Feature Engineering Based Human Brain Stroke Symptoms Prediction Using Machine Learning

Arjun Kumar Soni[1], Prof. Shobha Rajak[2], Nagendra Kumar[3]
[1, 3] Dept of CSE
[2]Prof., Dept of CSE
[1, 2, 3] Shri Ram Institute of Science and Technology, Jabalpur, M.P.

Abstract- *Brain stroke is a serious health issue that requires timely and accurate prediction for effective treatment and prevention. This study described a hybrid system that used the best feature selection method and classifier to predict brain stroke. The Stroke Prediction Dataset from Kaggle was used for this study. Synthetic minority over-sampling techniques (SMOTE), random Oversampling, Random Undersampling analysis was used to accomplish class balancing. Accuracy, sensitivity, specificity, precision, and the F-Measure were the main performance parameters considered for investigation. To determine the best combination for predicting brain stroke, the performance of five classifiers, Decision Tree (DT), random forest (RF), Logistic Regression (LR), LDA, and Adaptive Boosting (AdaBoost), was compared along with three data balancing techniques with different combinations. The performance parameters were assessed using k-fold cross-validation.*

Keywords**-** Brain Stroke symptoms, Feature extraction, AdaBoost Classifier, SMOTE, Training and Testing, LDA; Predictive analytics.

I. INTRODUCTION

Stroke is the leading cause of death and a major threat to global health known as cerebrovascular accident (CVA). According to the world health organization (who) definition, stroke is a serious vascular disease or dysfunction that causes paralysis, severe pain, and forgetfulness [1]. Stroke symptoms last more than 24 hours and can result in death within 3 to 10 hours [2]. A study by the world health organization showed that 15 million people have a stroke, and one person dies every 4 to 5 minutes (3). There are two types of strokes: ischemic stroke and hemorrhagic stroke. When the arteries that feed the brain become blocked or clogged due to blood clots, it is called an ischemic stroke and accounts for 87% of all strokes according to the American heart association (aha) [5]. In contrast, a hemorrhagic stroke occurs when blood vessels rupture or bleed, and 15% of strokes are hemorrhagic strokes [5]. The disease is rapidly increasing in developing countries such as China, where the burden is highest [6], while the United States is experiencing disability due to stroke. According to the world health organization, the mortality rate of stroke patients ranks 84th in the world, which is terrible news. Approximately 700,000 people are affected by this disease each year. Over the last few years, researchers have worked with different methods such as prospective studies, case-control studies, and case series and have identified signs of modifiable (genetic diseases, male gender, age) and modifiable risk factors (hypertension, smoking, diabetes) [8]. Figure 1.1 shows the risk factors for ischemic stroke. As the world population increases, the number of deaths and affected people from this disease is expected to increase. However, with early treatment and early diagnosis, this death can be prevented. There are tools such as the cox proportional hazard model that can be used to predict stroke. However, since it is a traditional method, it cannot use high-dimensional data in stroke prediction. In this context, machine learning can play an important role in stroke prediction effectively and efficiently at low cost. Different learning machines such as random forests, support vector machines, decision trees, and logistic regression have been used in clinical practice for many years to perform accurate analysis and predict accurate results due to the large inconsistency of patterns in the data. This study showed the highest results in stroke prediction using data analysis methods, machine learning algorithms including various risk factors, and non-uniform data.

The incidence of stroke is increasing worldwide and is now considered a leading cause of death and disability. Early intervention is critical to preventing long-term disability and death associated with stroke. However, the process of estimating stroke risk is often time-consuming and prone to error. Recently, machine learning algorithms have shown promise in estimating stroke risk based on a variety of clinical risk factors. Using these algorithms, doctors can identify high-risk patients and intervene earlier, reducing the number of stroke complications and improving patient outcomes. There is also a growing need for clarity and explanation of machine learning models in healthcare. Using machine learning models can help inform treatment decisions by providing doctors with valuable information about what puts a patient at risk for stroke. The world stroke organization estimates that 13 million people suffer a stroke each year worldwide, resulting in 5.5

