342437236

[9] Minhaz Uddin Emon, Maria Sultana Keya, Tamara Islam Meghla, Mahfujur Rahman, M Shamim Al Mamun, and M Shamim Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction", Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA-2020), IEEE Xplore Part Number: CFP20J88-ART; ISBN: 978-1-7281-6387

[10] Vamsi Bandi, Debnath Bhattachrayya, Divya Midhunchakravarthy, "Prediction of brain stroke using Machine Learning", International Information and Engineering Technology Association, Vol. 34, No.6, 2020, pp. 753-761, doi.org/10.18280/ria.340609

[11] Tasfia Ismail Shoily, Tajul Islam, Sumaiya Jannat, Sharmin Akter Tanna,Taslima Mostafa Alif, Romana Rahman Ema, "Detection of Stroke Disease using Machine Learning Algorithms", 10th ICCCNT 2019 IEEE - 45670, July 6-8, 2019, IIT - Kanpur, Kanpur, India

[12] S. H. Pahus, A. T. Hansen, and A.-M. Hvas, "Thrombophilia testing in young patients with ischemic stroke," Thrombosis research, vol. 137, pp. 108–112, 2016.

[13] R. Jeena and S. Kumar, "Stroke prediction using svm," in 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 600–602, IEEE, 2016

[14] P. A. Sandercock, M. Niewada, and A. Członkowska, "The international stroke trial database," Trials, vol. 13, no. 1, pp. 1–1, 2012

[15] A. Sudha, P. Gayathri, and N. Jaisankar, "Effective analysis and predictive model of stroke disease using

classification methods," International Journal of Computer Applications, vol. 43, no. 14, pp. 26– 31, 2012.

1. K.Veena Kumari
M.Tech ScholarCSE Department,
R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India.
veenakumarikotapati@gmail.com

2. Mr.K. Siva Kumar
Assistant Professor, CSE Department, R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India.
kom merlasivakumar@gmail.com

3. **Dr. M.Sreelatha**

Professor &HOD,CSE Department,
R.V.R & J.C Collegeof Engineering,
Guntur, Andhra Pradesh, India.
lathamoturicse@gmail.com