iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51431

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 V May 2023- Available at www.ijraset.com

Brain Stroke Prediction using Machine Learning

Prof. Deepali Deshpande[1], Shravani Bahirat[2], Vaisnavi Dalvi[3], Sakshi Darawade[4], Shravani Jagtap[5], Sakshi shakhawar[6]
Dept. of Electronics and Telecommunication, Vishwakarma Institute of Technology, Pune

Abstract: ***A brain stroke, is a saviour disease in which a blood cot or bleeding occur in the brain during a stoke, which can result in long-term harm. This can affects person's move, think, see, or communicate. The another medical name for the Brain Stroke is cerebrovascular or CVA is when The body part that the blood-starved brain cells regulate ceases to function. Brain Stroke is very critical situation as it directly leads to death or permanent disability. The stroke is classified into two areas ischemic stroke and Haemorrhagic. When blood clots re created in the brain and goes through the patient's streamed lodge in the brain then there is the high chances of ischemic stroke occur. Homographic stroke is second type of stoke which occurs when there is leak of blood or ruptures in the artery in brain. In early stage if the treatment get started we can treat to the ischemic stroke efficiently. If a stroke is feared, or bystanders should call emergency medical services right away. Symptoms disappear as their own in a Ischemic stroke reminded by Transient Ischemic Attack (TIA). As per the research of World Health Organization the 2nd leading cause of the death worldwide is Brain Stroke which is also responsible for the approximately 11% deaths. but can prevent up to 80% of stokes if they can be identified or predicted early stage stroke. Our ML model uses dataset to predict whether the person has any chances of getting stroke the parameters that are considered to predict stroke are gender, age, disease, smoking status, Cystatin-c , MMP10, Tau Our dataset focuses on major factors which has causes of brain stroke.***
Keywords: Machine learning, Brain Stroke. Ischemic Stroke, Transient ischemic attack.

I. INTRODUCTION

Machine Learning (ML) plays important role in medical field as it delivers an accurate and quick prediction outcome ,gives the specialized care for stroke patient. In healthcare application of Machine Learning and Deep learning increasing tremendously however such research areas does not get that much focus on the where there is actual need of research. In this project we have used 5 different ML algorithms such as Logistic Regression, KNN, Decision Tree, Random Forest and SVM to control and guess the risk of stroke. Results from the scientific web database Science Direct on ML in stroke from 2007 to 2019 identified a total of 39 studies. Out of them 10 studies Support vector Machine (SVM) is obtained as model for stroke issue. Furthermore, most studies are on the diagnosis of stroke, but the fewest In the treatment of stroke, identifying research gaps for next investigation. Nowadays, for diagnosis of stroke CT images are used. Finally, SVM and Random Forests are efficiently used techniques. In this study various approaches for stroke diagnosis and severity prediction are mentioned. To reduce the risk of future stoke this situation requires emergency assessment.. If all symptoms resolved within 24 hours then it will classified as a TIA. This project use various parameters gender , age, hypertension, heart disease, Blood pressure, ever married, work type, Residential type, average glucose level, BMI, smoking status this basic data is used to check whether the person has stroke or not, after he/she diagnosed with stroke then it will check how much sevier it is on the basis of the parameters such as Cystatin-c, Tau, MMP10 this is clinical data which is more related to brain blood cots. After that it will suggest which exercise should be done to reduce the possibility of brain stroke.

II. LITERATURE SURVEY

Existing literature was examined in order to gain the necessary knowledge about numerous ideas linked to the current investigation. The following s list of some of the significant conclusions reached through those.

1) Author[1]

In this paper authors discussed many problems related to stroke .Based on their similarities the review stdies were grouped in several categories. This paper targeted more than single study mentions research areas.

2) Author[2]

In this study they state tat how Machine learning with the help of pattern recognition becoming an important thing for the analysis, dealing, and prediction of difficulties and patient results in several neurological disease.

