Analysis of Brain Tumour and Stroke Prediction using Selected Machine Learning Algorithms

[1]Percy Okae, [1]Ernest Owusu Boadi, and [1]Prince Darko

[1]Computer Engineering Department, School of Engineering Sciences, University of Ghana, Accra, Ghana

***Corresponding Author:** pokae@ug.edu.gh

ABSTRACT

In this work, brain tumour detection and stroke prediction are studied by applying techniques of machine learning. To address challenges in diagnosing brain tumours and predicting the likelihood of strokes, this work developed a machine learning-based automated system that can uniquely identify, detect, and classify brain tumours and predict the occurrence of strokes using relevant features. The system utilizes advanced algorithms to analyze medical data and images to extract meaningful patterns and relationships that can assist in accurate prediction. The machine learning algorithms used for the brain tumour prediction is the convolutional neural network (CNN), whilst for the stroke prediction, support vector machine, random forest, decision tree and logistic regression were used for purposes of comparison. The dataset was partitioned into 80 % for training and 20 % for testing the brain tumour images using the machine learning programming language Python. It was observed that when the CNN model was trained up to 100 epochs, it achieved an overall accuracy of 95.78 %. The precision of the simulated model was 96.70 %, recall was 96.65 %, and F1-score was 96.17 %. For the stroke images, it was partitioned into 90 % for training and 10 % for testing and the decision tree algorithm gave the most accurate results among the four machine learning algorithms with an area under the curve (AUC) score of 0.97 on the original dataset and a value of 1.00 after hyperparameter tuning.

Keywords: machine learning, brain tumuor, stroke, MRI, convolutional neural network, Kaggle

1.0 INTRODUCTION

Brain tumours and strokes, two distinct yet equally critical neurological conditions, represent significant burdens on healthcare systems and patients' lives. A brain tumour, whether benign or malignant, can disrupt intricate neural networks, causing a cascade of neurological impairments according to (Ari, and Hanbay, 2018, Arora, and Sharma, 2021). Similarly, a stroke resulting from interrupting blood supply to the brain can lead to devastating consequences, including cognitive deficits and motor impairments (Chen-Ying Hung *et al.*, 2017, Ahmed *et al.*, 2019). The imperative to diagnose these conditions early and accurately is underscored by timely intervention, significantly influencing treatment outcomes. Brain tumours can be categorized based on their severity, including pituitary gliomas, meningioma, and glioblastoma (El-Feshawy, Saad, Shokair, and Dessouky, 2021).

In recent years, brain tumours and strokes have emerged as prominent contributors to mortality, impacting the central nervous system. Specifically, strokes encompass a significant proportion of these neurological challenges, with ischemic and hemorrhagic events inflicting substantial damage on the central nervous system (Puspitasari *et al.*, 2021). The global prevalence, as indicated by the World Health Organization (WHO), underscores the gravity of the situation: subarachnoid hemorrhage affects 3% of the population, intracerebral hemorrhage affects 10%, while the majority, a staggering 87%, are afflicted by ischemic strokes (Hossain *et al.*, 2021). This research integrates machine learning techniques to discern, categorize, and predict strokes and brain tumors using medical data and images. Traditional diagnostic methods for brain tumors and strokes rely on a combination of clinical assessments, medical imaging, and invasive procedures. While these approaches have contributed significantly to patient care, they often fail to provide rapid and precise predictions. The complex and multifaceted nature of these conditions necessitates a more nuanced and data-driven approach that can uncover hidden patterns in the vast datasets generated by medical imaging technologies, genetic analyses, and patient histories. In today's medical field, the collaboration between advanced technology and clinical know-how has made impressive progress in diagnosing and foreseeing illnesses. A groundbreaking example is the use of machine learning in neurology, specifically to detect and predict brain tumors and strokes as early as possible. Machine learning is a method in which a mathematical model employs patterns from input data attributes to predict or classify the output of new data points.

