Application of Machine Learning in Stroke Prediction: A Systematic Review

Shah Khushbu
Department of Computer Scienceand Engineering
Parul Institute of Engineering and Technology, Parul University
Vadodara, India
2203032010014@paruluniversity.ac.in

Amit Ganatra
Parul University
Vadodara, India
provost@paruluniversity.ac.in

Chintan Thacker
Department of Computer Scienceand Engineering
Parul Institute of Engineering and Technology, Parul University
Vadodara, India
chintan.thacker19435@paruluniversity.ac.in

2024 Parul International Conference on Engineering and Technology (PICET) | 979-8-3503-6974-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/PICET60765.2024.10716162

***Abstract*—A brain stroke is a dangerous condition in which there is insufficient blood flow to a part of the brain, frequently as a result of brain haemorrhage or clogged arteries. Long-term oxygen deprivation results from this, which damages the brain irreversibly and kills brain cells. Stroke patients should receive treatment as soon as possible since restoring blood flow might lessen the extent of the damage. In the world, stroke ranks as the second most common cause of death. Algorithms for machine learning can assist in both patient rehabilitation and the detection of brain strokes.Large datasets are being analysed using ML techniques to spot trends that might suggest a person's risk of having a stroke. This is a growing area of use in stroke prediction. Key methods include NB, SVM, GB, RF, DT, KNN, Clustering, and Ensemble Learning. By combining predictions from several trees, these methods can help pinpoint important risk factors, understand how variables interact, and increase model accuracy. Subgroups and patterns in the data are also made easier with their assistance. Using the stack ensemble model, which employedthree 1D convolutional models using neural networks from the survey done on the ECG dataset, the maximum accuracy of 99.70% was achieved..**

***Index Terms*—Artificial Intelligence, Machine Learning, Deep Learning, Stroke prediction.**

I. INTRODUCTION

Stroke emergencies are medical emergencies. In the US, it ranks as the fifth most frequent cause of death [1]. Over 795,000 Americans suffer from a stroke each year, of which 610,000 are first-time victims and 185,000 are repeat victims [2]. There are two different kinds of stroke: hemorrhagic stroke, which is caused by bleeding, and ischemic stroke, which is caused by an obstruction in a brain blood vessel[3]. The most prevalent kind of stroke (87%) is ischemicstroke [4]. Symptoms can vary depending on which area ofthe brain is impacted; these can include headaches, dizziness, weakness, numbness, altered speech, confusion, and visualabnormalities [5].

Hemorrhagic and ischemic strokes are neurological disorders caused by brain injury. Stroke is a serious illness, causing death and mental and physical complications [6]-[10].Stroke has been described as one of the most serious illnesses in the modern world since, in extreme situations, it can result

in death as well as mental and physical complications like paralysis and trouble speaking. [11].In 2019, 55.4 million fatalities were attributed to strokes. Ageing societies are defined by a 7-14% population age, with 14% being elderly and 20% being super-aged. [13]. This study explores brain stroke prediction using deep learning, artificial intelligence, and machine learning algorithms.NN, DT, and LR are a few examples of machine learning techniques that are essential for stroke prediction. They boost model performance, decrease overfitting, increase accuracy, and pinpoint trends and risk factors. In the end, these methods enhance patient care and stroke prediction by increasing interpretability, transparency.In order to determine stroke risk, discover early warning signals, customise risk assessments to each patient's unique profile, and increase accuracy using sophisticated models like ensemble methods and NN, ML techniques analyse large, complicated datasets.With the focus of healthcare resourceson high-risk patients and the reduction of needless tests and procedures, machine learning (ML) techniques improve model performance and explain factors that contribute to stroke risk. They also lower costs and integrate with remote monitoring.Various techniques, including DL, time-series analysis, transfer learning, and NLP, are being utilized to analyze medical imaging data for stroke risk prediction.

II. RELATED WORK

Leonardi et. al [14] used explainable artificial intelligence for trace classification of stroke patients, using CNN classifiers, Adversarial Autoencoders, and trace saliency maps. They achieved a maximum accuracy of 82% with 200 epochs.

Kunwar et al. [15] used the stack ensemble model to automatically predict strokes. The stacked ensemble model's test results indicate a 99.70% accuracy rate. The study's limitations include the need that the suggested model be tested on bigger ECG data sets in further research to confirm its applicability.