3) Author[3]

In this paper , The categorization into ischemic stroke(AIS) and non-AIS phenotypes is done from the brain MRI radiology on the basis of performance of natural language processing (NLP) and evaluated Machine algorithm.

4) Author[4]

They presented their work on Otsu Thresholding based SVM classifier for the detection of ischemic strokes. To train the model, 2 distinct kinds of brain MRI pictures in JPEG format were utilised as the dataset.

5) Author[8]

The goal of this paper is to calculate the probability of last 10 years stroke prediction and classify it's probability into 5 category. This paper is basically based on the biomedical field.

6) Author[9]

By taking help of Framingham Study cohort Help risk appraisal function has been developed. The probability of happening stroke with the help of Framingham Study cohort.

7) Author[10]

The study's paper sought to construct a model equation for an algorithm for predicting strokes using possibly adjustable risk variables.

8) Author[11]

In this study, back propagation neural network classification method is employed, together with decision tree algorithm, principal component analysis algorithm, and dimension reduction algorithm, to create a classification model.

9) Author[12]

Stroke MID means that the work which addresses the development of user interface or application interface to connect the data presentation and neurologist management to in a stroke cluster and prediction system .

10) Author[13]

This paper primary analysis algorithm is used to reduce dimensionality determine attributes that are more conducive to predating stroke disease, and predict whether a patient will suffer from Brain stroke.

11) Author[14]

In this article the current standard for initial imaging of patients with symptoms of Brain stroke is non-likeness CT scan of the head. The aim of this pare is for auto detection develop and validate set of deep learning algorithm.

III. PROPOSED SYSTEM

The proposed system serves as a predictive aid engine and will proves useful for user diagnosis. The algorithm which are used to diagnosis and find severity are much better than the existing one. On the basis of the severity we also suggest the exercise to overcome from the disease. The proposed system shortened the time required to put various collected data to practical use. The risk factor which are considered for the stroke and severity prediction are gender, age, hypertension, hart disease, average glucose level, CYSTATIN, MMP10, Tau these are some clinical terms which affects to stroke.

A. Advantages

1) Higher accuracy.
2) Higher Performance.
3) Easily available information is used for the prediction.
4) System provides the user different exercise for the user which are for reducing the risk of brain stroke on the basis of the severity.

This system is implemented using 5 Machine learning algorithm to get accurate results and accuracy.This project's Machine Learning model is developed using Logistic Regression algorithm, Support Vector Machine (SVM) algorithm, K Nearest Neighbour (KNN) algorithm, Decision Tree algorithm and finally Random Forest algorithm.

B. Front End Technology

HTML (Hyper Text Markup Language), CSS (Cascading StyleSheets) and Bootstrap.

C. Framework

Flask: Python API used for developing web-application.

D. Runtime Environment

Anaconda Environment: Anaconda allows you to perform different operations on environment such as copy, move, share and download. Specific collection of conda libraries and packages are collectively stored in a single environment. This allows them to maintain ad run independently.

The brain stroke Prediction Dataset has the total 5110 rows of data with 11 columns with attributes which are mentioned earlier.

The Severity prediction database is the another database that we used for our project with the 268 rows and 4 columns which are Cystatin_c, MMp10, Tau and class.

E. Libraries

Numpy, Pandas, Seaborn Sklearn/Scikitlearn.Pickle.Joblib, Matplotlib.