The automatic brain tumour segmentation (ABTS) method developed by Arora and Sharma, (2021), was employed for segmenting various constituents of the brain. They used four magnetic resonance images were to identify edema and gross tumour volumes (GTV). Thresholding, edema, and GTV segmentation were carried out in that order in this work. This method turned out to be fast and very accurate. Sehgal *et al.*, (2016), introduced a local independent projection-based into classification (LIPC) to derive a novel classification framework. In their approach, they considered locality as an important parameter whilst also prioritizing the data distribution of different classes using SoftMax Regression Model. The four stages in their method were preprocessing, tumour segmentation using the LIPC method, feature extraction, and post-processing using spatial constraints. Also, a multi-resolution framework was embedded in their work just to minimize the cost of computation. Their method was able to address the challenges posed by tumour segmentation which are generally caused by the complex characteristics of brain tumour MRI images. Gurbina, Lascu, and Lascu, (2019), used Field Programmable Gate Array (FPGA) implementation to study brain tumours classifications. This technology happens to be very suitable doing real-time analysis of algorithms used in image processing. This technique also happens to be cost saving when the need arises to implement new segmentation techniques on hardware. They use

SVM to classify the MRI data into tumourous brain and tumourless brain. Ahmadvand, (2016), applied Markov random field model, a wavelet-based method applied to MRI imaging to extract a feature vector. The feature that was extracted, called the modality fusion vector (MFV) detected the brain tumours as 3D images automatically. Subsequently, histogram matching, and bias field correction were applied on the images during which the regions of interest were separated from the backgrounds. The method applied for the segmentation of tumours was random forest. Local binary patterns in three orthogonal planes (LBP-TOP) and histogram of orientation gradients (HOG-TOP) of MR images were used as the classifier. The performance was tested by the brain tumor segmentation (BRATS) 2013 dataset. Tusher, Sadik, and Islam, (2022), collected their datasets from Turgut Ozal Medical Centre, Inonu University, Malatya, Turkey and used Support Vector Machine (SVM), Stochastic Gradient Boosting (SGB), and penalized logistic regression (PLR) to predict stroke. The findings of the research proved that SVM achieved the highest accuracy of 98%. Previous studies of stroke disease prediction and brain tumor classification only employed traditional machine learning algorithms to predict stroke. To predict stroke and classify brain tumor, distributed machine learning on different platforms was used here.

2.0 SYSTEM DESIGN AND DEVELOPMENT

The primary users of our healthcare application are frontline healthcare providers who interact directly with patients. To access the web-based program, doctors simply log in using their unique username and password, with an additional layer of security through email verification to confirm their account. The central focus of our work revolves around brain tumor classification and stroke prediction. The predictive tool is seamlessly integrated within the web application, offering doctors a user-friendly home page to select the specific prediction they require based on the patient's health issue.

2.1. Brain Tumor

The suggested method involves mapping data features to provide MRI images. The common texture-based selection technique of gray-level co-occurrence matrix is used to select features. Various feature selections, cleaning, standardization, and normalization approaches are used to pre-process the photos. Following that, the photos are downsized to 224 X 224 for use as input to multiple transfer learning models, and lastly, after training, the tumor class is classified using CNN. CNN generally learns to extract features completely and has fewer specific tasks compared to traditional methods. Figure 1 is a screenshot of the CNN process scheme.

The dataset used in this work was collected from the Kaggle dataset which is made up of training and validations sets. Each folder contains four MRI subsets: no tumor, glioma, meningioma, and pituitary

The MRI result image size is 512 x 512. The dataset includes three common MRI scan views: sagittal, coronal, and axial. Because we want our models to work on different scan perspectives, the image sequence in the dataset has been jumbled. Table 1 shows the datasets for the training and testing samples which consists of 1,621 slices of glioma, 1,644 slices of meningioma, 2,000 slices of no tumour and 1,757 slices of pituitary tumours. For this work, we divided the datasets into two: namely, training which contained 80 % of the datasets while the remaining 20% were used for testing. In total, 5,712 images were used for training and 1,310 for testing. Pre-processing techniques such as rescaling and normalization were used before putting medical images into the CNN. Medical photos are frequently of varying sizes and resolutions. Rescaling ensures regular proportions, while standardization decreases pixel intensity fluctuations. We stored all our images in a .jpg format after normalization and their respective annotations were deployed in an excel file.

