PREDICTIVE AND DETAILED ANALYSIS OF BRAIN STROKE USING DEEP LEARNING

[1] B. Divya Sri, [2] Dr. V. Usha Bala, [3] R. Udaya Lakshmi, [4] B. Charitha, [5]K. Madhu Mita,[6] D. Harshith
[1,3,4,5,6] BTech student, [2] Associate Professor

[1,2,3,4,5] Computer Science and Engineering, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, India

I. TITLE

The project Entitled as "Predictive and Detailed analysis of Brain Stroke" which falls under the "Healthcare" category, examining patient medical information performed at various hospitals.

II. ABSTRACT

***Abstract*—According to the World Health Organization (WHO), stroke is the second leading cause of death worldwide, accounting for about 11Stroke is a cerebro-vascular disease that affects the normal blood supply to the brain. Without proper supervision, it can lead to death or long-term disability. Stroke is classified into two types that are ischemic or hemorrhagic. Ischemic and hemorrhagic strokes can occur simultaneously. A stroke caused by a blood clot in a blood vessel is called an ischemic stroke, and a ruptured blood vessel is called a hemorrhagic stroke. Hypertension is the most important modifiable risk factor for stroke, but its contribution varies by subtype. Ischemic is the lack of oxygen in the cranial nervous system that ultimately leads to death. About 85% the ischemic category and are caused primarily by small vessel atherosclerosis, cardiac embolism, and aortic other thromboembolism. Ischemic stroke in young patients can occur for a variety of reasons such as extracranial dissection. The outcome for these patients is determined by the severity of the symptoms. MRI provides important physiological information through MR angiography and diffusion MRI. They are of high clinical value in identifying patients in need of immediate treatment. In current clinical practice, the time since stroke onset dominates the physiological information provided by MRI when making treatment decisions. This will change until clinical trials demonstrate that stroke physiology as revealed by MRI is superior to time from stroke onset in promoting favorable clinical outcomes. Its frequency is accelerating in developing countries like India due to unhealthy lifestyles. This guides appropriate investigation, and secondary prevention in localizing strokes and understanding their mechanisms.**

KEYWORDS: Cerebrovascular disease, intracerebral haemorrhage stroke, ischaemic stroke, stroke pathogenesis, stroke risk factors, transient ischaemic attack, MR angiography, diffused MRI, perfusion MRI.

III. INTRODUCTION

The project "Predictive and detailed analysis of brain stroke" in the healthcare category aims to utilize patient medical information obtained from diverse hospitals to detect the emergence of brain stroke. The project employs MRI reports or symptoms or lifestyle habits of the patient as input to generate a comprehensive and precise report on the identified stroke. Furthermore, the project presents the nature of detected stroke.

The objective of the project is to utilize methodologies to enhance stroke detection and offer workable insights to mitigate the vulnerability and intensity of strokes in patients. By using patient medical information, the project can aid healthcare professionals in making informed decisions and providing superior care to their patients. However, the decision to initiate treatment in the event of stroke onset is usually time-dependent. This is due to the fact that the clinical trials, which shaped the current stroke treatment protocols, relied on time and non-contrast CT scans for patient enrollment. Any changes in this approach will only emerge after clinical trials demonstrate the utility of MRI in identifying patients who can derive benefits from treatment.

IV. NEED OF STUDY

A. Background Study

The primary objective of the study is to promote awareness of stroke, a genetic origin that arises from the disruption or cessation of blood supply to the brain. This leads to insufficient oxygen and nutrient supply to the brain, resulting in impaired functioning and damage. Raising awareness about this ailment is crucial as early identification and timely treatment are essential for ensuring safe recovery from illness and preventing death, and the prediction and determination of a stroke event. Stroke, as defined by a focal neurological deficit of sudden onset, is characterized by symptoms that persist for more than 24 hours, or result in death before 24 hours. Currently, the available stroke detection applications are only capable of identifying stroke onset based on speech patterns. However, these definitions are no longer useful in clinical practice due to the time-sensitivity of stroke treatment, which needs to be initiated promptly after diagnosis.