IV. FLOW CHART

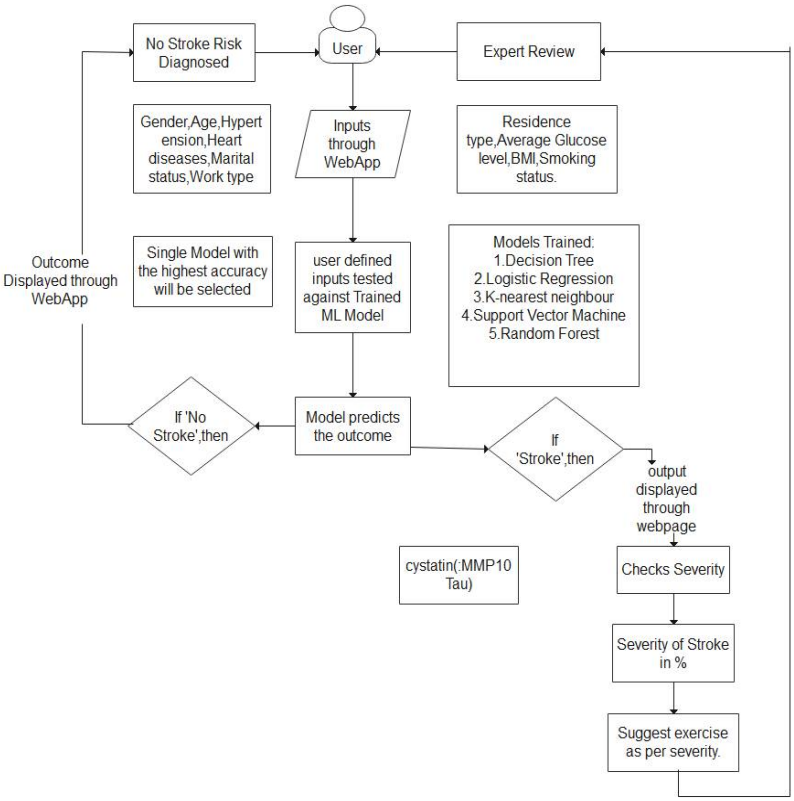


Fig 1. Flowchart of system

Detailed description of the flow chart:

1) *User:* The user is the person i.e. the end user who wants to know the stroke and its severity
2) *Accepting Inputs through WebApp:* The user will be requested to provide the specifics about their daily routine factor. Which are necessary to diagnosis of stroke.
3) *User Defined Inputs Tested Against ML Model:* The stroke possibility predicts against new data with the help of the 5 machine algorithm which are mentioned before these 5 machine learning algorithm are trained such that it will gives the best accuracy result which will be considered as the trained ML model which will help for prediction. Machine Learning Algorithms which are implemented are Decision Tree, Logistic Regression, K-Nearest Neighbour, Support Vector Machine and Random Forest.
4) *Outcome Forecasts Through Various Algorithms:* The trained Ml model is used to predict whether the person has the stroke or not and if it gives the result '1' the stroke risk is predicted and else if result is '0' then there is no any stroke risk.
5) *No Stroke Risk Diagnosed:* If the prediction model resulted nto '0' then the person does not have stroke so that "NO STROKE" is dislayed on the screen.
6) *Stroke Risk Diagnosed:* The user will get know through our web application that it has risk of stroke. If the input values are matched with the trained dataset values which will result into '1' then it will displays "Stroke Risk Diagnosed". With the consequences of modules which are predict the stroke of user working of web application is explained.

The module are:

a) To identify the stroke risk we must need to take some inputs from the user through web application. This input contains some basic information about the user's health and lifestyle. This input is taken to check it against the trained input data.
b) Processing training data against the input data The trained data is processed against the data which has been collected from user. and, as was stated in the further section of previous lesson, get the correct result at end.
c) Receiving the test results.The customer will receive exact and precise result from our web application as the last stage, enabling them to proceed as needed in according to what we found.

7) *Factors Affecting to Stroke:* After the Stroke risk get diagnosed it will also shows the factors which are affecting to the stroke depends on the database. It will checks the outliers for the each field ad compares if any input field value is grater than the normal one and which leads to Stroke it will get displayed on screen.
8) *Check Severity:* After the stroke get diagnosed and got the factors which will affect to stroke the user is able to check the severity of stroke y taking more clinical data from user. Which contains the parameter like Tau, CYSTATIN, MMP10. These parameters are explained below.
9) *Suggest Exercise:* When user gets the Severity of stroke it will be in the % form. On the basis of how much sever it is our application suggest them some exercise which will become beneficial to overcome from the stroke or to reduce the chances.

