OPEN ACCESS Freely available online

ISSN 2737-534X
Volume 6
Issue 2
*International Journal of Clinical Medicine and Bioengineering*

**Article**

# Enhancing Stroke Risk Prediction: Leveraging Machine Learning and MRI Data for Advanced Assessment

**Pranav Kunderu [1,*], Salik Mian [2], and Shivm Patel [3]**

[1] Biological Sciences, Independent, Los Angeles 91307, CA, USA
[2] Physical Sciences, Independent, Los Angeles 91326, CA, USA; salikmian75@gmail.com
[3] Molecular & Cell Biology, University of California, Berkeley, Berkeley 94720, CA, USA; shivmpatel@berkeley.edu
* Correspondence: pranavkunderu1209@gmail.com

**Received:** Sep 27, 2024; **Revised:** May 13, 2025; **Accepted:** May 28, 2025; **Published:** Jun 30, 2026

**Abstract:** As a leading cause of death across the globe, strokes have historically been regarded as a dangerously impactful condition with little to no predictability. Currently, the few ways that people can tell whether or not a stroke is taking place in a patient are through questionable methods, such as looking for warning signs and checking for hereditary factors. We aim to create a quantitative approach to this problem by offering a tool that neurologists will be able to take advantage of, providing them with the data they need to predict strokes before they happen. We intend to use MRI scan data obtained from OpenNeuro, specifically showing brains of adult patients pre-stroke and post-stroke. We will use this data to train a variety of machine learning models independently, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest. By training these models, we aim to accurately predict the risk of a stroke by inputting healthy brain imaging scans and determining their similarity to post-stroke brain scans. If implemented in neurologist offices, this model could have a considerable effect, allowing people to realize that a stroke is imminent and to begin preparing for it, potentially saving lives and improving outcomes.

**Keywords:** MRI, Stroke, Machine learning, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random forest, Healthcare, Statistical analysis

## 1. Introduction

When it comes to strokes, two main types occur. First is the hemorrhagic stroke, which is the deadliest type and occurs when a blood vessel in the brain randomly begins leaking or bursts [1]. The other type is an ischemic stroke, which happens when, in the brain, an artery becomes blocked [2]. Whether an individual experiences one or the other, it can end in tragedy or near-tragedy, demonstrating the idea that strokes remain a powerful condition waiting to wreak havoc upon the nervous system. Up until now, strokes have remained quite unpredictable, with families/friends of victims having to notice simple things occurring about the person, such as trouble speaking or walking. Some other common symptoms that may take place in stroke patients are numbness occurring on the face and arms, problems seeing in any eye, and even trouble understanding what someone else is saying [3]. Watching these can be a hassle and can remain quite troublesome since anyone can experience these difficulties. However, we propose a different method to this problem, a method utilizing both machine learning and MRI data to assess patients and, hopefully, assist neurologists.

MRI, or Magnetic Resonance Imaging, is a type of scan technique that creates an image of the entire structure of the human body [4]. It allows doctors to understand the specific positions of the organs, bones, and muscles of patients that they are treating. This can allow for surgeries to take place or for diagnoses to be put in place for patients. This technology can be used to create specific models for prediction, especially when used in combination with machine learning, like throughout this study [5]. Machine learning can be used with MRI scans to design certain algorithms that provide modeling applications for many types of conditions/diseases [6]. In this case, it can be used for strokes, as that remains the major topic of this research. We would like to create a model with a very low possibility of inaccuracy [7]. We hope to assist any doctor in need of a predictability device for their stroke patients [8], making sure that they are being as correct as possible and receiving the information that they need to proceed with their duties.

### 1.1. Problem Statement

To create a machine learning algorithm able to generate a stroke prediction confidence score using MRI scans for training data. By helping patients to see if they are at risk for stroke, we are able to run practical tests on our model before finally giving copies of it to neurologists and other general physicians who mainly need it, enabling them to use it to their best advantage. This program has a target goal to diagnose strokes quicker and allow patient safety through an accurate prediction.

### 1.2. Contributions

- To enhance specific patient awareness, we intend to create an algorithm meant to predict stroke risk.
- This model will use machine learning methods and MRI scans as its foundation.
- Some MRI scans are used as training data, while others are used as testing data. This will allow for an accurate model.
- We hope to depict the exact stroke risk in any specific patient through this model. It can be implemented in many doctors' offices across the world due to its high value when it comes to humanitarian causes.

## 2. Materials and Methods

The data that we trained and tested our models on was obtained from a publicly available database called OpenNeuro. There were two datasets that we utilized in our research, one of post-stroke patients, and one of healthy pre-stroke patients. The post-stroke images were obtained in a study that collected MRI scans which can be found here [9]. The pre-stroke images were acquired in another study that gathered MRI scans available here [10]. We downloaded the scans taken from the back of the head for each subject separating them into different folders corresponding to either healthy or stroke. We loaded all the data from the local hard drive into the program using code.

### 2.1. Data Description

In total there were 29 scans collected from the post-stroke database and 29 scans collected from the healthy pre-stroke database. 3 scans from both classes were purposely separated before training to be used in an additional functionality. This functionality was added to make the code more accessible for clinical purposes. It allows the user to add a link or file path to an image and it will return a stroke risk confidence score ranging from 0 to 1, 0 being healthy and 1 being stroke.

### 2.2. Data Preprocessing

Each image must be transformed into a format that the machine learning models can understand. To do this a number of steps must be taken, firstly to transform the scan into a grayscale image. The next step is to reshape the image into a 128x128 format to ensure uniformity between all the scans in the dataset. Since we have transformed all the images into grayscale, each pixel of each image can be represented as a singular number instead of 3 (corresponding to RGB in color images). The grayscale numbers range from 0 to 255 and are a measure of how dark or bright a pixel is. The 128x128 array is then flattened into a 1-dimensional array to be fed into the remainder of the model.

### 2.3. Variables

The independent variables in this research are the MRI scans that we obtained from the open-source databases and the new scans users input into the program. The dependent variable in this study is the similarity score that the model outputs for these brand new, unseen MRI scans. The similarity score, as mentioned earlier, varies between 0 and 1, where 0 is a very healthy patient, and 1 implies that the scan highly resembles that of a stroke patient. By determining how similar a pre-stroke patient's MRI scan is to a post-stroke patient, we can make a prediction regarding their risk for stroke.

### 2.4. Techniques Utilized

1. Transfer learning: When a model developed for a particular task is used as a starting point for another model in a second task. If a neural network has been trained on a large dataset of images, then one can fine-tune it on a much smaller dataset in a specific image classification task. Such a strategy makes use of the features and patterns learned from the original task, which will help get optimized performance on the new task where little training data is available. In our research, we utilized the VGG16 transfer learning model to assist with feature extraction. We inputted the 1D array of pixel values and the VGG16 base

model returned features such as edges and textures that were fed into the classifier models outputting the stroke risk confidence score for the scan. The classifier models we coded include SVMs, Random Forests, KNNs, and Adam Optimizers.

2. Cross Validation: Another technique that we employed was cross validation. The benefits of cross validation include a reduction in overfitting and the variance in the performance of the data is decreased. The cross-validation technique functions by splitting the data into several sections called folds. In our code, we chose to do 5-fold cross validation. This means that the dataset we provide is randomly split into 5 sections and 4 of them are used to train while the 5th is used to test. The model retrains 5 times over with a different testing fold in every iteration. The final performance of the model is calculated as the average of the performances of all 5 iterations. By training with different subsets of the data, the model improves its ability to classify new data it has never seen, also known as improved generalization.

### *2.5. Algorithms*

1. Support Vector Machines: SVMs are models of supervised learning applied in problems of classification and regression. They ensure that there exists a maximum distance between an optimized hyperplane (that separates the different classes of data) and the closest data points to the hyperplane. SVM is effective when the dimensionality of the data is high, and the number of samples is limited. They allow for both linear and nonlinear classification using kernel functions. In the context of our program, we utilized the Gaussian kernel also known as a RBF (radial basis function) kernel. This kernel is well suited for classification problems where the relationship between the independent and dependent data is highly non-linear.
2. K-Nearest Neighbors: KNN is a simple instance-based learning algorithm for classification and regression. When implementing KNN, all the input data is mapped into a feature space. A new data point is classified based on the majority class of its 'k' nearest neighbors in the feature space. For regression, this output is usually an average of values obtained from those neighbors. Basically, this algorithm uses metrics of distances, among which is the Euclidean distance, to determine how similar data points are to each other. Euclidean distance, which we utilized in our research KNN model is calculated as:

$$\text{distance}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}$$

   Where $x_1$ and $x_2$ are data points in the feature space and $x_{1i}$ and $x_{2i}$ are the *i*th feature of $x_1$ and $x_2$. Although KNN is easy to implement, it can be rather computationally expensive for large datasets since computing numerous distances among data points is involved.
3. Random Forest is an ensemble learning method where many decision trees combine to rise in accuracy and power of prediction. A decision tree is another type of machine learning model consisting of 3 main components, nodes, branches and leaf nodes. The full set of data starts at the first node (also called the root node) and the dataset is split based on a decision represented by the node. For example, a set of human individuals can be split based on age at a particular node. Then each subset travels along a branch until it encounters the next node which further splits the data into smaller subsets. This cycle repeats until the data reaches a leaf node which represents the classification of the data. In the case of random forests, each tree in the forest will have been trained on a random subset of the full training data. This randomness enables reduced overfitting and can increase generalization ability. In the process of prediction, all the predictions of individual trees are aggregated by the Random Forest to come up with a final decision. In other words, the most popular classification made by the trees within the Random Forest is what the model decides as the final classification.
4. The Adam optimizer is a popular machine learning algorithm, and the name Adam is short for Adaptive Moment Estimation. The purpose of this algorithm is to guide the gradient of the model. When training a model, we often take the difference between the predicted and actual result, which is called the error. The gradient tells us how to change the parameters (or weights) of our model to reduce the error as much as possible. What makes an Adam optimizer unique is that it looks at gradients from the past instead of only considering the current gradient. By evaluating the average direction and the average size of the previous gradients the Adam optimizer can choose when to make large or small changes to the weights intelligently. This allows for smooth changes that do not disrupt the accuracy of the model or the learning process. Due to this feature of the optimizer, it requires less manual tuning than many other models. Along with this optimizer, we chose to use binary cross entropy for our loss type. This denotes how the error of our model is calculated during the learning process. Binary cross entropy is especially well suited for classification models with two classes, which in our case are stroke and healthy.

### *2.6. Statistical Measures*

To evaluate the suitability of our models, we decided to use a classification report which collects the following statistical values: precision, recall, accuracy, and f-1 score. Each of these statistics can be calculated as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} = \frac{\text{True Positive}}{\text{Total Predicted Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{\text{True Positive}}{\text{Total Actual Positive}}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Where TP refers to true positives, TN refers to true negatives, FP means false positives, and FN means false negatives.

In addition to the statistical values recorded in the classification report, our program outputs the $R^2$ value and Mean Squared Error of each model. Mean Squared Error is found by subtracting the predicted result and actual result of a particular data point, then squaring it and gathering the average across all the data points. This process is defined in the following formula:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

In this equation, n is the number of data points, $y_i$ is the actual value corresponding to a data point, and $\hat{y}_i$ is the predicted value from the model. Of course, a smaller MSE value suggests the predictions made by the model are accurate to the actual classifications.

The $R^2$ value of a model is an indicator of how well the model can understand the patterns and relationship between our independent and dependent variables. This value can range between 0 and 1, with 1 representing a perfect model of the data, and 0 being an unreliable model. The $R^2$ value is calculated using the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Where $y_i$ corresponds to the actual value associated with a datapoint, $\hat{y}_i$ is the predicted value the model generates, $\bar{y}$ is the mean of the actual values across all data points, and n is the number of data points.

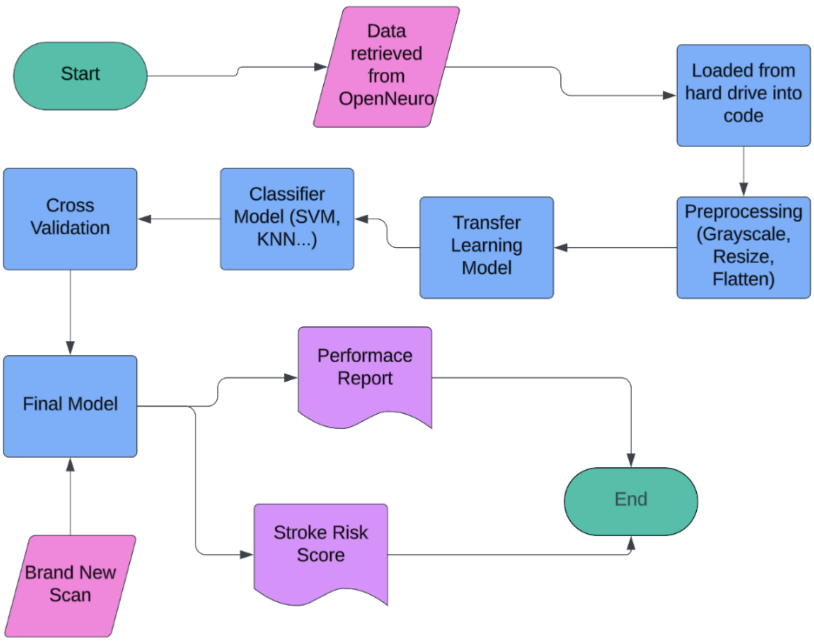


**Fig. 1.** This flow chart represents the necessary steps the algorithm will take to reach its end goal of computing the stroke risk score and generating the performance report.

### 3. Results

For this study four different types of machine learning models were used; Adam Optimizer (Adaptive Moment Estimation), Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN).

The confidence levels for the different scans using the Adam Optimizer model were mixed. In Scan 1, the confidence values lie within the range of 0.0039 to 0.0112, with an average of 0.0064, indicating a high probability that the patient is healthy. Scan 2 had higher confidence values: from 0.0196 to 0.0293, averaging 0.0243, still indicating a relatively healthy patient but with rising certainty of a possible condition. Scan 3 also showed a confidence of 0.0222, virtually identical to that of Scan 2. Scan 4 demonstrated high confidence levels of 0.9741-0.9854 with an average of 0.9806, a strong indicator for a stroke condition. Scan 5 had almost perfect confidence, with confidence levels ranging from 0.9968 to 0.9993, with an average of 0.9982, indicating a very high likelihood that the patient was a stroke patient. Scan 6 gave similar results of high confidence, averaging 0.9966, thus strongly indicating the likelihood of a stroke. Scan 7 had considerably lower confidence, averaging 0.1068, indicating a much healthier patient compared to the previous scans. Scan 8 was at a moderately high level of confidence, averaging 0.7591, indicating a higher likelihood of a stroke than the initial scans but less certain than Scans 4, 5, and 6.

The SVM model again followed through with flat confidence levels across the scans. Scan 1 had confidence ranges from 0.0489 to 0.0503, with an average of 0.0497. This indicated a healthy patient. Scan 2: The confidence levels ranged from 0.1091 to 0.1113 with an average of 0.1110. Thus, this was a relatively healthy patient with some certainty that some issues were present. Scan 3 had about equal confidence levels with an average of 0.1208. Scan 4 had very high confidence levels between 0.9059 and 0.9098 with an average of 0.9082 hence strongly indicating a stroke patient. Scan 5 confidence was kept very high, around 0.9686 to 0.9702, averaging 0.9695, very indicative of a stroke. Scan 6 also indicated quite high confidence with an average of 0.9560. Scan 7 had much lower confidences, averaging 0.3136, indicating a healthier patient but approaching worrying confidence levels. Scan 8 had relatively medium confidence levels, with an average of 0.7717, indicating a higher probability of stroke but less confidence compared to Scan 4 and Scan 5.

The Random Forest model returned the same confidence levels for all runs on any scan. Scan 1 returned a confidence level of 0.06 all the time. This means it is a very healthy patient. Scan 2 maintained a confidence level of 0.24 all the time. This means it belonged to a relatively healthy patient. Scan 3 had a confidence level of 0.25, where it indeed showed similar outcomes to Scan 2. Scan 4 had a rather high confidence level of 0.86, so it would strongly indicate the presence of a stroke. Scan 5 showed an even superior confidence level of 0.92, signifying a very high likelihood of stroke. Scan 6 maintained a confidence level of 0.87, also indicating a high likelihood of stroke. Scan 7 had a confidence value of 0.43, indicative of a healthier patient compared to Scans 4 and 5 but less healthy than earlier scans. Scan 8 had a confidence value of 0.55, showing a moderate risk of stroke.

Regarding the KNN, all trials for scans 1, 2, 3, and 7 returned a confidence level of 0.0; thus, these patients are very healthy. At the other extreme, perfect confidence levels of 1.0 for every trial suggest that scans 4, 5, 6, and 8 point rather considerably to stroke patients.

For all 4 models we found that the accuracy for a fold 5 was perfect at 1.0. In cross validation it turned in a precision, recall and F-1 Score of 1.0 for both the health and stroke classes. The cross-validation $R^2$ value was also 1.0 showing that all variance was accounted for in this model thus also producing a Mean Standard Error (MSE) of 0.0. By all metrics the pure classification performance of all 4 models was perfect.

### 4. Discussion

Ref. [1] talks about cerebrovascular accidents, or stroke, which can be either ischemic or hemorrhagic in nature. Hemorrhagic strokes account for 10-20% of strokes and include intracerebral hemorrhage and subarachnoid hemorrhage. Often caused by hypertension or another underlying condition, they are characterized by an initial compression from hematoma raising ICP, followed by secondary injury due to inflammation and oxidative stress. Diagnosis is mostly done by CT and MRI, whereas angiography finds the causes. Prognosis depends upon hematoma volume and age; recovery is good if properly managed and rehabilitated.

Ref. [2] is regarding an ischemic stroke which occurs when there is a sudden onset of focal neurological deficits due to the impaired blood flow to the brain caused by an embolic or thrombotic event. Prevention through lifestyle modification and management of risk factors such as hypertension and diabetes is also vital.

Ref. [3] explores how stroke is the second leading cause of death and disability worldwide and preferentially affects developing countries. Substantial advances in the understanding of stroke pathophysiology and treatment have concentrated on restoration of blood flow and management of neurological damage. Recently, when clinical trials reported limited success, it was found that researchers developed more refined animal models of stroke and targeted study designs, along with advanced technologies.

Ref. [4] considers MRIs, which provide relevant information for diagnosis. Though tPA is the only proven therapy for acute ischemic stroke, there are hemorrhagic risks with its use. It can also aid in the identification of patients who would benefit from thrombolysis by differentiating stroke from stroke mimics using MRI, particularly diffusion-weighted imaging and perfusion-weighted imaging, for the assessment of salvageable tissue and the prediction of clot characteristics and hemorrhagic risk. The pure dependence on MRI-based selection of patients for thrombolysis has translated into mixed clinical trial outcomes, which means a comprehensive approach should be used incorporating multiple MRI findings to guide treatment decisions.

Ref. [5] explores how machine learning has brought much hope to stroke medicine, as it enables efficient analysis of large data volumes and supports strategies in personalized medicine. Recent advances in ML research established high accuracy in the analyses of imaging, diagnosis of stroke subtypes, risk stratification, guidance of treatments, and prediction of patient outcomes. Broad-scale adoption of ML holds the future for stroke medicine, with potential enhancement of diagnosis, treatment selection, and prognostication to improve outcomes and quality of life.

Ref. [6] studies machine learning applications in stroke rehabilitation showed that random forests, logistic regression, and deep neural networks, including convolutional neural networks for image analysis, could be used to predict motor function recovery using clinical data and magnetic resonance imaging effectively. These models have potential in predicting functional outcomes, interpretation of video fluoroscopic swallowing studies, and maybe even in prognosis of depression, language, cognitive, and sensory recovery from stroke.

Ref. [7] looks into ten different Classifiers that were trained with Logistic Regression, Decision Tree, AdaBoost, and XGBoost and aggregated to maximize accuracy using weighted voting. The accuracy rate for the proposed model was 97%, far ahead of individual classifiers, while areas under the curve were high and false positive and negative rates were low. Thus, the weighted voting classifier has been very efficient and effective in this regard for the purpose of early stroke prediction, benefiting physicians and patients alike.

Ref. [8] explores how strokes are one of the critical medical conditions that needs immediate diagnosis and intervention to prevent permanent damage in the brain. This paper presents an advanced stroke detection algorithm for predicting stroke events using machine learning techniques. The dataset used contained parameters such as age, BMI, gender, heart disease, and smoking status. The data was then preprocessed for missing values, categorical features, and balance. Comparisons were made with different classifiers, including Naïve Bayes, logistic regression, XGBoost, decision trees, AdaBoost, K-Nearest Neighbor, random forests, voting classifier, and support vector machines, in order to create prediction models. In this research, the performance of the models is classified based on accuracy, F1-score, recall, precision, and an extra metric of specificity. A Support Vector Machine algorithm was run with an accuracy of 99.5%, precision of 99.9%, recall of 99.1%, F1-score of 99.5%, and specificity of 99%.

After reading through these 8 literatures, the biggest gap that we seemed to notice throughout all the papers is the absence of the stroke MRI scans that were used in combination with machine learning methods. We found studies that used stroke MRI scans and we found articles that used machine learning, but we never found anything that used both at the same time, which is where our study comes in, as a more up to date research paper regarding these topics.

## 5. Conclusion

The Random Forest model gives very consistent levels of confidence on all types of scans; thus, it shows that it is robust, reliable, and insensitive to intermediate risk levels at the same time. Such subtle output from the Adam Optimizer model may find value in the estimation of various levels of stroke risk, with SVM and Random Forest models being stalwart models. The result obtained by the Adam Optimizer model involved various confidence levels against the scans, where high values indicate a strong likelihood of stroke, and the lower values indicate a healthy status. This model was good at sorting out low- from high-risk cases, but the mixed confidence levels for the intermediate ones suggest that there may be space for improvement in precisely identifying varying degrees of stroke risk. The SVM model provided very consistent levels of confidence on similar scans and perfectly distinguished whether it is a stroke case or a healthy case. Even though this model is very stable, it might turn out to be less discriminatory on intermediate risk levels compared to the Adam Optimizer. The further development and testing of the models' robustness and generalizability are both important tasks.

The reason the accuracy, precision, recall, and f-1 score for the four models is very high is because all these metrics are measuring is how good proficient each model is at identifying if the patient is healthy or is a stroke patient. For example, if the model outputs a 49% confidence level the model will still identify it as a healthy patient, which explains why the classification report outputs such astounding values.

Many strokes occur without any obvious symptoms, making strokes some of the most unpredictable medical phenomena. Each person's stroke risk is influenced by a combination of genetic, environmental, and lifestyle factors, which can make it difficult to create a one-size-fits-all predictive model. While certain risk factors such as high blood pressure, diabetes, and smoking are known to increase stroke risk, predicting when a stroke will occur remains complex. Risk factors can be managed but are not always fully controlled.

Further research can focus on biomarkers. Research into biomarkers associated with stroke risk could provide predictive capabilities. Biomarkers might include specific proteins or genetic markers found in blood or cerebrospinal fluid. By combining and creating models that consider biometrics and patient medical history we can create more accurate predictive models.

Using machine learning is the first step in working towards a model that can predict strokes before they happen. The model's ability to predict stroke could further be enhanced if trained with scans from individuals who have had stroke, but the scans were taken before the stroke occurred. As usage of our program increases, the model will absorb more data and continue to add to its training. The more MRI scans that the model sees the more accurate it will become at predicting when someone is at elevated risk of having a stroke in the future.

**Author Contributions:** Conceptualization: P. Kunderu and S. Mian. Methodology: P. Kunderu and S. Mian. Software: S. Mian. Validation: P. Kunderu, S. Mian, and S. Patel. Formal Analysis: P. Kunderu and S. Mian. Investigation: P. Kunderu and S. Mian. Resources: P. Kunderu. Data Curation: P. Kunderu, S. Mian. Writing – Original Draft Preparation: P. Kunderu, S. Mian, and S. Patel. Writing – Review and Editing: P. Kunderu and S. Mian. Visualization: S. Mian. Supervision: P. Kunderu.

**Funding:** This research did not receive external funding.

**Data Availability Statement:** The data of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Unnithan, A.K.A.; Mehta, P.; Das, J.M. Hemorrhagic Stroke. *StatPearls [Internet]*, National Library of Medicine, 2023. Available online: https://www.ncbi.nlm.nih.gov/books/NBK559173/ (accessed on October 11, 2024).
2. Hui, C.; Tadi, P.; Patti, L. Ischemic Stroke. *StatPearls [Internet]*, National Library of Medicine, 2022. Available online: https://www.ncbi.nlm.nih.gov/books/NBK499997/ (accessed on October 11, 2024).
3. Kuriakose, D.; Xiao, Z. Pathophysiology and Treatment of Stroke: Present Status and Future Perspectives. *International Journal of Molecular Sciences* 2020, *21*(20), 7609. https://doi.org/10.3390/ijms21207609.
4. Kim, B.J.; et al. Magnetic Resonance Imaging in Acute Ischemic Stroke Treatment. *Journal of Stroke* 2014, *16*(3), 131–145. https://doi.org/10.5853/jos.2014.16.3.131.
5. Daidone, M.; Ferrantelli, S.; Tuttolomondo, A. Machine Learning Applications in Stroke Medicine: Advancements, Challenges, and Future Prospectives. *Neural Regeneration Research* 2024, *19*(4), 769. https://doi.org/10.4103/1673-5374.382228.
6. Choo, Y.J.; Chang, M.C. Use of Machine Learning in Stroke Rehabilitation: A Narrative Review. *Brain & Neurorehabilitation* 2022, *15*(3), e26. https://doi.org/10.12786/bn.2022.15.e26.
7. Emon, M.U.; Keya, M.S. Performance Analysis of Machine Learning Approaches in Stroke Prediction. IEEE Xplore, 2020. Available online: https://ieeexplore.ieee.org/document/9297525 (accessed on October 11, 2024).
8. Mushtaq, S.; Saini, K.S.; Bashir, S. Machine Learning for Brain Stroke Prediction. IEEE Xplore, 2023. Available online: https://ieeexplore.ieee.org/document/10151148 (accessed on October 11, 2024).
9. OpenNeuro. ds003999 (MRI data for stroke patients). Available online: https://openneuro.org/datasets/ds003999/versions/1.0.2 (accessed on October 11, 2024).
10. OpenNeuro. ds003469 (MRI data for healthy subjects). Available online: https://openneuro.org/datasets/ds003469/versions/1.0.0 (accessed on October 11, 2024).