V. OBJECTIVE OF THE STUDY

A. PRIMARY OBJECTIVE

The primary aim of this project is to detect the occurrence of based on the MRI scan and life styles. Multiclass Classification is used along with ML supervised Classification algorithms. The main purpose is to detect the occurrence of stroke using life styles or MRI scan of user.

B. SECONDARY OBJECTIVE

The purpose is to determine the type of stroke (if detected) by using a supervised ML classification algorithm that is SVM, Decision Tree and CNN for detecting occurrence of stroke.

VI. SYSTEM ANALYSIS

A. EXISTING SYSTEM

- Mircea Gurbin, Mihaela Lascu, and Dan Lascu et al. proposed a method consisting of Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT) and Support Vector Machine (SVM). It uses different levels of wavelets, and by training, the cancerous and

non-cancerous tumours can be identified. The computation time is longer for the proposed method.

- Somasundaram S. and Gobinath R. et al. explains the present status of detection and segmentation of tumour through deep learning models. For deeper segmentation, 3D based CNN, ANN and SVM is used.
- Damodharan S. and Raghavan D. et al. address segmentation of pathological tissues (Tumor), normal tissues (White Matter (WM) and Gray Matter (GM)) and fluid (Cerebrospinal Fluid (CSF)), extraction of the relevant features from each segmented tissues and classification of the tumor images with Neural Network (NN).

B. DISADVANTAGES OF EXISTING SYSTEM

- Owing to the small size of tumors compared to the rest of the brain, brain imaging data are imbalanced. Due to this characterization, existing networks get to be biased towards the one class that is overrepresented.
- Less accuracy was achieved.
- Higher false rates.
- Takes more time than the usual time.
- Complex in implementing real-time applications.

C. PROPOSED SYSTEM

- The aim of this system is to build a system that would help in brain tumour detection from MRI images through the convolution neural network. The proposed method was tested and compared with the existing classification techniques to determine the accuracy of the proposed method.
- Deep learning is a machine learning technique that instructs computers what to do as a human think and do in a scenario. In deep learning, a computer model is able to do classification tasks from images, sound or text. Sometimes human level performance is being exceeded by deep learning techniques. One of the most popular neural networks is an artificial neural network that has a collection of simulated neurons. Each neuron acts as a node and by links each node is connected to other nodes.
- The proposed system consists of image acquisition, preprocessing, segmentation, feature extraction and classification.
- This algorithm is faster in execution for normal MRI images.

D. ADVANTAGES OF PROPOSED SYSTEM

- Deep learning techniques will be effective here to classify tumor image without segmentation.
- In deep learning, the feature is extracted from the entire image automatically. Convolution in the CNN architecture performs this operation. Number of feature maps increases with the increase in CONV layer.
- Reduction of dimension is required to initiate training. Pooling layer down samples the feature dimension. Fully connected layers manipulate the score of each label. Softmax layers prepare the model with feature and class score.
- CNN is proved for better accuracies with supporting to the deep learning methods. It is also complemented with the light weight library in python for image processing which help us to classify the image and improves the speed of execution. System has used various parameters for classification between normal and tumorous brain [7].

VII. PROPOSED SYSTEM (FOR SYMPTOMS)

The proposed system aims to detect the occurrence of brain stroke. The input for this proposed system is either brain MRI scans or lifestyles and health related information of the users like BMI, age, gender, marital status, type of workplace and many more. The existing systems detect the occurrence of stroke through speech patterns but this system detects the occurrence of stroke through MRI scans. This system also detects the type of stroke which is the main spotlight of the proposed system.