million deaths [10]. Stroke affects all aspects of patients' lives, including family, social environment, and work, and is a leading cause of death and disability worldwide [10], [11]. A common misconception is that stroke only affects certain groups of people, such as the elderly or those with underlying medical conditions. In fact, anyone can be affected, regardless **of** age, gender or physical health. A stroke is a serious interruption of blood flow to the brain, depriving it of oxygen. There are two types: ischemic and hemorrhagic. Medium-grade strokes can cause permanent or temporary damage, depending on their severity. Hemorrhagic strokes are rare, but they are caused by ruptured blood vessels in the brain. This type of stroke usually occurs when blood vessels become blocked or narrowed, preventing blood flow to the brain. Being over 55 years of age, having a previous stroke or tia, irregular heartbeat, high blood pressure, atherosclerosis, smoking, high cholesterol, diabetes, obesity, impotence, medication estrogens, blood clots, carotid artery stenosis due to medications or amphetamines, and heart disease such as heart attack are risk factors for stroke [14], [15], [16]. A stroke can occur suddenly, and its symptoms can be variable and unpredictable. The main symptoms of a stroke include paralysis on one side of the body, numbness in the face, arms, or legs, difficulty speaking or walking, dizziness, blurred vision, headache, vomiting, loss of mouth, and in severe cases, loss of consciousness. These feelings can come on suddenly or gradually, and in some rare cases, make you aware [17], [18], [19]. Stroke affects both men and women, reducing their quality of life and affecting public health. The scientific community has developed advanced models to predict strokes in order to prevent strokes, and artificial intelligence plays an important role in this effort as it is widely used to prevent stroke pain. Many studies have been conducted to improve stroke diagnostic models [20], [21], [22], to predict treatment outcomes and patient responses, and to develop treatment strategies [23], [24], [25]. For example, Arslan et al. [26] Proposed a data mining system using data from 80 ischemic stroke patients and 112 healthy subjects to predict ischemic stroke, and the most accurate support vector machine (svm) classifier reached 97.89% and the auc reached 97.83%. This study also investigates how different factors affect the determination of risk factors for ischemic stroke. Cerebrovascular accident, also known as stroke, is the second leading cause of death and the third leading cause of disability worldwide [27]. Stroke is defined as the sudden death of some cells due to lack of oxygen and is usually asymptomatic. Stroke greatly affects the social and economic life of many countries. For example, according to the American heart association, more than 140,000 people died of stroke in the United States in 2016. This means that 1 in 19 deaths are caused by stroke. The direct and indirect costs of stroke were estimated at $45.5 billion in 2015. Stroke is reported by the world health organization (who) as a growing problem that has received little attention to date [29]. Early diagnosis of the disease has been shown to facilitate prevention and treatment [30]. 1.2 Methods for Stroke Risk Prediction Stroke is considered one of the most serious diseases of today's life. It can cause physical and mental deaths such as hemiplegia, speech disorder (aphasia), ataxia, blindness, forgetfulness, and dementia. According to the 2019 death index published by the world health organization (who) in December 2020, the top 10 causes of death accounted for 55% of the total number of deaths in 2019 (approximately 55.4 million people). Among them, 6 million people died of cerebrovascular disease, which was reported to be the second leading cause of death [32]. The united nations states that if the population aged 65 and over reaches 7 percent or more in a country, that country is classified as elderly; an old life; when the rate is more than 20%, it means that it has entered super old life [33]. For this reason, the social problems of the elderly society have begun to gain importance and the elderly can be defined as segmental. In addition, the aging report of the international credit rating company moody's shows that countries such as Japan, Germany, and Italy have entered the elderly population since 2013, with the elderly accounting for more than 20%. It is reported that as many as 34 countries will enter the elderly age by 2030 [33].

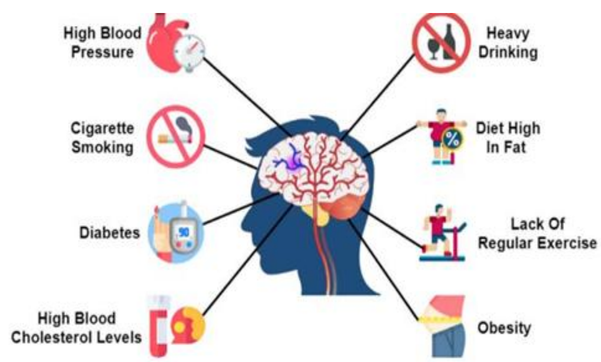

Figure 1.1: Risk factors for ischemic stroke [9].

II. LITERATURE SURVEY

The reliability of ai has been a useful tool for routine diagnosis and medical information processing (ai). Ai is expanded when human vision is limited. Currently, many applications using various machine learning algorithms are used for data processing and innovation in medical research. The use of machine learning techniques has been seen in many new health research applications, including disease screening and cancer diagnosis. Many previous studies have used machine learning in stroke prediction. This section describes the achievements of other researchers in this field. Minhaz et al. [34] also conducted a stroke study where they obtained data from various hospitals in Bangladesh. After the preliminary data, 10 algorithms were used for training. Then a weighting scale is used to improve the performance of each class. All models are then refined, and a weighted index is used to find the best model. The study concluded that the accuracy of the weight measurement is 97%. Yin ya et al. [35] Studied stroke, where the EEG (electroencephalogram) biometric signal during walking can detect stroke. This paper states that random forests can predict stroke using biometrics. In another study, priya et al. [36] Used mining tools and machine learning algorithms to predict stroke. The authors used 14 classification methods including simple tree, median tree, complex tree, logistic regression, linear svm, quadratic svm and neural network. Through this test, nn achieved 95.3% accuracy compared to others. Selasia et al. [37] Considered different data from Kaggle and performed the prior data including missing value handling, label coding and unequal data. After the prior data, six machine learning algorithms were applied to these data. After comparing the accuracy of these algorithms, naive bayes classification gave the highest accuracy of 82%. They also created an html page where the user can get the results by giving some parameters regardless of whether he/she has a stroke or not. Hager et al. [38] used four classification methods to predict stroke: logistic regression, random forest, decision tree, and support vector machine. They use hyperparameters tuning and cross validation in machine learning algorithms to obtain their results. Then they evaluated the performance of four models and found that random forest has the highest accuracy of 90% among the four models. Wu et al. [39] proposed a stroke prediction model with imbalanced data. In this study, they collected random data from the China longitudinal health and longevity study and successfully compared the data using random oversampling, random Undersampling (rus) and smote techniques. In this study, the authors used continuous logistic regression, support vector machine and random forest models to predict strokes in univariate data as well as balanced data where the best result was compared with a dataset. When compared, they show that svm and lr can achieve up to 95% accuracy on random data, but the sensitivity is. Badria et al. [40] Collected ct scan data from stroke patients in hajj hospital in Surabaya, Indonesia. In the first stage, image operations such as data transformation, cropping, scaling, grayscale and reliable data are performed here to improve the image quality. In this experiment, random forest achieved the highest accuracy of 95.97% compared to other classifications. Jaehak et al. [41] Also conducted a stroke study where they used real-time bio signals of ai to predict stroke. The system uses (machine learning) and short (deep learning) algorithms, and lstm achieves the best results (98.58%). Yin a et al. [42] published another report using eeg data and deep learning models, where lstm showed the best results with 94% accuracy.

III. PROPOSED WORK

We propose the detailed design process will be described in the subsequent sections.

In the following section, we will have a look at some of the main steps of a typical machine learning task, and the diagram below should give us an intuitive understanding of how they are connected.

1. Preparation of dataset the dataset is downloaded from kaggle. It has following task to peform:

- Data Manipulation.
- Missing Values Handling.
- Feature Generation.
- Dimensionality Reduction.
- Outlier Removal.
- Normalization.
- Partitioning.

2. Model Training

- Model Selection.
- Hyper Parameter Optimization.
- Model factory.

3. Model Optimization.

- Parameter Tuning.
- Parameter Optimization.
- Model Size.
- No iterations.

4. Model Evaluation

- Performance Measures.
- Accuracy.
- Precision.
- Recall.
- ROC_AUC score.

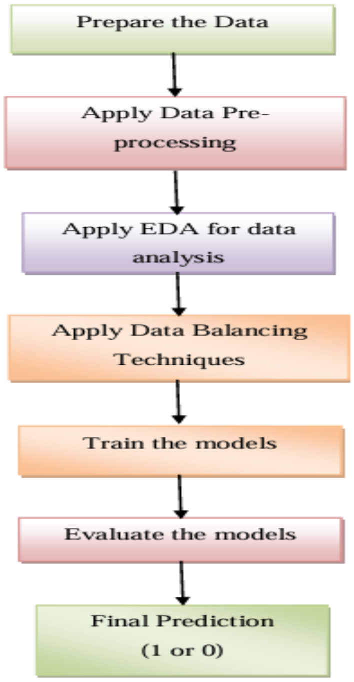

Figure 3.1: Proposed work.

4.2 Proposed Algorithm:
Step 1: Read the dataset of stroke related information.
Step 2: Clean the dataset.
Step 3: Categorize data into numerical and Categorical columns. Step 4: Apply Label encoding on Categorical columns.
Step 5: Normalize all data into a predefined range.
Step 6: Split the dataset into train_set and test_set.
Step 7: Train the model using various classifiers.
Step 8: Evaluate the models for metrics.
Step 9: Apply Grid Search and optimize the models.
Step 10: Again, train and test the models with optimized models. Step 11: Apply SMOTE and rebalance the dataset.
Step 12: Train and test the models to get more accurate results.
Step 13: Compare the performance of all the Classifiers.
Step 14: Exit.

IV. RESULTS

Implementation Detail:

Dataset To proceed with the implementation, different datasets were considered from Kaggle. This has 5110 rows and 12 columns. The columns have'id','gender','age','hypertension’,heart_disease','ever_married','work_type','residence type', 'avg_glucose_level', 'bmi', 'smoking_status' and 'stroke' as the main attributes. The output column 'stroke' has the value as either '1' or '0'. The value '0' indicates no stroke risk detected, whereas the value '1' indicates a possible risk of stroke. This dataset is highly imbalanced as the possibility of '0' in the output column ('stroke') outweighs that of '1' in the same column. Only 249 rows have the value '1' whereas 4861 rows with the value '0' in the stroke column. For better accuracy, data pre-processing is performed to balance the data.

The dataset is taken from Kaggle in the form of csv format. The snapshot of the dataset used is shown below in figure 6.2.

```
stroke.head(10)
```

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
5	56669	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1
6	53882	Male	74.0	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
7	10434	Female	69.0	0	0	No	Private	Urban	94.39	22.8	never smoked	1
8	27419	Female	59.0	0	0	Yes	Private	Rural	76.15	NaN	Unknown	1
9	60491	Female	78.0	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1

Figure 4.1: Dataset snapshot. The dataset has various columns like id, gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, smoking_status and stroke.

Classifier	Accuracy	Precision	Recall	ROC-AUC
LR(Scaled)	93.933464	0.000000	0.000000	0.500000
LR	94.031311	100.000000	1.612903	0.508065
LR(SMOTE)	75.989583	76.180483	75.625000	0.759896
Adaboost	93.737769	0.000000	0.000000	49.895833
Adaboost (Scaled)	93.737769	0.000000	0.000000	49.895833
RF	93.933464	0.000000	0.000000	50.000000
RF(Scaled)	93.933464	0.000000	0.000000	50.000000
RF(Grid)	93.933464	0.000000	0.000000	50.000000
LR(Grid)	93.933464	0.000000	0.000000	50.000000
DT(Grid)	93.737769	33.333333	3.225806	51.404570
LDA (Scaled)	93.444227	27.272727	4.838710	52.002688
LDA	93.444227	27.272727	4.838710	52.002688
DT(Scaled)	91.095890	20.408163	16.129032	56.033266
DT	91.878669	25.581395	17.741935	57.204301
LDA(SMOTE)	75.989583	75.176589	77.604167	75.989583
DT(SMOTE)	80.104167	88.328912	6.937500	80.104167
RF(SMOTE)	80.625000	95.230769	64.479167	80.625000
Adaboost(SMOTE)	86.145833	85.773196	86.666667	86.145833

Figure 4.2: Combined Result of all the Classifiers.