To predict stroke, Dev et al. [16] employed a predictive analytics approach. Neural networks and ML algorithms come in several varieties. RF, DT, and NN are utilized for classification. Neural networks provide the best accuracy of 78% among them when paired with a significant set of feature combinations. The observed miss rate was 19%. The suggested method will be externally validated as part of their future work.

Using a stroke prediction dataset, Ahmed et al. [17] have used a variety of ML classifiers. A decision tree, Support Vector Machines (SVM), RF and Logistic Regression classifiers are some of the classifiers available. The random forest classifier yielded the best accuracy of 96%, followed by the decision tree with 82%, throughout the evaluation of all the classifiers. SVM and LR both produced results with 81% accuracy.

Tazin et al. [18] employed a number of ML methods, including voting classifiers, DT classifiers, RF, and LR. Given how substantially unbalanced the dataset is, the SMOTE technique is used to balance it. The classifier with the highest accuracy among the ones used is the RF, which yields 96%. DT come in second with 94%, voting classifiers come in third with 91%, and LR yield 79%

Islam et al.'s [19] study looked into the usage of XAI for quick patient diagnosis. The study attempted in predicting strokes in the brain using EEG signal information from patients having stroke and those who does not have strokein different circumstances using XAI and ML models. The latter achieved 80% accuracy by using the adaptive gradient boosting, Xgboost, and LightGBM features of the algorithm.

Dritsas et al. carried out early stroke identification [20]. The authors employed several ML algorithms. Numerous techniques were employed, such as stacking, SGD, MLP, KNN, LR, RF, and NB. When the brain's blood supply is abruptly cutoff, a stroke occurs. 3254 people who were atleast 18 years old were included in the datasets used to assess these models' accuracy. Among these, the stacking method had the highest accuracy at 80%.

Effective techniques for predicting strokes include explainable AI and ML. It causes early mortality and serious economic consequences, such as 12 billion euros in productivity losses in Europe in 2017 and 27 billion euros in predicted healthcare costs [21]. The efficacy of the proposed ML method was assessed by comparing it to six popular classifiers. The six popular classifiers are as follows: SVM, RF, XGBoost, KNN, MLP, and LR.

Islam et al.'s research [22] examined the application ofML methods for prediction of stroke. The study assessed the prediction of stroke risk using the DT, KNN, SVM, Adaboost, voting classification algorithm, Artificial Neural Network (ANN), RF, and XGBoost algorithms. Stroke accounted for 13% of the total mortality in 2016. Of the 8600 patients in the stroke datasets used in the study, 2500 had suffered a stroke. They were gathered from several Bangladeshi hospitals. Among all the algorithms used, voting classifier attained maximum accuracy of 98% with 3% FN rate and 2% FP rate.

The Darabi et al. study aimed to identify individuals at higher risk for specialized care after ischemic stroke, using machine learning techniques and patient-level data to create 15 models for readmission prediction, reducing annual costs.The research examined the efficacy of ML techniques, such as LR, RF, XGBoost, SVM, and Gradient Boosting Machine.[23]

ML is used by Choi et al. [24] to assess the prediction ability of DT and improve their comprehension of the variables involved in stroke modelling. The study employed two algorithms to create decision trees: Cart and ID3. Given that the total rate of accuracy is 0.981, a 0.019 error rateis inferred. A stroke was correctly predicted in 98.17% of patients, but only 16.67% of those who were expected to have one were correctly identified.

Harshitha et al. used DT, LR, RF, SVM, KNN, achieving 95% accuracy. RF was the most effective method with the lowest FN rate.

Saleh et al. [17] assessed the effectiveness of distributed ML approaches for stroke prediction using the Healthcare Dataset Stroke. Four classification techniques were employed, including DT, SVM, LR, and RF Classifier. In comparison to the other models, the RF Classifier outperformed them all, achieving 90% accuracy in terms of Accuracy.

The article [26] offers insightful information on the significance of making sure that medical picture analysis using deep learning algorithms is transparent and easy to understand for both patients and medical professionals. The challenges and possible biases of DL techniques in the processing of medical images are discussed by the authors of the research [26].

Stroke risk prediction (SRP) using Hybrid deep transfer learning framework was carried out by authors Chen et al.[27]. In actual experiments, the model's maximum accuracy was 78%. The accuracy attained in artificial experiments is 72%. Future work will involve integrating the system of other diseases and expanding the NWT framework to consider numerous chronic illnesses at the same time.