Let's have a look on a data which decides the severity of Stroke:

A. Cystatin

Cystatin C is a kind of protein that is produced by the cell in your body. Cystatin C has been proposed as a more accurate indicator of renal function than the creatinine-based formulation reason behind this is that the blood concentration being independent muscle mass and appearing to be unaffected by age , Cystatin C. Renal dysfunction appears to be the most logical process linking inflammation, atherosclerosis, and vascular risk factors to high amount of cystatin C's documented favourable association with an elevated risk of stroke, cardiovascular morbidity, and death.

B. TAU

Tau is a protein mainly expressed in adult brains. In neurodegenerative disease and stroke it plays very important role. In all cases ischemic stroke almost account 80%. Stroke becomes important cause behind the adult death and disability. While tau protein's main physiological role is to maintain microtubule networks in neurons, the situation known as hyperpsphorylation will greatly diminish tau protein's biological activity. One of the harmful protein that builds in neurons and the cerebrovascular after he ischemia is dysfunctional tau. It is also closely linked to a variety of degenerative changes associated with ischemic stroke.

C. MMP 10

In translation neurology Ischemic stroke continues to be the most challenging disease. Matrix metalloproteinases, also called matrixins, cleaves proteins using a preserved process that includes the actication of site-bound water molecule by a Zn2+ ion. From 23 different proteases Human MMPs are made. Excreted MMps can be divided into four classes based on how specific they are to theirbase on which organism live: collagenases(MMP-1,-8, and -13), gelatinses(MMP -1, -8, and -13), gelatinases(MMP-2 and -9), stromelysins(MP-3, -10, and -11) and a heterogeneous group that include matrilysin(MMP-7), metallo-elastase (MMP-12), enamelysin (MMP-2 MMP-28)

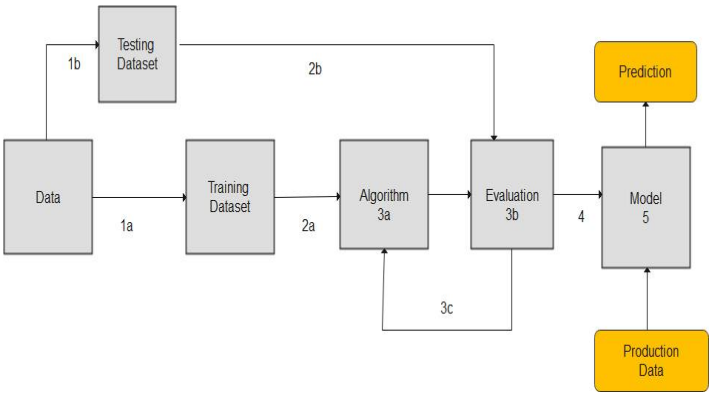


Fig 2. Work Flow

The work flow of the ML project is given below:

1) Firstly we have to clean the values i.e. deleting missing values and from training and testing data unwanted content is removed.
2) Label encoder is applied to encode to convert objects into integer apply.
3) Further data is split into the test data and Training data.
4) ML Models are trained using following algorithms:
a) Support Vector Machine(SVM) Algorithm.
b) Random Forest Algorithm.
c) Decision Tree Algorithm
d) KNN(K-nearest neighbour) Algorithm.
e) Logistic Regression.

5) Calculate the accuracy score for each model.
6) The model is selected which has the highest accuracy.
7) Crete GUI then Extract a model into that GUI.
8) Enter the input for the given parameters.
9) Result: - Predict data with respect to similar model.

Afterwards applying modules, algorithm and codes expected outcome yields. The homepage will assist users in entering the information needed for the connected stroke prognosis system, and the GUI portion is designed to be user-friendly for regular people. Knowing the results you want to achieve will help you get there more quickly. Using our dataset, the system has been developed using 5 distinct ML method, as stated in the implementation.

The following Results have been generated:

• During the process we got the algorithm which gives the lowest accuracy is Decision Tree i.e. (76.9%).
• During the process we got the algorithm which gives the highest accuracy is Logistic regression i.e. (98.56 %).