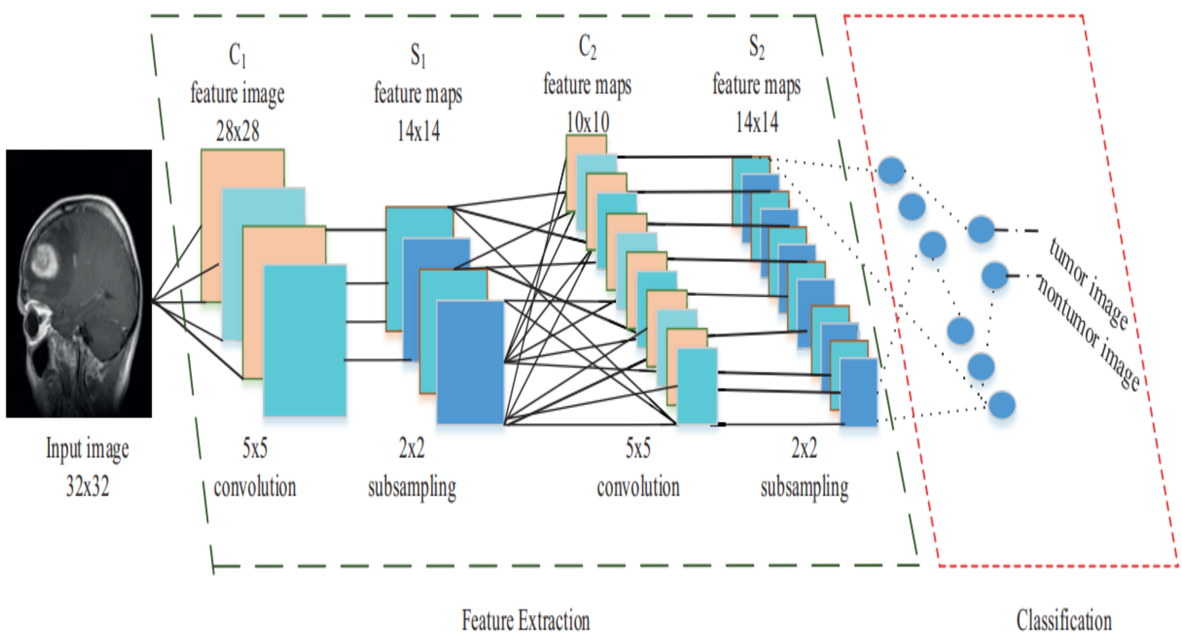

Figure 1: CNN process scheme showing the stages of feature extraction

Table 1: Dataset summary of various brain tumours

	Glioma tumour	Meningioma tumour	Pituitary tumour	No tumours	Total no. of samples
Training samples	1321	1339	1457	1595	5712
Testing samples	300	305	300	405	1310

2.2 Stroke

Four different types of machine learning algorithms were applied on the stroke dataset. These were decision tree, support vector machine, random forest and logistic regression (Zhang, Gan, and Huang, 2019). Hyperparameter tuning and cross-validation were also integrated into the algorithms to enhance our results. The features of the datasets that were used in the stroke analysis are age, gender, hypertension, heart disease, body mass index, ever_married, work_type, residence_type, smoking_status, and avg_glucose_level. The datasets were collected from Kaggle which is one of the most popular big data platforms that handle big data and includes an MLlib library (www.kaggle.com). Accuracy, Precision, Recall, and F1-measure were used to calculate performance measures of machine learning models.

The web applications developed to enhance ease of access and use of the system have the interfaces shown in Figures 2 and 3

.

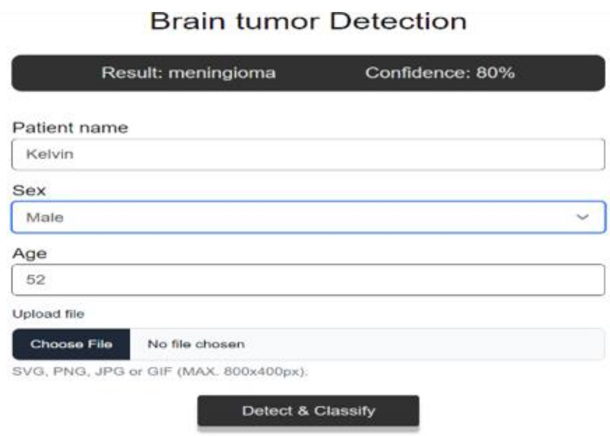

Figure 2a: Brain tumor classification results page

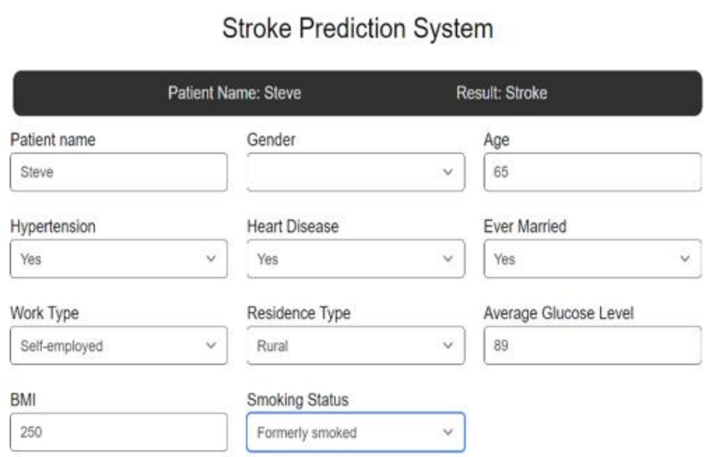

Figure 2b: Stroke prediction results page

Figure 2: Web interfaces of brain tumour classification results page and stroke prediction results page

3.0 SYSTEM IMPLEMENTATION AND TESTING

3.1. Testing of design and results

3.1.1. Brain Tumor

For the brain tumour, we used the CNN model to train up to 100 epochs for the classification of the brain tumour datasets. It achieved an accuracy of 95.78 % when it was run on the testing set. The dataset consisting of 5,712 images were split into a training set of 4,570 images (80% of the total MRI scans) and a testing set of 1,142 images (20% of the total MRI scans). A precision of 96.70 % was obtained and the recall was also 95.65 %. Also, an F1 score of 96.1 % was obtained. Table 2 gives the summary of the results. This simulated how the model would perform in real life scenarios with previously unseen MRI scans. Table 3 also gives the mean average precision for overall performance measurement of the various types of brain tumor.

Table 2: Model performance on the brain samples

Models	Performance
Epochs	100
Accuracy	94.9%
Precision	95.9%
Recall	96.7%
F1 - Score	96.17%

Table 3: Calculation of mean average precision for overall performance measurement

Number	Tumor Type	Average precision (%)
1	Glioma	85.89
2	Meningioma	91.23
3	Pituitary	94.45
4	No tumor	96.49
Mean Average Precision (%)		92.01

3.1.2. Stroke

The primary evaluation metric used for the stroke dataset was area under the curve (AUC). This metric gives the probability that a model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. The AUC metric was used because it gives both the scores and the plot on how each model performed. The results using the four ML algorithms adopted in this work indicated that decision tree performed best in both situations when the datasets were under sampled and oversampled. Comparison was made to choose the best predictive model using the AUC score as metric and comparing the metric with other metrics to further established how good each model has performed. The higher the score of AUC, the better the model is when predicting stroke cases and no stroke cases.

4.0. DISCUSSION OF RESULTS AND ANALYSIS

4.1. Brain Tumor

From the results obtained from the testing, glioma tumour churned out the least precision of 85.89 % compared to the rest of the samples. The underlying cause was that the glioma tumor was very shallow, and among the brain images most of the scans did not show as clearly a tumor as the other tumor classes (Hu, Li, Yao, and Yao, 2021). The testing results show that most gliomas were predicted as "no tumor" matching our guessing by inspecting the prediction result. Even increasing the layers of the CNN model did not help. The errors persisted after continuously changing and testing the parameters of the model to achieve optimum performance (Hu, Li, Yao, and Yao, 2021, Akinyelu, *et al.*, 2022, and Khairandish, *et al.*, 2022). In all, the metrics obtained proved that our model is very accurate and less susceptible to errors, which is the focus of this work. The model correctly detected and classified all the images in the dataset. In Figure 4, the confusion matrix for the brain tumour detection is shown.

