International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Brain Stroke Prediction Using Random Forest Algorithm

M.N.V.S.S. Abhiram[1], T. Sudheer[2], Y. Shanmuk Ram Tej[3], A.N.V.K. Swarupa[4]

[1,2,3]Students, Department of Computer Science and Engineering, Sasi Institute of Technology & Engineering, Tadepalligudem
[4]Assistant Professor, Department of Computer Science and Engineering, Sasi Institute of Technology & Engineering, Tadepalligudem
[1]maddula.abhiram@sasi.ac.in, [2]sudheer.thota@sasi.ac.in, [3]shanmuk.yepuganti@sasi.ac.in, [4]swarupa508@sasi.ac.in

Abstract

Brain stroke is one of the most prevalent reasons for disability and death worldwide, and thus the early detection of brain stroke is crucial for treating it and enhancing patient outcomes. The current research work utilizes recent deep learning and machine learning approaches for stroke predictive accuracy from Computed Tomography (CT) scans. A VGG16 CNN-based model is used for feature extraction due to its robust ability to learn important patterns in medical images. The features are then classified using Random Forest (RF) Algorithm. This model is trained and cross-validated with a large dataset, comparing predictive accuracy among models. Our experiments indicate that Random Forest provides the highest accuracy and performs better compared to other boosting techniques. The current research provides a novel method of stroke prediction by combining deep learning and traditional machine learning classifiers, which provides evidence of the potential of AI-based approaches for medical diagnosis and decision-making. The current research work results in quicker and accurate stroke detection, which will eventually contribute to early medical treatment.

Keywords: Stroke Prediction, CT images, VGG16, Random Forest Algorithm

1. Introduction

Stroke is a cardiovascular disease that occurs when blood flow to the brain becomes abnormal. It is a leading cause of disability and death. There are different types of strokes: ischemic and haemorrhagic. The first, which occurs when a blood vessel in the brain becomes blocked, is the most common and is the most dangerous. The second, which occurs when there is bleeding, is the most uncommon. There are specific treatments that should be used immediately after a stroke, and in both cases, a mistake can be fatal for the patient. Currently, doctors use magnetic resonance imaging (MRI) and computed tomography (CT), two well-established imaging methods, to pinpoint the location and type of a stroke. These methods are effective, but the imaging systems that go along with them are time-consuming, costly, not portable, and, in the case of CT scanning, dangerous. As a result, a variety of imaging methods have been investigated over the past few years, one of which is microwave imaging (MWI), which holds promise for imaging strokes. The general concept of IMV is that at microwave frequencies there is a dielectric difference between healthy brain tissue and the site of an ischemic or haemorrhagic stroke. Specifically, the dielectric constant of an ischemic stroke is lower than the dielectric constant of the surrounding brain tissue; in contrast, the dielectric constant of a haemorrhagic stroke (i.e., the dielectric constant of blood) is higher than the dielectric constant of brain tissue.

On top of the MVI device, there is a collection of antennas, and typically each antenna serves as both a transmitter and a receiver. A visualization algorithm is used to process the obtained images to identify the stroke's location and type. MWI gadgets use low-power microwave radiation, so they can be transported cheaply and safely. However, stopping the stroke type in real time is difficult for imaging algorithms that process measured data. Machine learning (ML) algorithms are an innovative method currently emerging in the world of biomedical imaging: they greatly reduce statistical processing time and constitute a promising alternative to deterministic imaging algorithms. To cut down on intervention time, a special microwave device that uses ML-based class of stroke type detection methods at the prehospital level has been proposed. Salucci et al. instead A machine learning algorithm was used to classify strokes step by step using a traditional diagnostic approach (i.e., detection, identification, and localization). Vectors of support. A complex network approach was used to sort stroke types in a small microwave diagnostic device for internal bleeding that uses deep neural network inference.

However, one major drawback of using machine learning (ML) algorithms for stroke classification is that ML algorithms can only be properly trained with a very large dataset. The dataset can be created using clinical facts, laboratory measurements with anthropometric head phantoms, or electromagnetic (EM) modelling of the entire machine. However, collecting a large number of clinical facts is complex and requires long runtimes for more than one measurement with unusual head conditions (health and stroke) and full-wave models. Brain symmetry, which was also used as a kind of signal correlation in the transmission path without the need for a training phase, was used in an unsupervised classification method to solve this issue.