V. CONCLUSION

The application of Machine Learning (ML) for stroke prediction has demonstrated its potential as a valuable tool in the early detection and risk stratification of individuals at risk of stroke. Machine Learning models: have shown powerful statistical results in distinguishing between stroke and non-stroke patients based on multiple input features, such as demographic information, medical history, and clinical factors (e.g., hypertension, diabetes, cholesterol levels). Key findings and contributions from studies using Machine Learning for stroke prediction include: 1. Effectiveness in Dimensionality Reduction: Normalization has been shown to perform well in reducing the complexity of high-dimensional data while retaining the most relevant information for classification, making it useful in medical diagnostics where numerous factors contribute to stroke risk. 2. Interpretability: One of the key advantages of machine learning classifiers are its simplicity and interpretability, which is crucial in healthcare settings. The results from various classifiers model can be understood in terms of which variables (e.g., blood pressure, age, and smoking status) contribute most to the likelihood of a stroke, providing actionable insights for clinicians. 3. Classification Accuracy: The accuracy is being calculated for various classifiers. After balancing the dataset, some classifiers have reduced the accuracy scores but there is a significant improvement in other parameters like precision, recall and ROC-AUC score. Finally, the AdaBoost classifiers have better results in terms of all parameters as compared to other models.

REFERENCES

[1] Valliani, A., & Ranti, D., & Oermann, E. (2019) Deep learning and neurology: A systematic review, Neurology and Therapy, Springer, vol. 8, issue. 2, pp. 351–365.

[2] A. Charidimou, A. Krishnan, D. J. Werring, and H. R. Jager, “Cerebral microbleeds: a guide to detection and clinical relevance in different disease settings,” Neuroradiology, vol. 55, no. 6, pp. 655–674, 2013.

[3] A. Charidimou and D. J. Werring, “Cerebral microbleeds: detection, mechanisms and clinical challenges,” Future Neurology, vol. 6, no. 5, pp. 587–611, 2011.

[4] M. Vernooij, A. van der Lugt, M. A. Ikram, P. Wielopolski, W. Niessen, A. Hofman, G. Krestin, and M. Breteler, “Prevalence and risk factors of cerebral microbleeds the rotterdam scan study,” Neurology, vol. 70, no. 14, pp. 1208–1214, 2008.

[5] C. Cordonnier, R. A.-S. Salman, and J. Wardlaw, “Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting,” Brain, vol. 130, no. 8, pp. 1988–2003, 2007.

[6] A. Charidimou and D. J. Werring, “Cerebral microbleeds and cognition in cerebrovascular disease: an update,” Journal of the neurological sciences, vol. 322, no. 1, pp. 50–55, 2012.

[7] Z. Wang, Y. O. Soo, and V. C. Mok, “Cerebral microbleeds are antithrombotic therapy safe to administer? Stroke, vol. 45, no. 9, pp. 2811–2817, 2014.

[8] M. Akter, T. Hirai, Y. Hiai, M. Kitajima, M. Komi, R. Murakami, H. Fukuoka, A. Sasao, R. Toya, E. M. Haacke et al., “Detection of hemorrhagic hypointense foci in the brain on susceptibility-weighted imaging: clinical and phantom studies,” Academic radiology, vol. 14, no. 9, pp. 1011–1019, 2007.

[9] J. D. Goos, W. M. van der Flier, D. L. Knol, P. J. Pouwels, P. Scheltens, F. Barkhof, and M. P. Wattjes, “Clinical relevance of improved microbleed detection by susceptibility-weighted magnetic resonance imaging,” Stroke, vol. 42, no. 7, pp. 1894–1900, 2011.

[10] S. M. Greenberg, M. W. Vernooij, C. Cordonnier, A. Vishwanathan, R. Al-Shahi Salman, S. Warach, L. J. Launer, M. A. Van Buchem, and M. Breteler, “Cerebral microbleeds: a guide to detection and interpretation,” The Lancet Neurology, vol. 8, no. 2, pp. 165–174, 2009.

[11] S. Gregoire, U. Chaudhary, M. Brown, T. Yousry, C. Kallis, H. Jager, and D. Werring, “The microbleed anatomical rating scale (MARS) reliability of a tool to map brain microbleeds,” Neurology, vol. 73, no. 21, pp. 1759–1766, 2009.

[12] A. Fazlollahi, F. Meriaudeau, V. L. Villemagne, C. Rowe, P. Yates, O. Salvado, P. T. Bourgeat et al., “Efficient machine learning framework for computer-aided detection of cerebral microbleeds using the radon transform,” in Proceedings of the IEEE-ISBI conference, 2014.

[13] H. J. Kuijf, J. de Bresser, M. I. Geerlings, M. Conijn, M. A. Viergever, G. J. Biessels, and K. L. Vincken, “Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform,” NeuroImage, vol. 59, no. 3, pp. 2266–2273, 2012.

[14] W. Bian, C. P. Hess, S. M. Chang, S. J. Nelson, and J. M. Lupo, “Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images,” NeuroImage: Clinical, vol. 2, pp. 282–290, 2013.

[15] B. Ghafaryasl, F. van der Lijn, M. Poels, H. Vrooman, M. A. Ikram, W. J. Niessen, A. vander Lugt, M. Vernooij, and M. de Bruijne, “A computer aided detection system for cerebral microbleeds in brain MRI,” in Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on. IEEE, 2012, pp. 138–141.

[16] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, “A convolutional neural network cascade for face detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp.5325–5334.

[17] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[18] D. Loyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. D. Simone, T. B. Ferguson, E. Ford, K. Furie, C. Gillespie et al., “Heart disease and stroke statistics 2010 update a report from the American heart association,” Circulation, vol. 121, no. 7, pp. 46–215, 2010.

[19] T. Thom, N. Haase, W. Rosamond, V. J. Howard, J. Rumsfeld, T. Manolio, Z.-J. Zheng, K. Flegal, C. Odonnell, S. Kittner et al., “Heart disease and stroke statistics 2006 update a report from the American heart

association statistics committee and stroke statistics subcommittee," Circulation, vol. 113, no. 6, pp. e85–e151, 2006.

[20] H. Masoumi, A. Behrad, M. A. Pourmina, and A. Roosta, "Automatic liver segmentation in mri images using an iterative watershed algorithm and artificial neural network," Biomedical Signal Processing and Control, vol. 7, no. 5, pp. 429–437, 2012.

[21] X. Luo, Q. Hu, X. He, J. Li, J. S. Jin, and M. Park, "Automatic liver parenchyma segmentation from abdominal ct images using support vector machines," in Complex Medical Engineering, 2009. CME. ICME International Conference on. Arizona, USA: IEEE, 2009, pp. 1–5.

[22] A. C. Kak and M. Slaney, Principles of computerized tomographic imaging. New York, N.Y.: IEEE press, 2001.

[23] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," Medical Image Analysis, vol. 14, no. 3, pp. 390–406, 2010.

[24] N. Farid, B. M. Elbagoury, M. Roushdy, and A.-B. M. Salem, "A comparative analysis for support vector machines for stroke patients," in WSEAS Proceedings of the 7th European Computing Conference, 2013, pp. 71–76.

[25] T. Mroczek, J. W. Grzymała-Busse, Z. S. Hippe, and P. Jurczak, A Machine Learning Approach to Mining Brain Stroke Data. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 147–158.

[26] O. Maier, C. Schroder, N. D. Forkert, T. Martinetz, and H. Handels, "Classifiers for ischemic stroke lesion segmentation: A comparison study," PLoS ONE, vol. 10, pp. 1–16, 2015.