With the help of Matplotlib library in Python all trained models accuracy is shown in below image.
So, Random forest generation model is imported for testing compared to given data that is shown in Figure 3.

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 V May 2023- Available at www.ijraset.com

```
import joblib
model_path=os.path.join('D:/Python37/Projects/Stroke Prediction/','models/dt.sav')
joblib.dump(dt,model_path)

[523]

['D:/Python37/Projects/Stroke Prediction/models/dt.sav']
```

Fig 3. Exporting Trained Model

So, the proposed system help us to analyse which algorithm gives more accuracy to get the the accurate results and also which parameters should be considered to predict if there is possibility of Brin stroke or not and also give the factors which are affecting to it,. This system also predict how much savvier this stroke becomes and what precautions should be take t reduce the risk of Brain stroke.

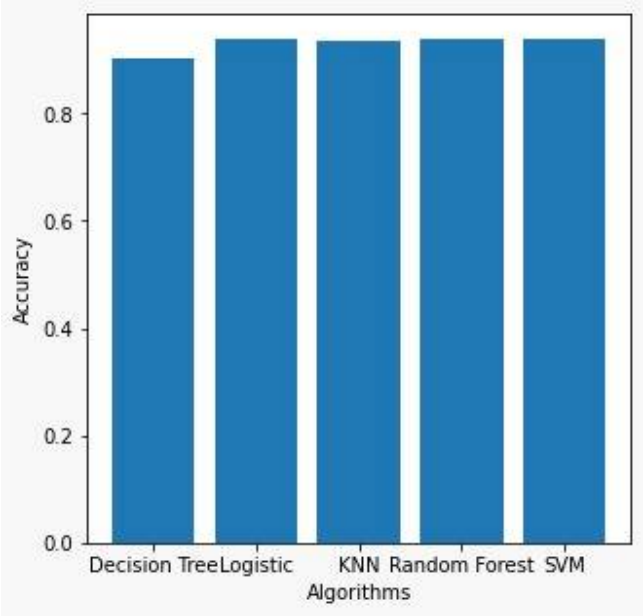


Figure. 4 – Trained Models Accuracy Score

V. OUTPUT

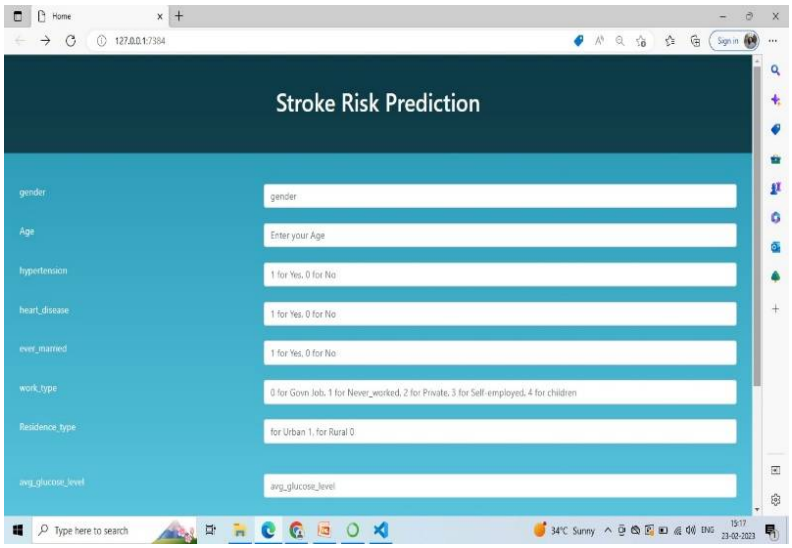


Figure5-Stroke Risk Prediction