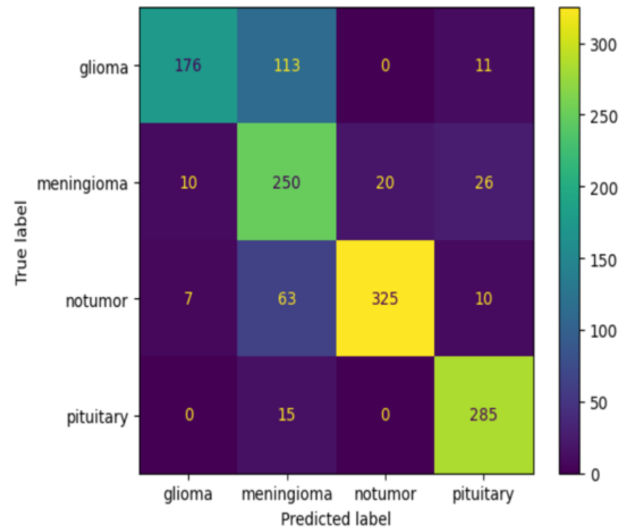

Figure 4: Confusion matrix for brain tumor

4.2. Stroke

We made analysis on our stroke prediction model based on the dataset and the performance metrics used to measure how each algorithm performed. Based on the performance of our model with different datasets that explored using the AUC score to evaluate the performances and pick the overall best model. We observed that on both the original dataset and sampled dataset, the decision tree model performed very well compared to the other models using the AUC score, accuracy, precision, recall and F1- score to compare between them (Jalaja Jayalakshmi, Geetha, and Ijaz, 2021, Mainali, Darsie, and Smetana, 2021). Below is a Table 4 showing all the performance metrics of the algorithms used for this work after hyperparameter tunning.

Table 4: The performance matrix after hyperparameter tuning of stroke images

Methods	Accuracy	Precision	AUC score
Support vector machine	0.9522	0.9881	0.9982
Random Forest	0.8179	0.7562	0.8761
Decision Tree	0.9999	1.0000	1.0000
Logistic Regression	0.8769	0.8943	0.9146

Comparing all these results of how each ensemble machine learning algorithms have performed based on the metric, Decision Tree has the highest AUC score, accuracy, and precision, as shown in Table 6. Hence, we have selected the Decision tree algorithm as the best model for predicting stroke.

4.3. Comparative analysis and evaluation

The findings above demonstrate that the models were able to precisely identify and classify incidences of brain tumor and stroke, demonstrating that the method mostly worked as intended. Easy access to these models for precise prediction and categorization was made possible by the online application. Accuracy, precision, recall, and F1-score were used to assess the performance of our model. We compared our model to current ones. In Puspitasari *et al.*, (2021), analysis of stroke utilizing machine learning algorithms such as decision trees and support vector obtained a 92% accuracy in 2021. A Deep Learning-Based Brain Tumor Classification Using MRI Images was also carried out by (Dipu, Shohan, and Salam, 2021) and an accuracy of 89% was attained. When comparing the performance and analysis of earlier works, it becomes clear that our system outperforms them because earlier works did not account for the use of multiple algorithms to assess the model's performance, and some did not account for the various types and classifications of brain tumors.

4.0 CONCLUSION AND RECOMMENDATIONS

The main objective of this work was to minimize the traditional diagnostic methods for brain tumor and stroke prediction and diagnosis.

These objectives were to:

1. Develop and implement machine learning based data driven models for brain tumor and stroke prediction systems.

2. Utilize machine learning algorithms to analyze stroke instances and identify image patterns for brain tumor prediction.

3.Train models using images and Patients data to improve accuracy and effectiveness of the model in the prediction brain tumor and stroke.

4. Evaluate the performance of the machine learning model in terms of the performance metrics and overall efficiency of the model.

5. Design and develop a web application.

6. Integrate model machine learning with web application for deployment.

The system developed was able to efficiently predict and classify stroke and brain tumor at good performance rates. The results obtained in this work will put health sectors one step ahead through easy and efficient prediction and classification of strokes and tumors. The work can also serve as a diagnostic tool to aid in the advancement of medical research.

However, integration of this model into healthcare systems is recommended to validate the model against real-world patient outcomes. Integrating this model into clinical practice may help healthcare providers make informed decisions regarding early detection and risk assessment of brain tumors and stroke.

REFERENCES

Ahmadvand K.P., "Multispectral MRI image segmentation using Markov random field model," *Signal, Image and Video Processing*. 2016.

Ahmed, H., Abd-El Ghany, S.F., Youn, E.M.G., Omran, N.F., and Ali, A.A., "Stroke prediction using distributed machine learning based on Apache spark," *International Journal of Advanced Science and Technology*, vol. 28, no. 15, pp. 89–97, 2019, doi: 10.13140/RG.2.2.13478.68162.

Akinyelu, A.A., Zaccagna, F., Grist, J.T., Castelli, M., and Rundo, L., "Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey," *Journal of Imaging 2022, Vol. 8, Page 205*, vol. 8, no. 8, p. 205, Jul. 2022, doi: 10.3390/JIMAGING8080205.

Ari, A., and Hanbay, D., "Deep learning-based brain tumor classification and detection system," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, no. 5, pp. 2275–2286, 2018, doi: 10.3906/elk-1801-8.