VIII. INPUT DESIGN AND OUTPUT DESIGN

A. INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data into a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps, and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

B. OBJECTIVES FOR INPUT DESIGN

- Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
- It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
- When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate

messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

C. OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

D. OBJECTIVES FOR OUTPUT DESIGN

- Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
- Select methods for presenting information.
- Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

E. PROPOSED ARCHITECTURE:

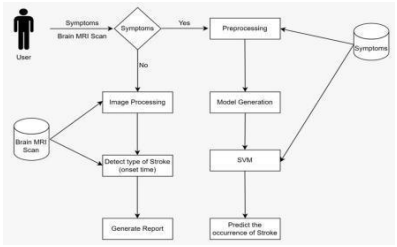


Fig. 1.*Proposed Archietecture*

- Preprocessing Methods For Symptoms: Preprocessing methods involve the techniques used for filtering the data and filling the missing values. These techniques can significantly affect the accuracy of the project.
- Data cleaning: The null values are identified using appropriate methods from the pandas library. Then the null values are filled with the average value of the column.
- Feature selection: The most important features that affect the result are identified and used for analysis. For example, Name of the user does not affect the result and hence can be ignored. In this way the important features are identified.
- Outliers: Outliers are identified using boxplot and these are handled using machine learning algorithms like SVM.
- Normalization: Normalization is mainly performed to transform the feature to a similar scale and distribution. Here sklearn is used for normalization which uses several techniques to normalize the data like z-score, min-max normalization.

IX. METHODOLOGY

A. Techniques tools:

- SVM: Support vector machines, or SVMs, are one of the most popular supervised learning algorithms used for both classification and regression problems. However, it is mainly used for machine learning classification problems. The goal of the SVM algorithm is to create optimal lines or decision boundaries that can divide the n-dimensional space into classes so that new data points can be easily placed in the correct category in the future. This optimal decision boundary is called a hyperplane. SVM chooses extrema/vectors to help create hyperplanes. These extreme cases are called support vectors, and the algorithm is called a support vector machine. Consider the following figure, which has two different categories classified by decision boundaries or hyperplanes. After the preprocessing and normalization of data then the SVM model is generated and trained with the labelled data. This helps the model to learn the detection based on trained data. So training data plays a vital role in performance of model. SVM is majorly used for classification of data. This is the best algorithm for detecting stroke or no stroke. This detects the occurrence of stroke based on the distance between the new input and existing features that are used for training.
- Decision tree: This model works by recursively splitting the data into subsets based on the values of the features until a final decision or prediction is made. The final decision is made at the leaf nodes of the tree, which represent the predicted outcome.
- CNN: In a convolutional network, there are basically three types of layers:
  1.Convolution layer 2.Pooling layer 3.Fully connected layer
- Numpy: Numpy is a widely used Python library for scientific computing and data analysis. It provides powerful tools for working with arrays, numerical computing, and mathematical operations.we used the numpy module for working with arrays.
- OS Module: The os module in Python provides a way of using operating system dependent functionality like reading or writing to the file system, interacting with the file system. We used OS module to retrieve the images dataset from the directory.

- Keras: Keras is a high-level neural networks API, written in Python and capable of running on top of various lower-level deep learning frameworks, such as TensorFlow and Theano. It provides a user-friendly interface to build, train, and evaluate deep learning models, making it easier for developers to create and experiment with different neural network architectures. We have used Keras module to build the model called sequential model and add the layers like Conv2D,Flatten,Dense,MaxPooling2D,Dropout to the CNN model.Keras is also used for saving the model.
- Cv2: OpenCV (cv2 for short) is a popular computer vision library for Python. It provides a wide range of image and video processing algorithms and tools for feature extraction, object detection, and more.
- sklearn.metrics: sklearn.metrics is a module in scikit-learn that provides various metrics for evaluating the performance of a machine learning model.

X. MODULES

A. Data acquisition

In the first module, we developed the system to get the input dataset for the training and testing purpose.We have taken the dataset brain stroke MRI Scans Image dataset consists of three classes:

- 1.No stroke
- 2.Ischemic stroke
- 3.Hemorrhagic stroke

First, we have to load the Symptoms csv file to an array. The dataset consists ofattributes like age, gender, hypertension, work type, BMI a with their respective values Stoke or No Stroke.

B. Importing the necessary libraries

We are using Python language for this. First we are importing the necessary libraries such as keras for building the main model, sklearn for splitting the training and test data, PIL for converting the images into array of numbers,CV2 for resize the image, Flask for backend connection and other libraries such as pandas, numpy ,matplotlib and tensorflow.

C. Data Preprocessing and Feature Extraction

We retrieved the images and their labels. Then resize the images to (150,150) as all images should have the same size for recognition. Then converted the images into a numpy array. As the Symptoms data has some missing values,we filled the missing values with mean of the column values and Removed the Outliers to make the dataset more efficient,Label Encoding: allocated labels as 0 for no Stroke and 1 for Stroke.

D. Model Building

Split the dataset into train and test. 80building the model we used a sequential model from keras library. Then we added the layers to make a convolutional neural network. In the first 2 Conv2D layers we have used 32 filters and the kernel size is (3,3) for one and 64 filters for other. In the MaxPool2D layer we have kept pool size (2,2) which means it will select the maximum value of every 2 x 2 area of the image. By doing this dimensions of the image will reduce by factor of 2. In dropout layer we have kept dropout rate = 0.25 that means 25layers again with some change in parameters. Then we apply a flatten layer to convert 2-D data to 1-D vector. This layer is followed by dense layer, dropout layer and dense layer again. The last dense layer outputs 2 nodes as the Type of Stroke or No Stroke This layer uses the softmax activation function which gives probability value and predicts which of the 3 options has the highest probability. From the Sklearn package,we imported the Decision tree model and SVM model for training the model.

E. Model Training

We compiled the model and applied it using the fit function. The batch size is 2. We set the epoch value to 20. Then we plotted the graphs for accuracy and loss. We applied the data for the models using fit function

F. Model Evaluation

By importing SKlearn Module, we calculated the Accuracy, Precision, Recall.These are the commonly used metrics to evaluate the performance of a classification model in machine learning.

G. Saving the Trained Model

Once we're confident enough to take the trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle or keras.We can import the module and dump the model into .h5 By importing SKlearn Module, we calculated the Accuracy,Precision,Recall.These are the commonly used metrics to evaluate the performance of a classification model In machine learning.

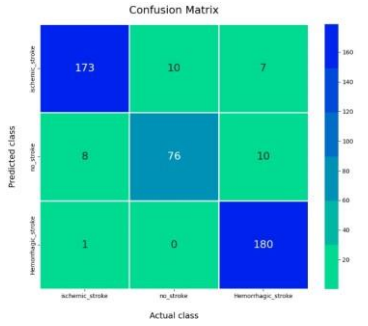
XI. IMPLEMENTATION:

A. For symptoms:

The input to this module is the lifestyle habits of users like BMI, Glucose levels, past heart stroke history, age, Work environment and living area like urban or rural. The dataset used for the model generation is taken from kaggle and the data is preprocessed inorder to increase the accuracy of the model. The model is generated using the training data and the model is downloaded to the local system. This model is generated using SVM and decision tree. The input data is then given to the generated model and the result page is displayed based on the result. Here we used SVM and Decision tree algorithm for training the model. Two models are used here in order to increase the accuracy of the system.

B. For MRI:

The input for the model are the MRI Scans retrieved from the Directories using OS Module.The images dataset is chosen from the Kaggle.The Preprocessing like Converting image to Binary Image,resize the image to a specific size is done by using CV2 module. The CNN model is build using the Sequential model and added the Conv2D,Flatten,Dense,MaxPooling2D,Dropout layers to the model Using the the Keras module on top of tensorflow. Splitting the data as 20model with 20 epoch for improving efficiency of the model.Once the model is done with compiling we saved that model in the H5 file model using Keras module.



	precision	recall	f1-score	support
ischemic_stroke	0.95	0.91	0.93	190
no_stroke	0.88	0.81	0.84	94
Hemorrhagic_stroke	0.91	0.99	0.95	181
accuracy			0.92	465
macro avg	0.92	0.90	0.91	465
weighted avg	0.92	0.92	0.92	465

Fig-2 [a] Confusion Matrix

Fig-2[b] Accuracy Analysis

XII. CONCLUSION

This paper provides a new method to detect brain tumours using deep learning methods. Early detection of stroke contributes to timely and effective treatment. In current clinical practice, time overrides physiology when making decisions about treatment delivery. There is evidence to suggest that, over time, physiology revealed by MRI may represent improved clinical decisions regarding the management of severe stroke patients. However, this requires clinical trials demonstrating the value of using MRI to improve outcomes when used to select patients for treatment. Kaggle's dataset contains high-quality MRI images for research purposes. Various segmentation algorithms were tested. From this, a convolutional neural network using the modified approach helped to obtain results with 92interface support. It can also identify evidence of various diseases from MRI images.