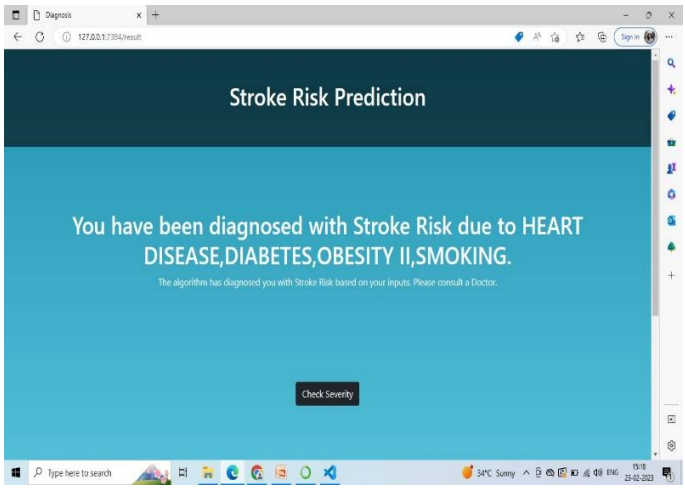


Figure 6 - For Stroke

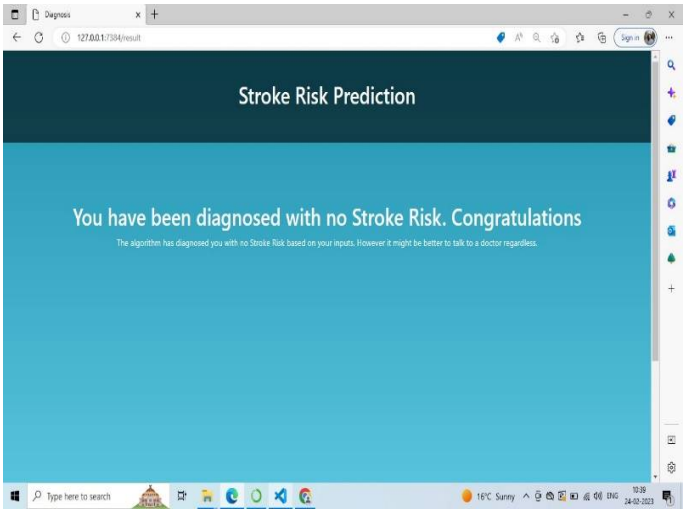


Figure7-For No Stroke

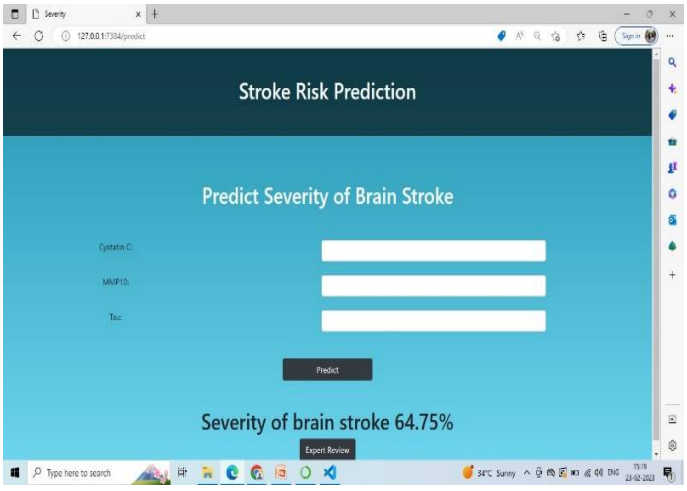


Figure8 – For Severity Prediction

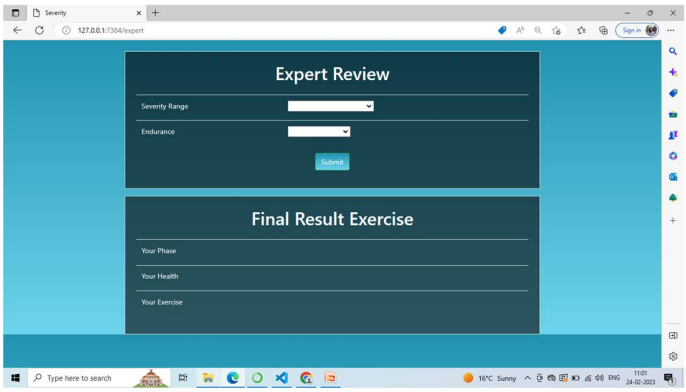


Figure 9- For Taking input to suggest exercise

VI. CONCLUSION

As we seen in literature survey every system has it's own pros and cons thus proposed a system that assistances to predict brain stroke efficiently and easily. Stroke prone patients were identified with acceptable accuracy using several assessments and prediction models. Hence this project helps to predict the whether the person has stroke or not and if person has stroke it also gives how much savvier it could be and also suggest some exercise to reduce the chances of stroke. This system the saves the time of the patients and also gives the result based on the clinical data which was provided by that user. This encourages medical user to become more motivated to control their health and to alter their health related habits.

VII. FUTURE SCOPE

The future enhancement that can be done in this project that can be:

1) The correctness of model will be increased by applying more powerful algorithms.
2) It can be possible to provide more information regarding to the brain stroke
3) For more improvements one can add visualizing results for more and effective understanding.
4) Stroke can be predict with help of more clinical data like MRI, CT Scan, EEG etc.,

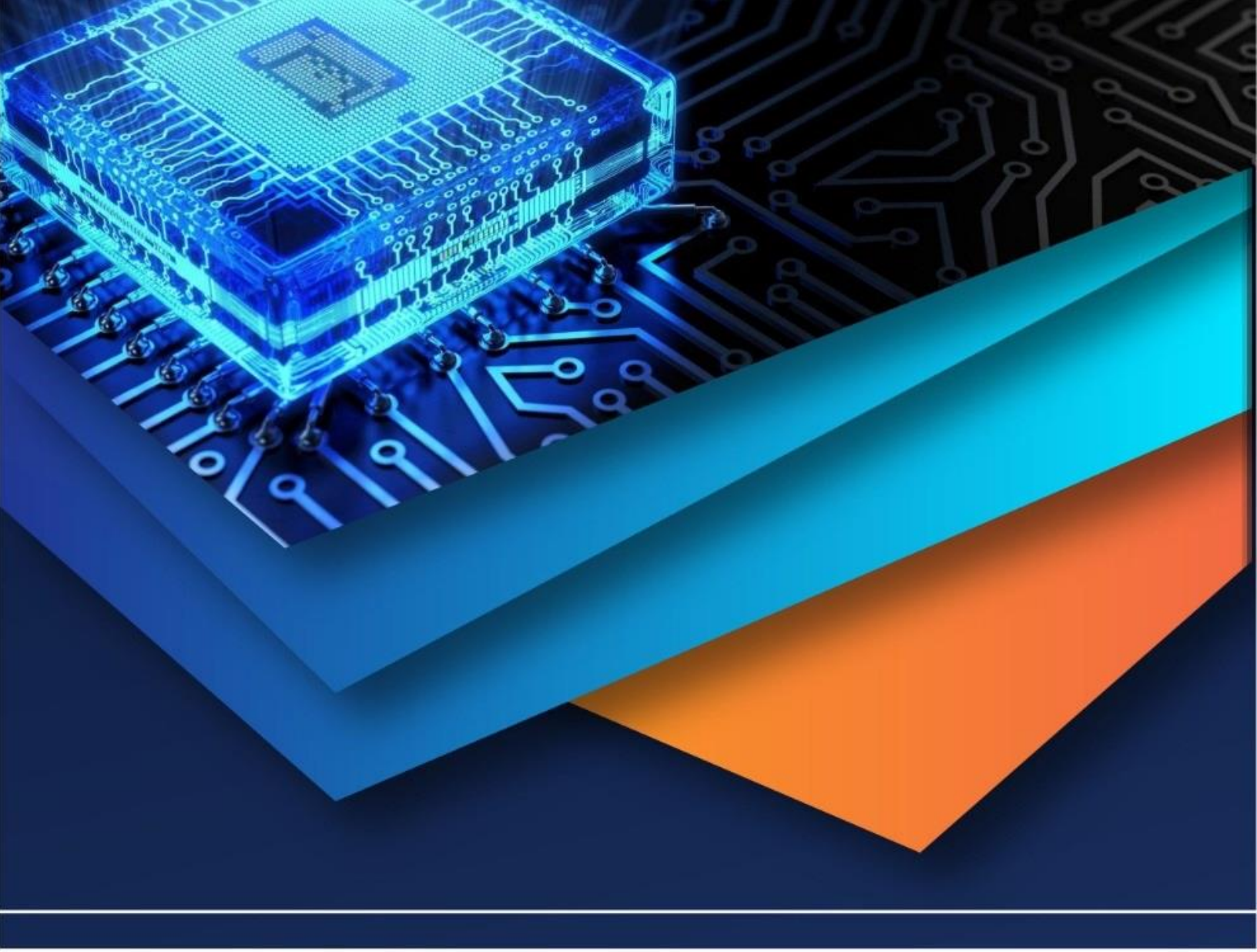
VIII. ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to the individuals and organisations who provided invaluable assistance and contributions to the project. Our project guide Prof Deepali Deshpande for her expert guidance, and unwavering support throughout the project. Staff at Artificial Intelligence and data Science Engineering Department at Vishwakarma Institute of Technology for providing us with the necessary resources and facilities to assembling this project. Our team members who worked tirelessly on this project, and whose unique skills and perspectives greatly contributed to its success. Without the support of all these individuals, this project would not have been possible.