Stroke risk prediction using interpretable classifier was performed by Penafiel et al. [28]. The authors opted for the Dempster-Shafer using Gradient Descent Classifier (DSDG)which makes use of Dempster-Shafer theory and gradient descent classifier. DSDG achieved 85.4% accuracy and 83.8% AUC.

Yu et al. [29] performed stroke prediction system based on Artificial Intelligence that can make prediction and semantically interpret stroke prognostics systems. According to the trial results, while senior people were walking, C4.5 DThad a prediction accuracy of 91.56% and RF had a prediction efficiency of 97.51%. Furthermore, the CNN LSTM model with raw ECG and PPG data showed a respectable 99.15% accuracy in predicting. Consequently, actual-time predictions made by senior stroke patients showed good performance and accuracy.

In the work by Lu et al. [30], a unique eye-tracking technology task with 5 color stimuli in TCM was usedto recognize strokes.They chose six exceptionally effective supervised algorithms: DT, XGBoost, KNN, RF, CatBoost, and gradient boosting classification algorithm (GBC).With Precision of 86.48%, and eye-movement algorithm trained by RF, it was determined to be the best ML model for the red stimulus. The eye-movement model's rate of accuracy using the CatBoost, RF, KNN algorithms were greater than 87%.

ML approaches were utilized by Mridha et al. [31] for automated prediction of stroke. In the study, ML methodssuch as SVM, KNN, NB, RF, LR, and XGB classifier were employed. The experiment's findings demonstrated that simpler ones were outperformed by more sophisticated models, with RF coming in first place and attaining an accuracy of roughly 91%, while the other models' accuracy ranged from 83 to 91%.

Srinivas et. al[32] introduced a brain stroke detection model that combined predictions from multiple methods using an ensemble machine learning classifier based on soft voting. With swarm intelligence to be used in future optimisation, the model's accuracy was 96.88%.

A hybrid system was designed using five different classifiers to predict stroke in article by Bathla et al. [33]. The study's data source was the Kaggle Stroke Prediction Dataset. The five classifiers used were SVM, NB, RF, Adaboost, and XGBoost. The feature reduction ratio offered by FI was 36.3%. The combined system that predicted brain strokes most accurately with accuracy of 97.17% employed RF as the classifier and FI as the feature selection method

Using MRI-VPD with automatic carotid plaquesegmentation and deep learning-based feature classification, Wang et al. [34] assessed the risk of stroke. The approachthat was used involved segmenting the plaque pictures and classification using trained ResNet-50 models after the U-Net model, from the MRI scans, automatically retrieved the regionof interest. Using their suggested framework, they were able to get the greatest accuracy of 94.11%, demonstrating theeffectiveness of their strategy. [34]

Mak et al. in [35] tried with the help of augmented reality (AR) and EEG, they anticipate that to create a system that can detect the existence of spatial neglect in visual perception and depict the estimated peak of the neglected visual field. According to their first findings, the method has the potential to accurately identify SN and forecast visual target reactionsin individuals with stroke who have SN.

Alruily et al. [36] made use of combined ML methodsto forecast cerebral stroke.. The likelihood of a cerebral stroke was predicted using earlier mentioned dataset usingthe ML methods RF, XGBoost, and light gradient-boosting machine.All of the ML algorithm's performance measures showed encouraging outcomes after the hyper-parameters were tuned. The accuracy obtained by the algorithm is 96.34%

III. METHODOLOGY

Stroke prediction makes use of various algorithms and methods. Dataset plays important role in stroke prediction. There are various methods that makes use of ML, artificial intelligence and deep learning techniques.

A ML method called ensemble learning mixes several models to enhance prediction performance. Boosting, stacking, and bagging are common varieties. While boosting concentrates on training weak learners and merging them into a strong learner, bagging applies the same base learning algorithm to numerous instances on distinct data subsets. On the other hand, stacking captures the complimentary qualities of various models by training a meta-model to integrate the predictions of multiple base models.

A. Logistic Regression

Logistic regression is a statistical method used to analyze datasets in which the outcome is influenced by a number of independent variables the outcome variable is known when the logistic regression model is trained on a labelled dataset. [17], [18], [20], [21], [23], [25], [31].

B. K-Nearest Neighbor (KNN)

KNN is a is a ml method that fall under supervised ml for classification and regression.[20], [21], [22], [25],[30], [31] In KNN classification tasks, KNN determines the class of a latest data point by casting a majority vote. KNN employs the average of the KNN target values as the prediction for a new data point when doing regression work.

C. Support Vector Machine (SVM)

A supervised ML approach called SVM is utilized for classification and regression problems. SVM is very useful for binary classification, but it can also be used to solve multi-class issues. Due to its reliance on the nearby data points (support vectors), this hyperplane is referred to as a "support vector". [17], [21], [22], [23], [25], [31].

TABLE I. SUMMARY OF STROKE PREDICTION METHODS

Authors	Year	Technologies used	Dataset used	Results
Leonardi et. al [14]	2022	CNN classifiers,Adversarial Autoencoders, trace saliency maps.	Italian Stroke Center	Accuracy 82%
Kunwar et. al [15]	2023	Stack ensemble model make use of three 1D convolutional models with neural networks	ECG Dataset	Accuracy 99.70 %.
Dev et al. [16]	2022	ML algorithms like DT and RF, and NN	Kaggle,5110 instances	Accuracy 78%
Ahmed et. al [17]	2019	ML classifiers: LR, DT, SVM, RF classifier	Kaggle dataset	Accuracy 96%,
Tazin et. al [18]	2021	ML techniques such as DT, RF, LR, classifier, voting classifier	Kaggle dataset	Accuracy 96%
Islam et al. [19]	2022	XAI and ML models,adaptive gradient boosting, Xgboost, and LightGBM, XAI techniques used were Eli5and LIME.	EG Dataset CA, USA.	Accuracy 80%
Dritsas et al. [20]	2022	RF, NB, LR, KNN, stacking, SGD, and MLP.	Kaggle Dataset	Accuracy 80%
Kokkotis et al. [21]	2022	LR, RF, XGBoost, KNN,SVM, MLP	Stroke prediction dataset	Total FN rate 18.60%

Islam et al. [22]	2021	DT, KNN, SVM, Adaboost, SGD, Voting classifier, ANN, RF and XGBoost classifier.	Stroke disease dataset is collected from the different hospitals in Bangladesh	Accuracy 98%
Darabi et al. [23]	2021	LR, RF, XGBoost, SVM, Gradient boosting machine	EHR Dataset from Geisinger Health System	AUC 0.66
Choi et al. [24]	2020	Cart (Classification and Regression Tree) and ID3, Decision tree	Stroke Prediction Dataset	Accuracy 98.17%
Harshitha et al. [25]	2021	DT, LR, RF, SVM, and KNN.	Kaggle dataset	Accuracy 95%.
Saleh et al. [17]	2019	DT, SVM, LR, and RF Classifier.	Kaggle dataset	Accuracy90%
Chen et al. [27]	2022	HDTL-SRP, DNN	Dataset collected from various hospitals	Accuracy 78%.
Penafiel et al. [28].	2021	Dempster-Shafer using Gradient Descent Classifier (DSDG)	Dataset collected from Japanese workers checkup	Accuracy 85.4%
Yu et al. [29]	2022	CNN and LSTM	Raw collected data of PPG and ECG	Accuracy 99.15%
Lu et al. [30]	2023	RF, KNN, DT, gradient boosting classifier (GBC), XGBoost, and CatBoost.	Data collected from Xi'an Jiaotong University	Precision 86.48%
Mridha et al. [31]	2023	RF, LR, SVM, KNN, NB, and XGB classifier	Kaggle Dataset	Accuracy 91%
Srinivas et al. [32]	2023	RF, Extremely Randomizeds, and Histogram-Based Gradient Boosting	Stroke prediction dataset	Accuracy 96%
Bathla et al. [33]	2023	SVM, RF, NB, XGBoost And Adaboost	Stroke prediction dataset	Accuracy (95%)
Wang et al. [34]	2023	U-net and ResNet-50	MRI image dataset	Accuracy 94.11%
Alruily et al. [36]	2023	LightGBM, RF, XGBoost	Stroke prediction dataset	Accuracy 96.34%.

D. *Random Forest:*

The RF method is regarded as being extremely strong and robust, user-friendly. In comparison to a single decision tree, it is also less vulnerable to overfitting. [16],[17], [18], [20], [21], [23], [25], [30], [31] Different subsets of the data are used for each tree's training by a collection of decision trees that is called a random forest classifier. The end result is basedon the agreement of every individual tree in the forest, eachof which came up with a prediction.

E. *Naive Bayes*

Although it can be modified for various purposes like text creation or recommendation systems, the probabilistic ML method Naive Bayes is generally used for classification tasks. It makes the assumption of feature independence and is deriving from the Bayes theorem, which makes the modelling process easier. The approach computes the odds of each characteristic occurring inside each class during training. [20],[31].

F. *Decision Tree*

Using decision trees is a popular ML technique for re-gression and classification applications. They are useful tools for data analysis and interpretation because they are intuitive, simple to comprehend, and can be visually represented. Adecision tree is a hierarchical model for displaying a seriesof choices and their outcomes. It has a tree- like structure, with roots (decisions) leading to leaves (results or predictions).[16], [17], [18], [22], [24], [25], [30].

G. *Voting Classifier*

A voting classifier is a ML model that combines the outcomes of various separate classifiers (or models) to arriveat a single outcome. [18], [22].

IV. DATASETS MATERIALS

A. *Stroke Prediction dataset*

The Kaggle dataset has been analyzed in numerous studies [39]. They concentrated on using this dataset's over-18 individuals. The authors in [16], [17], [18], [20], [25], [30] have used stroke prediction dataset for the prediction of stroke.

1) *Gender [40]:* The characteristic relates to the genderof the person. Men make up 1260 of the total, while women make up 1994.
2) *Age (years) [40]:* The participants who are over the age of 18 are identified by this attribute.
3) *Hypertension [41]:* It indicates whether or not the subject has hypertension. 12.54% of the participants have high blood pressure.
4) *Heart disease [42]:* It indicates whether or not the person has heart disease. Participants who have heart illness make up 6.33% of the total.
5) *Ever married [43]:* The participants' marital status is represented by this attribute, with 79.84% of them being married.
6) *Work type [44]:* The participant's employment status is represented by this characteristic, which contains 4 categories (private 65.02%, self-employed 19.21%, gov't job 15.67%, and never worked 0.1%).
7) *Residence type [45]:* It which has two groups (urban 51.14% and rural 48.86%), describes the participant's living situation.

8) *Average blood glucose level (mg/dL) [46]:* It measures the participant's typical blood glucose level.
9) *BMI (Kg/m2) [47]:* It records the participants' body mass indices.
10) *Smoking Status [48]:* It records the participant's level of smoking, which may be divided into three categories: current smoker (22.37%), never smoked (52.64%), and previous smoker (24.99%).
11) *Stroke:* Whether the subject has previously experienceda stroke is indicated by this feature. percent of individuals report having had a stroke.

V. FUTURE CHALLENGES AND POSSIBLE SOLUTIONS

Future work should focus more on utilizing more sophisticated AI techniques [17]. Future work on this project aims to improve the framework models by utilizing a bigger dataset and ML models like AdaBoost, SVM, and Bagging. In [18] The ultimate goal of this research is to apply deep learning techniques to improve the ML framework. The final, difficult yet interesting approach is to gather picture data from brainCT scans and assess how well deep learning models forecast the occurrence of strokes[20].The difficulty is in guaranteeing model performance from inconsistent, partial, or biassed data, which can be resolved by using preprocessing methods, data cleaning, and artificial data creation. Robust models for streaming data processing are necessary for real-time monitoring and prediction. Possible fix: Make use of lightweight models and maximize the scalability and efficiency of the model design.

VI. CONCLUSION

In the review, various and adaptable ML techniques are highlighted as they are used to anticipate strokes. The significance of feature selection and data integration in improving prediction performance is emphasized. Conventional risk assessment technologies may overlook subtle patterns and interactions that ML is able to identify. Additionally, it draws attention to how ML may be used to clarify the underlying processes and biomarkers linked to the incidence of stroke.

ACKNOWLEDGMENT

The authors wish to thank Parul university for providing all the necessary facilities that were required for this study.

REFERENCES

[1] S. L. Murphy, K. D. Kochanek, J. Xu, and E. Arias, "Mortality in the united states, 2020," 2021.

[2] S. S. Virani et al., "Heart disease and stroke statistics—2020 update: a report from the American Heart Association," Circulation, vol. 141, no. 9, pp. e139–e596, 2020.