The aim of this paper is to propose an alternative and efficient method based on a fully decomposed generalized approximation and linearization of the growth operator to generate the training dataset. The generation time of the following dataset for each major country is almost four orders of magnitude faster than that of typical full-wave simulations of the entire MWI apparatus, and the computational requirements are significantly lower (c, or seconds as opposed to hours). Support Vector Machine (SVM), Multilayer Perceptron (MLP) from Circle of Relatives, and k-Nearest Friends Neural Network (k-NN) algorithms were trained using the proposed approach. To the authors' knowledge, this is the first time that a k-NN rule set has been applied to stroke detection using a microwave imaging approach. Statistics from full-wave simulations of the entire MWI system in a three-dimensional anthropometric multi-tissue head model were used to study MO. By making a small change in the antenna geometry (possibly in the antenna implementation section), more experimental data were obtained and the information scattering became less, i.e., the phase was reduced. The tested ML algorithms were able to accurately identify the type, location, and presence of the brain in the majority of cases. Multi-tissue head models from real and closed training and testing sets were also used to test the algorithms' ability to effectively generalize and classify models with different head models. In this case, both the SVM and MLP algorithms achieved high-quality results; in fact, they were able to classify samples from the validation set that corresponded to head samples that were not included in the training set. The single-head version was used in recent preliminary analyses.

2. Related Work

Early and accurate diagnosis of brain tumors is essential to provide high-quality treatment. MRI plays a key role in this process. However, there may be difficulties in interpreting it. DL is now a useful instrument for analyzing brain tumors. In the quest to automate and enhance the accuracy of brain tumor analysis, the following are recent advancements. To identify and classify brain tumors in MRI images, a revolutionary Internet of Things (IoT) computing device employs a hybrid technology that combines convolutional neural networks (CNN) and long short-term memory (LSTM). When tested on the 3,264 MRI scans in the Kaggle dataset, the device performed better than conventional CNN models. By reducing the lengthy process of acquiring knowledge, this research boosts productivity. The authors propose a state-of-the-art real-time intraoperative analysis of brain tumors using stimulated Raman histology (SRH) and CNNs. They achieved a high accuracy level of 92.6%, reducing the detection time to less than 150 seconds. This review proposes to explore the place of SRH in molecular diagnosis by integrating spectroscopic detection.

Using a deep multi-level convolutional neural network, the researchers proposed a brand-new, fully automated method for dividing and dividing brain tumors into different types. On a 3,064-slice MRI dataset, it outperformed previous methods and achieved 87.3% accuracy. The results are enhanced by the variety of datasets, alternative format, and version evaluation. For automated brain tumor segmentation on 3D image datasets, this review proposes the development of a fully semantic DL-based segmentation method. This method shows excellent 3D visualization and accurate tumor prediction with an average prediction rate of 91.71%. It also has excellent IOU and BF scores. The authors of this article conducted a comprehensive analysis of 147 recent studies on systems learning and deep learning strategies for diagnosing Alzheimer's disease, brain lesions, epilepsy, and Parkinson's disease. The study, which looked at 22 datasets, sheds light on effective diagnostic methods, highlights recommendations for future research, and highlights problems with analyzing brain disorders that remain unsolved.

The study's outstanding findings lay the groundwork for future advancements in this field. Using three unusual 3D CNN architectures and a variety of MRI scans, the author conducted in-depth research on gliomas. The MICCAI Braats 2018 study exhibited excellent results from this study. It came in second in survival prediction and fifth in the category out of more than 60 groups. It performed exceptionally well in this competitive field, with a survival type classification accuracy of 61.0%. This research result may also be improved by a variety of social structures and learning strategies. They used the DL version of CNN to detect a robotic brain tumor using MRI recordings measured from the Br35H dataset, resulting in an accuracy of 90.9%. In order to make this view easier to understand and translate into more effective clinical applications in the diagnosis of brain tumors, upcoming research may uncover intricate systems and incorporate multimodal recordings. From this point of view, the author also improves the classification of brain tumors based on MRI scans by employing artificial intelligence (AI) techniques with pre-trained models like Xception, ResNet50, InceptionV3, VGG16, and Mobile Net. F1 rates that were unexpectedly high ranged from 91.25 percent to 89.75% using this method. The view's overall performance is improved by learning biases, refining strategies, exploring hyper parameters, and enhancing datasets.