REFERENCES

[1] Manish Sirsat Eduardo Ferme, Joana Camara, "Machine Learning for Brain stroke: A Review, " Journal of stroke and cerebrovascular disease: the official journal of National Stroke Association(JSTROKECEREBROVASDIS), 20220
[2] Harish Kamal, Victor Lopez, Sunil A. Sheth, "Machin e Learning in Acute Ischemic Stroke Neuroimaging, " Frontiers in Neurology (FNEUR) 2018.
[3] Chuloh Kim, Vivienne Zhu, Jihad Obeid and Leslie Lanert, "Natural Language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke , "public Library of Science One (PONE) 2019
[4] R.P. Lakshmi , M.S. Bbu and V.Vijayalakshmi "Voxel based lesion segmentation through SVM classifier for effective brain stroke detection," International Conference on wireless Communications, Signal Processing and Networking (WiSPNET), 2017
[5] J.Yu et al, "Sementic Analysis of NIH Stroke Scale using Machine Learning Techniques" International Conference on platform technology and service(PlatCon), 2019
[6] Gangavarapu Sailasys and Gorli L Aruna Kumari "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms, "International Journal of Advanced Computer Science and application (IJACSA), 2021
[7] "Stroke Prediction Dataset", KAggle.com 2021 https://www.kaggle.com/fedesoriano/strokeprediction- dataset. Accessed 6 Oct 2021.
[8] "Computer Methods and Programs in the Biomedical"- Jae-woo Lee, Hyun-sun lim, Dong-wook kim, Soon-ae Shin.
[9] "Probability of Stroke: A risk Profile from the Framingham Study"- Philip A. Wolf, MD; Ralph B.D' Agostino, PhD, Albert J.Belanger, MA; and William B. Kannel, MD.

[10] "Development algorithm for Stroke Prediction: A national Health Insurance Database Study" – Mn SN, Park SJ,Kim DJ, Subraniyam M,Lee KS
[11] "Stroke Prediction using Artificial Intelligence"- M.Sheetal Singh, Prakash Choudhary.
[12] "Medical Software user interfaces, stroke MD ppliction design(IEEE)" Elena Zamsa.
[13] "Effective analysis and predictive Model of Stroke Disease using classification methods."-A.Sudha, P.Gayathri, N.Jaisankar.
[14] "Deep Learning alohorithm for detection of critical findings in head for CT scan: a retrospective study"- Rohit Ghosh, Sweetha Tanamla, Mustafa Bivaji, Norbert G Campeau, Vasantha Kumar Vanugopal.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY