iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50262

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Review on Evaluation of Stroke Prediction Using Machine Learning Methods

Asha Gaikar[1], Dr. Uttara Gogate[2], Amar Panchal[3]
[1]M.E.Student, [2]Associate Professor, [3]Assistant Professor, Department of Computer Engineering, Mumbai University

***Abstract:** This research proposes early prediction of stroke disease using different machine learning approaches such as Logistic Regression Classifier, Decision Tree Classifier, Support Vector Machine and Random Forest Classifier. In this paper we have used different Machine Learning algorithm and by calculating their accuracy we are proposing fair comparisons of different Stoke prediction algorithms by using same dataset with same no of features. The researcher will help to predict the Stroke using best Machine Learning algorithm.*
***Keywords:** Stroke Prediction, Machine Learning Algorithm, Logistic Regression Classifier, Decision Tree Classifier, Support Vector Machine, Random forest Classifier.*

I. INTRODUCTION

Stroke is a very important medical and social problem. It is the main cause of permanent disability and lack of independence in the group of adults. It is estimated that 15 million people fall for a stroke every year in the world, and about 5 million people die (Sacco et al., 2013). This disease is in third place among causes of death (after cardiovascular disease and cancer) (Mackay and Mensah, 2004).

A stroke occurs when the blood supply to various areas of the brain is interrupted or decreased,the cells in these areas are not supplied with nutrients and oxygen and begin to die. A stroke is a medical emergency that requires immediate treatment. There are two forms of stroke: ischemic and hemorrhagic. In an ischemic stroke, the drainage is blocked by blood clots, and in a hemorrhagic stroke, a weak blood vessel bursts and bleeds into the brain. Stroke can be prevented by maintaining a healthy/balanced lifestyle, i.e., avoiding poor lifestyle habits such as smoking and drinking, controlling body mass index (BMI) and average blood glucose levels, and maintaining good heart and kidney health. Predicting stroke is necessary and must be treated to prevent permanent damage or death. In this project, hypertension, BMI, heart disease, and average blood glucose level were considered as parameters for stroke prediction.

Moreover, machine learning can play an important role in the decision-making processes of the proposed prediction system. There are very few research papers in the literature that use machine learning models for stroke prediction. Machine learning algorithms include artificial neural networks (ANN), stochastic gradient descent, decision tree algorithm, KNN (Knearest Neighbor), PCA (Principal Component Analysis), CNN (Convolutional Neural Network), Naive Bayes, etc. There is a relationship between the diseases/attributes such as hypertension, BMI level, average glucose level and heart disease and stroke.

A weighted voting classifier is proposed to predict stroke using diseases/attributes such as hypertension, body mass index, heart disease, average glucose level, smoking status, previous stroke, and age.

II. LITERATURE REVIEW

Amini et al [1] investigated the prediction of stroke incidence and included 807 healthy and unhealthy subjects in their study, categorizing 50 risk factors for stroke, diabetes, cardiovascular disease, smoking, hyperlipidemia, and alcohol consumption. They used two techniques, with the c4.5 decision tree algorithm having the best accuracy, 95%, and the K-nearest neighbor algorithm having an accuracy of 94%.

Sung et al [2] performed an observation to increase a stroke severity index. They collected data from 3577 sufferers with acute ischemic stroke. For their prediction models, they used numerous statistical mining techniques and linear regression. The best predictive result was provided by the k-nearest neighbor model (95% CI).

Govindarajan et al [3] studied stroke categorization using a text mining aggregate and a machine classifier and collected data from 507 patients. For their evaluation, they used various gadget learning methods for training purposes using ANN, and the SGD set of rules gave them the first-class price that became ninety five%.

©IJRASET: All Rights are Reserved

Cheng et al [4] published a data file to assess the prognosis of ischemic stroke. In their evaluation, 82 data of ischemic stroke sufferers were used, two ANN models were used to find precision, and seventy nine percent and 95 percent were used.

Cheon et al [5] performed an observation to predict mortality in stroke patients. In their study, they used 15099 patients to determine the occurrence of stroke. They used a deep neural network to detect strokes. The authors used PCA to extract clinical records and predict strokes. They have a region under the curve (AUC) cost of 83%.

Singh et al [6] studied stroke prediction using artificial intelligence. In their studies, they used a special method to predict stroke in the cardiovascular fitness examination dataset (CHS). And they took the selection tree algorithm to extract the main aspect of the analysis. They used a neural network category algorithm to construct the version that achieved ninety seven% accuracy.

Chin et al [7] performed an investigation to stumble on automated early ischemic stroke. The main motive of their investigation was to develop a device that CNN used to automatically detect ischemic stroke. They collected 256 images to train and test the CNN model. In their device image advantageously feature the impossible area that couldn't occur of stroke, they used the information extension method to increase the collected image. Their CNN technique has given 90% accuracy.

Monteiro et al [8] conducted a study to obtain functional outcome prediction for ischemic stroke using machine learning. In their studies, they applied this approach to a patient who died three months after admission. They obtained an AUC value of more than 90%

Jae-woo Lee et al [9] The aim of this work was to calculate the 10-year stroke prediction probability and classify the user's individual stroke probability into five categories.

Philip A. Wolf et al [10] A health risk estimation function was developed for predicting stroke using the Framingham study cohort.

Adam et al [11] the study has developed a classification model for ischemic stroke using decision tree algorithm and k nearest neighbour. They have collected 400 cases from different hospitals. They have configured the parameters of k- nearest neighbor algorithm as 80% of the dataset for training the data.

Zulfiker st al [12] conducted research to determine the predictability of a stroke patient death. They identified the stroke incidence using 15,099 individuals in their research. They detected strokes using a deep neural network method.

Kansadu et al [13] conducted research to determine the risk of stroke. The authors of the research analysed the data to predict strokes using Naive Bayes, decision trees, and neural networks. They assessed their pointer's accuracy and AUC in their research.

Ali et al [14] This method improved generalization across different scanner vendors but had challenges in images with pathology. While such methods are promising, convolutional neural networks (CNN) can adapt to highly variable biomedical imaging data, are achieving state-of-the-art performance for a variety of MRI segmentation applications.

Noor et al[15] , compared performances of the existing deep learning (DL)-based methods for detecting neurological disorders from MRI data acquired using different modalities.

Table I
REVIEW OF EXISTING METHODS

Sr.No	Title & Publication year	Methodology	Disadvantages/Limitations
1	Prediction and control of stroke by data mining(1)	conducted research to predict stroke incidence, collected 807 healthy and unhealthy subjects in their study categorized 50 risk factors for stroke	Security, lack of confidence in the results of data mining and the desire to retain them exclusively for their next possible studies
2	Developing a stroke severity index based on administrative data was feasible using data mining techniques.(2)	carried out a observe to increase a stroke severity index. They accrued 3577 affected person's facts with acute ischemic stroke	it is not valid for measuring variations in stroke risk at the time of hospital admission

3	Classification of stroke disease using machine learning algorithms(3)	performed a look at to categorize stroke disorder the usage of a textual content mining aggregate and a machine studying classifier	lack of good data. While enhancing algorithms often consumes most of the time of developers in AI, data quality is essential for the algorithms to function as intended
4	Prediction of the prognosis of ischemic stroke patients after intravenous thrombolysis using artificial neural networks(4)	Two ANN models were used to find precision	the relatively small sample size may restrict the generalization of the study results
5	The Use of Deep Learning to Predict Stroke Patient Mortality(5)	They used a deep neural network approach to detect strokes.	a lack of input data separation and the lack of longitudinal data. They employed survey data, which has drawbacks which include the binary format
6	Stroke prediction using artificial Intelligence(6)	they took the decision tree algorithm to feature extract to principal component analysis	In AI sophisticated and expensive processing resources needed are unavailable to the majority of businesses.
7	An automated early ischemic stroke detection system using CNN deep learning algorithm(7)	develop a device using CNN to automatic number one ischemic stroke	a lot of training data is needed for the CNN to be effective and that they fail to encode the position and orientation of objects.
8	Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients(8)	Functional outcome prediction of ischemic stroke using machine learning	To incorporate the use of image and genetic information and to take advantage of the longitudinal aspect of the data
9	Computer Methods and Programs in the Biomedicine(9)	To calculate and predict the probability of stroke within 10 years	a limitation in accurately identifying the stroke type that was caused due to the factors, the model produced an AUC of 80percent
10	Probability of Stroke: A Risk Profile from the Framingham Study(10)	prediction of stroke using the Framingham Study cohort.	predicts only future coronary heart disease (CHD) events, however, it does not predict future total cardiovascular events, meaning that it does not predict risk for stroke, transient ischemic attack (TIA), and heart failure.

11	Classification of Ischemic Stroke using Machine Learning Algorithms(11)	has developed a classification model for ischemic stroke using decision tree algorithm and k nearest neighbour	Decision tree algorithm is not good for regression and also very expensive.
12	Predicting students' performance of the private universities of bangladesh using machine learning approaches(12)	seven different classifiers have been trained, namely: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Extra Tree Classifier seven different classifiers have been trained, namely: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Extra Tree Classifier seven different classifiers have been trained, namely: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Extra Tree Classifier seven different classifiers have been trained, namely: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Extra Tree Classifier seven different	Assumptions of some classifies are not appropriate

		classifiers have been trained, namely: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Extra Tree Classifier seven different classifiers have been trained, namely: Support Vector Machin seven different classifiers trained namely:Support vector machine,K-Nearest neighbor,Logical Regression,Decision Tree,AdaBoost,Multilayer Perception & extra tree classifier	
13	Stroke risk prediction model based on demographic data(13)	Decision tree,naïve bayes and Neural network used for predicting Stroke	Naive Bayes assumes that all predictors are independent also Neural Network don't give any insights on the structure of the function being approximated
14	Application of convolutional neural network in segmenting brain regions from mri data(14)	Convolutional Neural Network & Deep Learning technique used to segment brain regions from Magnetic Resonance Imaging	Deep learning works only with large amounts of data. It also needs extensive hardware to do complex mathematical calculations
15	Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of alzheimer's disease, parkinson's disease and schizophrenia(15)	Machine learning,data Preprocessing and Deep learning methods are used	Deep learning requires massive amount of data with long training time.

III. OUTCOME OF SURVEY

In this section, analysis is being done on the previously published papers related to work on prediction of stroke disease using different machine learning approaches and algorithms.

In paper[1] and [2] different data mining techniques have studied and the most accurate model is obtained i.e. Decision tree algorithm. It was found most accurate than others. In[3] and [11] data is collected from different places then the case sheets were utilizing and handled information were taken care of into different AI calculations. So accurate calculation with high order precision is done in this paper.

In [4]Artificial neural networks are used to establish prediction models with good performance to predict thrombolysis outcomes. So we have used these models to help physicians to discuss and explain the likely outcomes to patients and their families before stroke treatment.We have seen Deep Neural network used in paper [5],to analyse the disease with huge amount of data. The area under the curve (AUC) value of this method was 83.4%.

[7]For training and testing the convolutional neural network, patch images were used as input. Here, 256 patch images were used to train and test the CNN module, which was able to detect the ischemic stroke. It was found that the proposed method provided more than 90% of the results.In [8] a model is developed for scanning the report which can perform the CT and MRI scanning. By entering the detailed set of information and the x- ray of a patient by using image processing it divides the whole image in pixels and perform the operation.So this paper is also useful in proposed system. In paper [9] includes the Risk factors like age, BMI, cholesterol, hypertension, diabetes, smoking status and intensity, physical activity, alcohol drinking, past history (hypertension, coronary heart disease) which is we are using in our paper.

Table II
COMPARISION OF STROKE PREDICTION METHODS

Ref. No.	Dataset Size	Accuracy
1	807	95%
2	3577	95%
3	507	95%
4	82	95%
5	15099	83%
7	256	90%

There are different Machine Learning Algorithms are implemented by different authors to predict Stroke but the dataset size used by various authors is different and no of features are also different. So it is difficult to compare these methods for stroke prediction. In our proposed method we will be implementing some classifier like Logistic Regression Classifier, Decision Tree Classifier, Support Vector Machine, Random forest Classifier on same dataset with no of features so it will be fair to compare these methods on the basis of Accuracy

IV.PROPOSED SYSTEM

We have considered the previously published works to review the Machine learning techniques used for Stroke Predictions. It's been found that the majority of the research work was done on mortality rate and functional outcome as the predicted outcomes. The most commonly used techniques were random forest, support vector machines, decision trees and neural networks.

The agenda of this Proposed System is to identify the better machine learning techniques used to predict stroke, which will also help to understand and resolve the problem in more effective ways.

The proposed weighted classifier has considered gender, age, hypertension, heart disease, average glucose level, BMI, smoking status feature attributes to predict stroke. The performance evaluation reveals that weighted voting provided the highest accuracy of about 97% compared to the commonly used other machine learning algorithms

V. CONCLUSIONS

This paper showing the performance of various Machine Learning algorithms to predict the Stroke based on different Attributes. Out of all these algorithms we have chosen one algorithm with best accuracy.

We are trying to give fair comparison of all the Machine Learning algorithms for stroke prediction by implementing the same data set & same no of features.

Several assessment and prediction models, decision tree, Naive Bayes and neural network, showed acceptable accuracy in identifying patients at risk for stroke. Therefore, this paper helps to predict stroke risk using a predictive model and provides a personalized warning and lifestyle correction message through a web application. In this way, medical users are encouraged to strengthen their motivation for health management and change their health behaviour.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

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

[2] 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 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.

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

[4] C.-A. Cheng, Y.-C. Lin, and H.-W. Chiu, "Prediction of the prognosis of ischemic stroke patients after intravenous thrombolysis using artificial neural networks," Studies in Health Technology and Informatics, vol.202, pp. 115–118, 2014.

[5] S. Cheon, J. Kim, and J. Lim, "The Use of Deep Learning to Predict Stroke Patient Mortality," International Journal of Environmental Research and Public Health, vol. 16, no. 11, 2019.

[6] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), Aug. 2017, pp. 158–161.

[7] C. Chin, B. Lin, G. Wu, T. Weng, C. Yang, R. Su, and Y. Pan, "An automated early ischemic stroke detection system using CNN deep learning algorithm," in 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), Nov. 2017, iSSN: 2325-5994.

[8] M. Monteiro, A. C. Fonseca, A. T. Freitas, T. Pinho e Melo, A. P.Francisco, J. M. Ferro, and A. L. Oliveira, "Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients," IEEE/ACM Transactions on Computational Biology and Bioinformatics,vol. 15, pp. 1953–1959, Nov. 2018.

[9] "Computer Methods and Programs in the Biomedicine" - Jae–woo Lee, Hyun-sun Lim, Dong-wook Kim, Soon-ae Shin, Jinkwon Kim, Bora Yoo, Kyung-hee Cho.

[10] "Probability of Stroke: A Risk Profile from the Framingham Study" - Philip A. Wolf, MD; Ralph B. D'Agostino, PhD, Albert J. Belanger, MA; and Willim B.Kannel,MD.

[11] S. Y. Adam, A. Yousif, and M. B. Bashir, "Classification of Ischemic Stroke using Machine Learning Algorithms," International Journal of Computer Applications,vol.149, no.10,pp.26–31, Sep.2016.

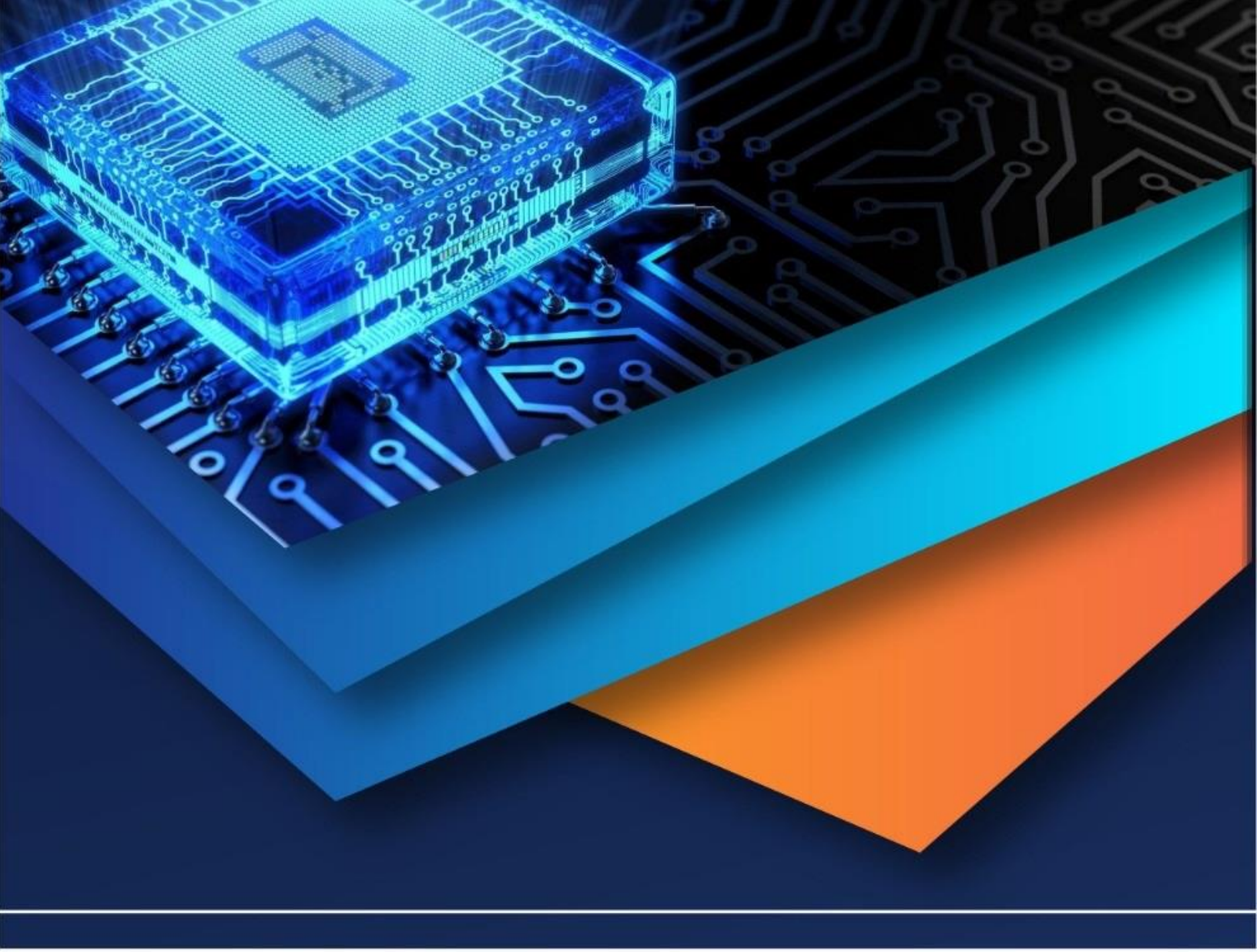
[12] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty, and M. M.Rahman, "Predicting students' performance of the private universities of bangladesh using machine learning approaches," International Journal of Advanced Computer Science and Applications, vol. 11, no. 3, 2020.

[13] T. Kansadub, S. Thammaboosadee, S. Kiattisin, and C. Jalayondeja, "Stroke risk prediction model based on demographic data," in 2015 8th Biomedical Engineering International Conference (BMEiCON), Nov.2015, pp. 1–3.

[14] H. M. Ali, M. S. Kaiser, and M. Mahmud, "Application of convolutional neural network in segmenting brain regions from mri data," in International Conference on Brain Informatics. Springer, 2019, pp.136–146.

[15] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of alzheimer's disease, parkinson's disease and schizophrenia," Brain Informatics, vol. 7, no. 1, pp. 1–21, 2020.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)