Arora, S. and Sharma, M., "Deep Learning for Brain Tumor Classification from MRI Images," *Proceedings of the IEEE International Conference Image Information Processing*, vol. 2021-Novem, pp. 409–412, 2021, doi: 10.1109/ICIIP53038.2021.9702609.

"Brain Tumor | Kaggle." https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor. (accessed Sep. 12, 2023).

Chen-Ying Hung, Wei-Chen Chen, Po-Tsun Lai, Ching-Heng Lin, Chi-Chun Lee. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. Annu Int Conf IEEE Eng Med Biol Soc. 2017 Jul; 2017:3110-3113. doi: 10.1109/EMBC.2017.8037515. PMID: 29060556.

Dipu, N.M., Shohan, S.A., and Salam, K.M.A., "Deep Learning Based Brain Tumor Detection and Classification," *2021 International Conference on Intelligent Technologies, CONIT 2021*, pp. 1–6, 2021, doi: 10.1109/CONIT51480.2021.9498384.

El-Feshawy, S.A., Saad, W. Shokair, M., and Dessouky, M., "Brain tumour classification based on deep convolutional neural networks," *ICEEM 2021 - 2nd IEEE International Conference on Electronic Engineering*, pp. 1–5, 2021, doi: 10.1109/ICEEM52022.2021.9480637.

Gurbina, M., Lascu, M., and Lascu, D., "Tumor detection and classification of MRI brain image using different wavelet transforms and support vector machines," *2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019*, pp. 505–508, 2019, doi: 10.1109/TSP.2019.8769040.

Hossain, M.S., Saha, S., Paul, L.C., Azim, R., and Al Suman, A., "Ischemic Brain Stroke Detection from MRI Image using Logistic Regression Classifier," *International Conference on Robotics, Electrical and Signal Processing Techniques*, pp. 763–767, 2021, doi: 10.1109/ICREST51555.2021.9331090.

Hu, H., Li, X., Yao, W., and Yao, Z., "Brain Tumor Diagnose Applying CNN through MRI," *Proceedings - 2021 2nd International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2021*, pp. 430–434, 2021, doi: 10.1109/ICAICE54393.2021.00090.

Jalaja Jayalakshmi, V., Geetha, V., and Ijaz, M.M., "Analysis and Prediction of Stroke using Machine Learning Algorithms," *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2021*, pp. 1–5, 2021, doi: 10.1109/ICAECA52838.2021.9675545.

Khairandish, M.O., Sharma, M., Jain, V., Chatterjee, J.M., and Jhanjhi, N.Z., "A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images," *IRBM*, vol. 43, no. 4, pp. 290–299, Aug. 2022, doi: 10.1016/J.IRBM.2021.06.003.

Lee *et al.*, "Machine learning approach to identify stroke within 4.5 hours," *Stroke*, pp. 860–866, Mar. 2020, doi: 10.1161/STROKEAHA.119.027611.

Mainali, S., Darsie, M.E., and Smetana, K.S., "Machine Learning in Action: Stroke Diagnosis and Outcome Prediction," *Frontiers in Neurology*, vol. 12, Dec. 2021, doi: 10.3389/FNEUR.2021.734345.

Puspitasari, D.I., Riza Kholdani, A.F., Dharmawati, A., Rosadi, M.E., and Mega W., Pradnya Dhuhita, "Stroke Disease Analysis and Classification Using Decision Tree and Random Forest Methods," *2021 6th International Conference on Informatics and Computing, ICIC 2021*, 2021, doi: 10.1109/ICIC54025.2021.9632906.

Sehgal, A., Goel, S., Mangipudi, P., Mehra, A., and Tyagi, D., "Automatic brain tumor segmentation and extraction in MR images," *Conference on Advances in Signal Processing, CASP 2016*, pp. 104–107, 2016, doi: 10.1109/CASP.2016.7746146.

Tusher, A. N., Sadik, M. S., and Islam, M.T., "Early Brain Stroke Prediction Using Machine Learning," *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, pp. 1280–1284, 2022, doi: 10.1109/SMART55829.2022.10046889.

Zhang, D., Gan, Z., and Huang, Z., "Study on classification model of traditional Chinese medicine syndrome types of stroke patients in convalescent stage based on support vector machine," *Proceedings - 10th International Conference on Information Technology in Medicine and Education, ITME 2019*, pp. 205–209, Aug. 2019, doi: 10.1109/ITME.2019.00053.