A CNN-based deep learning model for classifying brain tumors based on MRI images was presented by the study's author. In two separate studies, their method achieved first-order accuracy of 91.3% and 92.7%, respectively. This study's productivity could be increased by improving the model's generalizability, its ability to be evaluated using existing methods, and its transferability to clinical settings. To accurately segment brain tumors in MRI and CT images, the researchers developed a novel approach that combines ResNet50 models and 3D U-Net design. With accuracies of 89.02 percent and 91.99%, this method proved to be very effective. Model development, real-time training, and taking into account class imbalance conditions all lead to better outcomes, as other studies in this article demonstrate. To make it easier to accurately classify brain tumors in MRI images, we present a hybrid model that combines a single CNN model with cutting-edge machine learning techniques in this paper.

The proposed hybrid version outperforms the complex CNN models in terms of temporal performance, with an average accuracy of 92.15 percent. By testing the hybrid model on numerous and extensive datasets, the overall performance is enhanced by combining various deep learning models and machine learning algorithms. In order to enhance the precision of MRI image segmentation for brain tumors, the author of this study presents a modified U-Net framework. The subpixel addition and convolution techniques are combined in this version. On the BraTS Venture datasets, the proposed model outperforms the most recent methods with high accuracies of 91.40% and 92.20%. The exploration of model generalization and the enhancement of performance in a variety of scientific image segmentation tasks are made possible by the use of massive datasets and unrestricted computational evaluation. The author of this newsletter provides a general explanation of the postal and majority voting procedures. They added datasets for the type of brain tumor and size-weighted focal loss and target mismatch. When compared to conventional convolutional neural networks, this approach resulted in significant advancements.

The difficulties that arise when determining size weights and enhancing accuracy can be solved through alternative fusion techniques and computational efficiency. In their review, the authors present a robust brain tumor segmentation framework consisting of four MRI images from a single ensemble and an optimized version of the neural network. The proposed method highlights its effectiveness by demonstrating exceptional performance on the BRATS 2018 dataset with good levels of precision, recall, and DICE score. The model's overall performance is enhanced by incorporating additional preprocessing techniques, addressing scaling issues, and investigating quantitative and scientific issues. Using MRI images, the author of this review presents a version of DL fusion with an impressive accuracy of 90.98% for each type of brain tumor. Features from VGG16, ResNet50, and Convolutional Deep Concept Networks (CDBN) are combined in this release.

The statistical augmentation techniques can be improved and the version's applicability to various clinical datasets and imaging modalities evaluated to further enhance its overall performance. In this study, the researchers apply the FPCIFHSS version to predict brain tumor sensitivity using Complex Intuitive Fuzzy Numbers (CIFN) and introduce a related approach to address diagnostic uncertainties. In systems with limited resources, this approach will be extremely beneficial. The version should be made better so that it can be used in a reasonable clinical setting, and future research should test how well it works by comparing it to established diagnostic methods. When it comes to cancer volume discrimination using multi-omic data from the Cancer Genome Atlas, the models presented in this paper are as follows: Xgboost achieved the highest overall performance with an accuracy of 89.62%. Additionally, the Deep Forest Cascade group successfully identified cancer subtypes from the MetaBrick dataset, with 83.45% accuracy for five subtypes and 77.55% accuracy for 10 subtypes. The immense power of AI in the study of genomic archives is demonstrated by these advancements. Improving AI algorithms, enhancing preprocessing techniques, evaluating models on a variety of datasets, and carrying out rigorous scientific experiments to guarantee robust validation should be prioritized in future research. Low accuracy, fine-tuning bias, high computational requirements, poor interpretation, and limited generalization plague previous brain tumor detection techniques. Our version of ViT-GRU addresses these problems by using XAI methods such as LIME, SHAP, and attention mapping, which achieve high accuracy, effectively remove class imbalance, and improve interpretability. It easily processes information from multiple sources, increasing its versatility and relevance across different populations and imaging modalities.

3. Methodology

In our research work, we employed the effective pre-trained classifier VGG16, which is taken from a CNN architecture, in order to recognize brain strokes using CT scan image data. The proposed model includes four important steps: The system's block diagram is presented in Figure 1.

1. Data collection
2. Data cleaning and preprocessing.
3. Feature selection.
4. Model Evaluation (VGG16+RF)

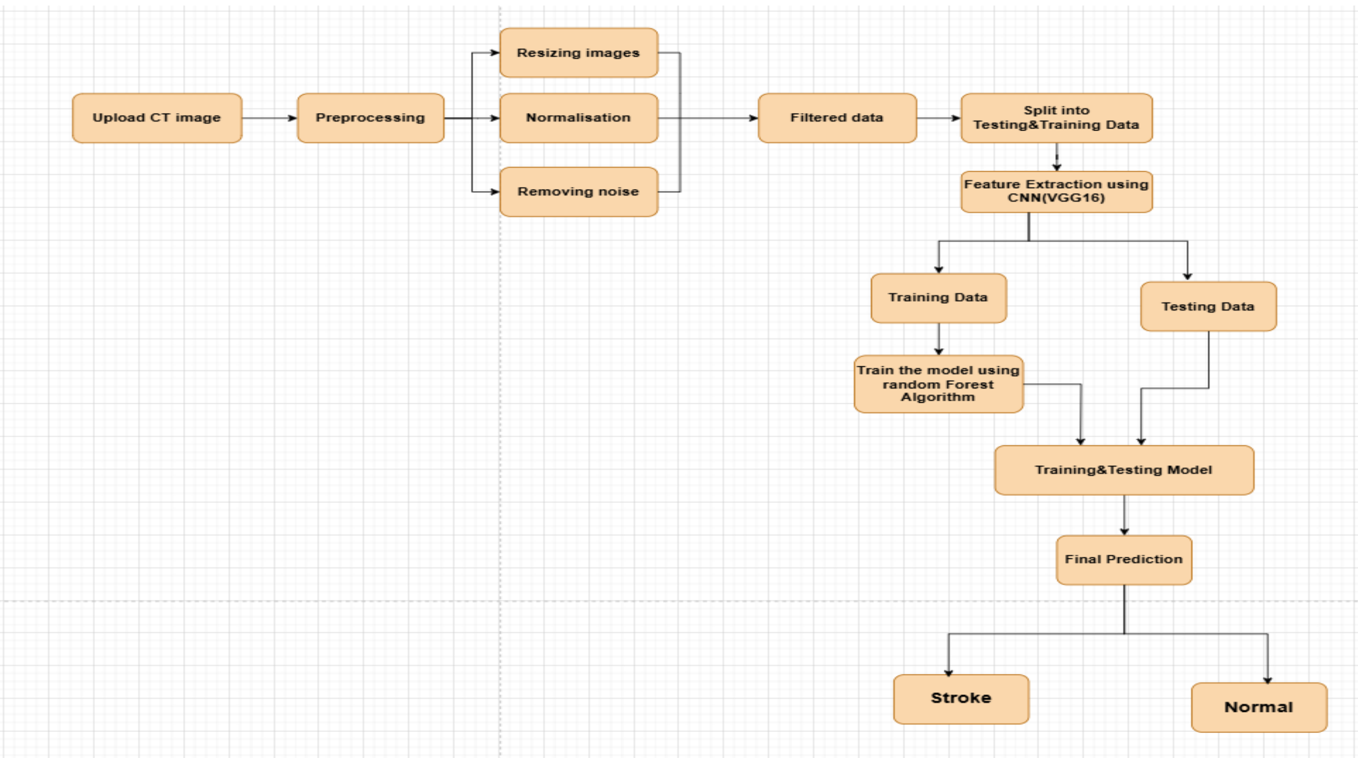


Figure1: System Architecture of proposed model

3.1 Dataset

In this study, Brain Stroke CT Image Dataset was utilized for training the VGG16 and Random Forest model. It consists of 1000 normal images and 1000stroke images, which were divided into two classes: Stroke and Normal. Images from the dataset were employed for training, validation, and testing to determine the performance of the proposed model.

3.2 Pre-processing of the Image

Image preprocessing is used to improve image quality prior to inputting them into the model. Preprocessing includes resizing, normalization, noise filtering, and rescaling (grayscale conversion, if necessary). Preprocessing is done under two broad dimensions:

3.2.1 Cleaning the Data

Data cleaning is the elimination of noisy or irrelevant data through different transformations to enhance model learning. It is an important step in making the visualizations more precise and the trained model more trustworthy.

3.3 VGG16 for Feature Extraction

The Visual Geometry Group (VGG) model, VGG16, is most commonly employed for feature extraction and image classification. VGG16 has 16 layers (13 convolutional and 3 fully connected layers) and is recognized for its deep architecture, which facilitates efficient feature extraction. VGG16 has been employed in this work to extract key features from brain CT scans prior to sending them to the Random Forest classifier for stroke detection.

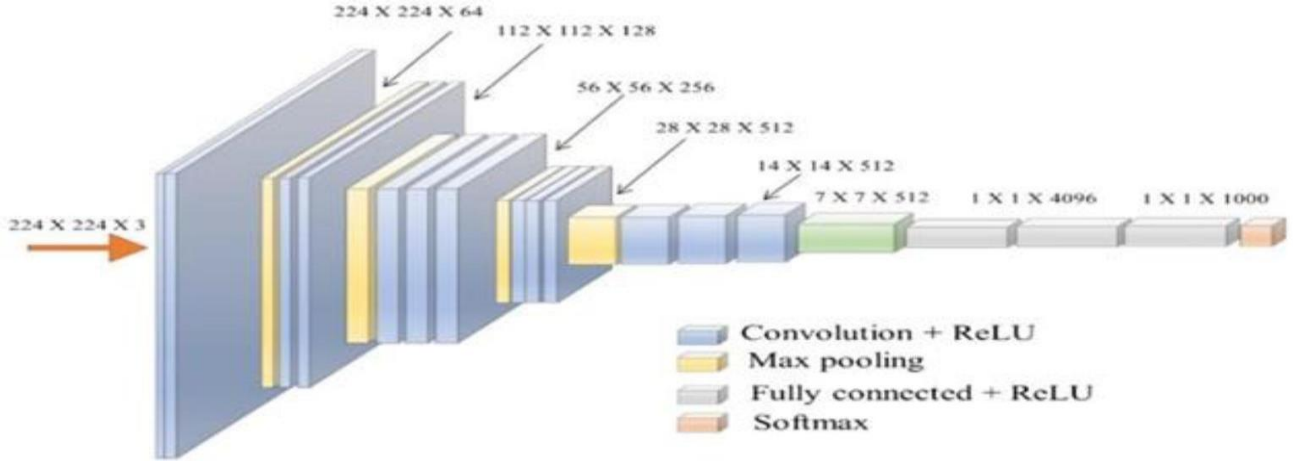
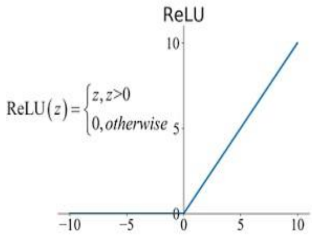


Figure 2: Classification process of brain CT images using the VGG16 model.

Following pre-processing (resizing, normalization, removal of noise), the input image goes through some convolutional layers to extract meaningful features. Convolution is followed by an activation function, ideally ReLU (Rectified Linear Unit), to bring in non-linearity and to enable the model to learn more complex patterns.

ReLU Activation Function

ReLU function sends only positive values and gives zero to negative values, which helps in faster and more efficient training. The function is mathematically expressed as:



Once the feature maps are created, the feature maps are flattened and passed to the fully connected layers for the classification. The final classification is done by the SoftMax activation function, which scales the output to probability values of different classes (Stroke/Normal).

SoftMax Activation Function

SoftMax function is applied to the last classification layer to forecast the probability that the image will belong to a specific class. It is defined as:

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \text{ for } j = 1,\ldots,K$$

Finally, the extracted features are input to the Random Forest classifier, which performs the final classification to identify whether the CT image belongs to the stroke or normal class.

Random Forest (RF):

The Random Forest algorithm is used in the implementation of the proposed brain stroke prediction system. It is an extension of the decision tree model enhancing accuracy by aggregating several decision trees. Random Forest uses feature randomization and bagging to build an ensemble of decision trees and enhance classification accuracy. In contrast to a decision tree, which makes splits using all features, RF uses a random subset of features, decreasing correlation between trees and making the model more robust. Each tree in the Random Forest is learned from a bootstrap sample, i.e., a random sample of a portion of the dataset with replacement, and the remaining data is used as the out-of-bag (OOB) set for testing. It prevents overfitting and generalizes well. For brain stroke detection, the RF model processes CT images and predicts them as stroke or normal based on a majority voting of decision trees. It enhances the accuracy of prediction, hence an ideal choice for stroke detection at an early stage.

4. Result & Discussion

1. Model Performance

The performance of the RF classifier was assessed through accuracy, precision, recall, and F1-score. The model was accurate to the tune of 95.85%, demonstrating efficiency in separating the stroke and normal images.

The Figure3 represents the detailed classification report, including precision, recall, and F1-score for each class.

```
Accuracy: 95.85365853658536

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.98      0.96       220
           1       0.97      0.94      0.95       190

    accuracy                           0.96       410
   macro avg       0.96      0.96      0.96       410
weighted avg       0.96      0.96      0.96       410
```

Figure 3: Classification Report and Model Accuracy

The excellent 0.94 good stroke case recall shows that the model accurately classifies stroke cases with very few false negatives. Macro and weighted averages remain at 0.96, i.e., well-balanced performance for both classes.

5. Confusion Matrix Analysis

To further evaluate the performance of the Random Forest (RF) classifier, we analysed the confusion matrix, as shown in Figure 4.

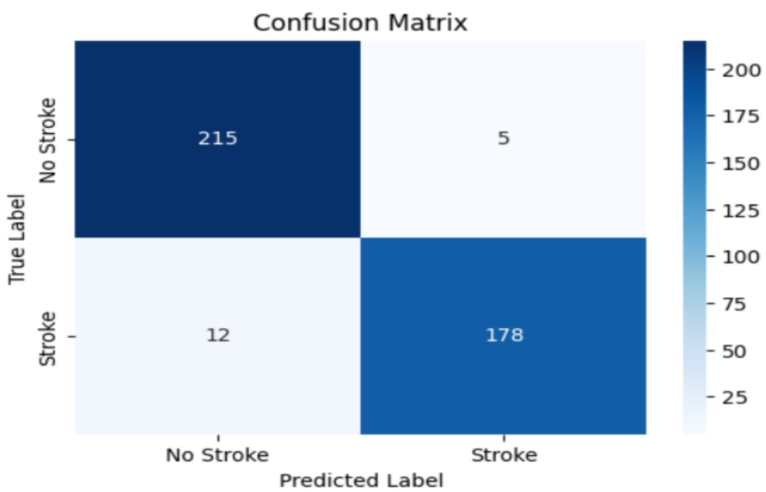


Figure 4: Confusion Matrix Representation

The confusion matrix shows that the model has low false positive and relatively low false negative rates, which signify high classification ability. The higher the values of true positives and true negatives, the more robust the dependability of the model in stroke case identification.

6. Conclusion

To visually inspect the predictions of the Random Forest (RF) model, the system returns a classification result along with the respective CT scan image. Figure 5 provides an example of a correctly classified case, wherein the model classifies "No Stroke" for the provided input. By analyzing such predictions, we can validate the model's effectiveness in identifying stroke and non-stroke cases, supporting the quantitative evaluation presented in the classification report and confusion matrix.