[27] M. Havaei, N. Guizard, H. Larochelle, and P. Jodoin, "Deep learning trends for focal brain pathology segmentation in MRI," arXiv, vol. abs/1607.05258, 2016. [Online]. Available: http://arxiv.org/abs/1607.

[28] G. Altan, Y. Kutlu, and N. Allahverdi, "Deep belief network based brain activity classification using EEG from slow cortical potentials in stroke," in Proceedings of the International Conference on Advanced Technology & Sciences, 2016, pp. 233–239.

[29] Lee, H., Yune, S., Mansouri, M., Myeongchan, K., Tajmir, S., Guerrier, S., Ebert, S., Pomerantz, S., Romero, J., Kamalian, S., Gonzalez, R., Lev, H., & Do, S.(2019) An explainable deep-learning algorithm for the detection of acute intracranial hemorrhage from small datasets, Nature Biomedical Engineering, vol. 3, pp. 173–182.

[30] Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N., Venugopal, V., Mahajan, V., Rao, P., Warier, P.(2018) Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study, The Lancet, Elsevier, vol. 392, issue 10162, pp. 2388-2396.

[31] Grewal, M., Srivastava, M. M., Kumar, P. & Varadarajan, S.(2018) RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans, IEEE International Symposium on Biomedical Imaging, pp. 281–284.

[32] Desai, V., Flanders, A. E. & Lakhani, P. (2017) Application of deep learning in neuroradiology: automated detection of basal ganglia hemorrhage using 2D-convolutional neural networks, arXiv, website. arxiv.org/ftp/arxiv/papers/1710/1710.03823.pdf.

[33] G. J. S. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[34] M.U. Emon, M.S. Keya, T.I. Meghla, M.M. Rahman, M.S. Al Mamun, M.S. Kaiser, Performance analysis of machine learning approaches in stroke prediction, in: Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020, no. January, 2020, pp. 1464–1469, http://dx.doi.org/10.1109/ICECA49313.2020. 9297525.

[35] Y.A. Choi, et al., Machine-learning-based elderly stroke monitoring system using electroencephalography vital signals, Appl. Sci. 11 (4) (2021) 1–18, http://dx. doi.org/10.3390/app11041761.

[36] P. Govindarajan, R.K. Soundarapandian, A.H. Gandomi, R. Patan, P. Jayaraman, R. Manikandan, Classification of stroke disease using machine learning algorithms, Neural Comput. Appl. 32 (3) (2020) 817–828, http://dx.doi.org/10.1007/ s00521-019-04041-y.

[37] G. Sailasya, G.L.A. Kumari, Analyzing the performance of stroke prediction using ML classification algorithms, Int. J. Adv. Comput. Sci. Appl. 12 (6) (2021) 539–545, http://dx.doi.org/10.14569/IJACSA.2021.0120662.

[38] H. Ahmed, S.F. Abd-El Ghany, E.M.G. Youn, N.F. Omran, A.A. Ali, Stroke prediction using distributed machine learning based on apache spark, Int. J. Adv. Sci. Technol. 28 (15) (2019) 89–97, http://dx.doi.org/10.13140/RG.2.2.13478. 68162.

[39] Y. Wu, Y. Fang, Stroke prediction with machine learning methods among older chinese, Int. J. Environ. Res. Public Health 17 (6) (2020) 1–11, http://dx.doi. org/10.3390/ijerph17061828.

[40] T. Badriyah, N. Sakinah, I. Syarif, D.R. Syarif, Machine learning algorithm for stroke disease classification, in: 2020 International Conference on Electrical, Communication, and Computer Engineering, ICECCE, IEEE, 2020, pp. 1–5.

[41] J. Yu, S. Park, S.H. Kwon, C.M.B. Ho, C.S. Pyo, H. Lee, AI-based stroke disease prediction system using real-time electromyography signals, Appl. Sci. 10 (19) (2020) http://dx.doi.org/10.3390/app10196791.

[42] Y.A. Choi, et al., Deep learning-based stroke disease prediction system using real-time bio signals, Sensors 21 (13) (2021) http://dx.doi.org/10.3390/ s21134269.