XIII. FUTURE WORK :

- 1.This system can be improvised by adding the algorithms for detecting the BMI and the GLUCOSE LEVELS of the user when the MRI scans are given.
- 2.This can also further be improved to detect the onset time of stroke when a stroke is detected.
- 3. This can further be improved to predict whether thrombolysis is advisable or not.
- 4. This can further be extended by adding security and payment sections.

REFERENCES

[1] R. Lakshmi, M. S. Babu, and V. Vijayalakshmi, " Voxel based lesion segmentation through SVM classifier for effective brain stroke detection," Proceedings of International Conference on Wireless Communications, Signal Processing

[2] S. Keerthana and K. Sathiyakumari, "Brain Strokes Prediction using fuzzy C-means Clustering" International Journal of Computer Applications, vol. 154, no. 4 , pp. 26-30, 2016

[3] Y.-S. Tyan, M.-C. Wu, C.-L. Chin, Y.-L. Kuo, M.-S. Lee, and H.-Y. Chang,"Ischemic stroke detection system with a computer-aided diagnostic ability using an unsupervised feature perception enhancement method," International journal of biomedical imaging, vol. 2014, pp. 1-12, 2014.

[4] A. A. Mamun and M. S. Hossain, "Ulcer Detection in Image Converted from Video Footage of Wireless Capsule Endoscopy," Proceedings of 1 st International Conference on Advances in Science, Engineering and Robotics Technology, 3-5 May 2019, Dhaka, Bangladesh, pp. 1-4

[5] A. Subudhi, M. Dash, and S. Sabut, " Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier," Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 277-289, 2020

[6] Li, Shutao, JT-Y. Kwok, IW-H. Tsang and Yaonan Wang. "Fusing images with different focuses using support vector machines." IEEE Transactions on neural networks 15, no. 6 (2004): 1555-1561.

[7] Vaishali et al. (2015) Wavelet-based feature extraction for brain tumor diagnosis—a survey. Int J Res Appl Sci Eng Technol (IJRASET) 3(V), ISSN: 2321- 9653 .

[8] S. Roy And S.K.Bandyopadhyay, "Detection And Qualification Of Brain Tumor From MRI Of Brain And Symmetric Analysis," International Journal Of Information And Communication Technology Research, Volume 2 No.6, June 2012, Pp584-588.

[9] R. Noureddine, K. Tarhini, and S. Saleh, " Segmentation and extraction of brain injury lesions from MRI images: Matlab implementation," Proceedings of International Conference on Advances in Biomedical Engineering, 16-18 September 2015, Beirut, Lebanon, pp. 45-48

[10] Pradeepa, S., Manjula, K. R., Vimal, S., Khan, M. S., Chilamkurti, N., Luhach, A. K.: DRFS: Detecting Risk Factor of Stroke Disease from Social Media Using Machine Learning Techniques. In Springer (2020)

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

[12] Nwosu, C.S., Dev, S., Bhardwaj, P., Veeravalli, B., John, D.: Predicting stroke from electronic health records. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE (2019).

[13] Fahd Saleh Alotaibi: Implementation of Machine Learning Model to Predict Heart Failure Disease. In: International Journal of Advanced Computer Science and Applications (IJACSA) (2019).

[14] Shanthi, D., Sahoo, G., Saravanan, N.: Designing an artificial neural network model for the prediction of thrombo-embolic stroke. Int. J. Biometric Bioinform. (IJBB) (2009).

[15] Singh, M.S., Choudhary, P., Thongam, K.: A comparative analysis for various stroke prediction techniques. In: Springer, Singapore (2020)