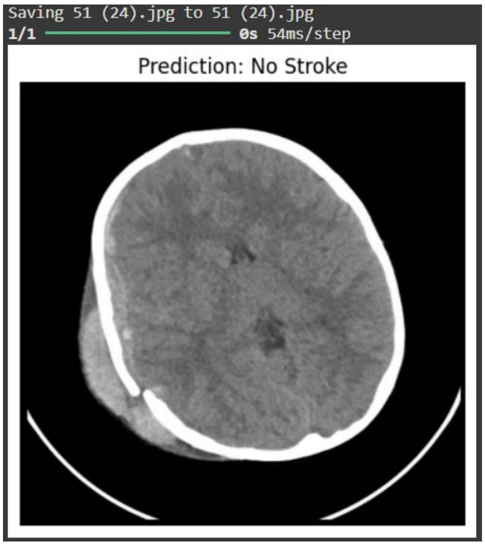


Figure 5: Predicted output of the RF model for a given CT scan image.

References

1. Celik, M. & Inik, O. Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor multi-classification. Expert Syst. Appl. 238, 122159 (2024).
2. Sharif, M., Amin, J., Raza, M., Yasmin, M. & Satapathy, S. C. An integrated design of particle swarm optimization (pso) with fusion of features for detection of brain tumor. Pattern Recogn. Lett. 129, 150–157 (2020).
3. Mudda, M., Manjunath, R. & Krishnamurthy, N. Brain tumor classification using enhanced statistical texture features. IETE J. Res. 68, 3695–3706 (2022).
4. Sajid, S., Hussain, S. & Sarwar, A. Brain tumor detection and segmentation in mr images using deep learning. Arab. J. Sci. Eng. 44, 9249–9261 (2019).
5. Lan, Y.-L., Zou, S., Qin, B. & Zhu, X. Potential roles of transformers in brain tumor diagnosis and treatment. Brain-X 1, e23 (2023).
6. Bhadra, S. & Kumar, C. J. An insight into diagnosis of depression using machine learning techniques: A systematic review. Curr. Med. Res. Opin. 38, 749–771 (2022).
7. Ranjbarzadeh, R., Zarbakhsh, P., Caputo, A., Tirkolaee, E. B. & Bendechache, M. Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm. Comput. Biol. Med. 168, 107723 (2024).
8. Castiglioni, I. et al. Ai applications to medical images: From machine learning to deep learning. Physica Med. 83, 9–24 (2021).

9. Castiglioni, I. et al. Ai applications to medical images: From machine learning to deep learning. Physica Med. 83, 9–24 (2021).
10. Hossain, S., Chakrabarty, A., Gadekallu, T. R., Alazab, M. & Piran, M. J. Vision transformers, ensemble model, and transfer learning leveraging explainable ai for brain tumor detection and classification. IEEE J. Biomed. Health Inform. 28, 1261–1272 (2023).
11. Padmapriya, S. & Devi, M. G. Computer-aided diagnostic system for brain tumor classification using explainable ai. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) Vol. 2 (ed. Padmapriya, S.) 1–6 (IEEE, 2024).
12. Mahim, S. et al. Unlocking the potential of xai for improved alzheimer's disease detection and classification using a vit-gru model. IEEE Access (2024).
13. Younis, A., Qiang, L., Khalid, M., Clemence, B. & Adamu, M. J. Deep learning techniques for the classification of brain tumor: A comprehensive survey. IEEE Access (2023).
14. Vankdothu, R., Hameed, M. A. & Fatima, H. A brain tumor identification and classification using deep learning based on cnn-lstm method. Comput. Electr. Eng. 101, 107960 (2022)
15. Hollon, T. C. et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nat. Med. 26, 52–58 (2020).
16. Díaz-Pernas, F. J., Martínez-Zarzuela, M., Antón-Rodríguez, M. & González-Ortega, D. A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. Healthcare 9, 153 (2021).
17. Karayegen, G. & Aksahin, M. F. Brain tumor prediction on mr images with semantic segmentation by using deep learning network and 3d imaging of tumor region. Biomed. Signal Process. Control 66, 102458 (2021).