[3] D. O. Kleindorfer et al., "2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline from the American Heart Association/American Stroke Association," Stroke, vol. 52, no. 7, pp. e364–e467, 2021.

[4] W. L. Baker et al., "Neurothrombectomy devices for the treatment of acute ischemic stroke: state of the evidence," Annals of internal medicine, vol. 154, no. 4, pp. 243–252, 2011.

[5] D. Kleindorfer et al., "Which stroke symptoms prompt a 911 call? A population-based study," The American journal of emergency medicine, vol. 28, no. 5, pp. 607–612, 2010.

[6] S. De Raedt, A. De Vos, and J. De Keyser, "Autonomic dysfunction in acute ischemic stroke: an underexplored therapeutic area?," Journal of the neurological sciences, vol. 348, no. 1–2, pp. 24–34, 2015.

[7] C. Warlow, "Epidemiology of stroke," The Lancet, vol. 352, pp. S1–S4, 1998.

[8] K.-D. Seo, M. J. Kang, G. S. Kim, J. H. Lee, S. H. Suh, and K.-Y. Lee, "National trends in clinical outcomes of endovascular therapy for ischemic stroke in South Korea between 2008 and 2016," Journal of Stroke, vol. 22, no. 3, p. 412, 2020.

[9] T. D. Musuka, S. B. Wilton, M. Traboulsi, and M. D. Hill, "Diagnosis and management of acute ischemic stroke: speed is critical," Cmaj, vol. 187, no. 12, pp. 887–893, 2015.

[10] Q. Song et al., "Long sleep duration and risk of ischemic stroke and hemorrhagic stroke: the Kailuan Prospective Study," Scientific reports, vol. 6, no. 1, p. 33664, 2016.

[11] J. Yu, S. Park, H. Lee, C.-S. Pyo, and Y. S. Lee, "An elderly health monitoring system using machine learning and in-depth analysis techniques on the NIH stroke scale," Mathematics, vol. 8, no. 7, p. 1115, 2020.

[12] W. Factsheet, "The top 10 causes of death," Geneva: World Health Organization, 2020.

[13] U. N. D. of Economic and S. Affairs, World social report 2020. United Nations New York, NY, USA, 2020.

[14] G. Leonardi, S. Montani, and M. Striani, "Explainable process trace classification: An application to stroke," Journal of Biomedical Informatics, vol. 126, p. 103981, 2022.

[15] P. Kunwar and P. Choudhary, "A stacked ensemble model for automatic stroke prediction using only raw electrocardiogram," Intelligent Systems with Applications, vol. 17, p. 200165, 2023.

[16] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," Healthcare Analytics, vol. 2, p. 100032, 2022.

[17] A. A. Ali, "Stroke prediction using distributed machine learning based on apache spark," Stroke, vol. 28, no. 15, pp. 89–97, 2019.

[18] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M.

[19] M. Khan, "Stroke disease detection and prediction using robust learning approaches," Journal of healthcare engineering, vol. 2021, 2021.

[20] M. S. Islam, I. Hussain, M. M. Rahman, S. J. Park, and M. A. Hossain, "Explainable artificial intelligence model for stroke prediction using EEG signal," Sensors, vol. 22, no. 24, p. 9859, 2022.

[21] E. Dritsas and M. Trigka, "Stroke risk prediction with machine learning techniques," Sensors, vol. 22, no. 13, p. 4670, 2022.

[22] C. Kokkotis et al., "An explainable machine learning pipeline for stroke prediction on imbalanced data," Diagnostics, vol. 12, no. 10, p. 2392, 2022.

[23] R. Islam, S. Debnath, and T. I. Palash, "Predictive analysis for risk of stroke using machine learning techniques," in 2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), 2021, pp. 1–4.

[24] N. Darabi, N. Hosseinichimeh, A. Noto, R. Zand, and V. Abedi, "Ma- chine learning-enabled 30-day readmission model for stroke patients," Frontiers in neurology, vol. 12, p. 638267, 2021.

[25] Y. Choi and J. W. Choi, "Stroke Prediction Using ML based on Artificial Intelligence," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 5, pp. 1-6, 2020.

[26] H. H. KV, G. Gupta, and P. KB, "Stroke prediction using ML algorithms," International Journal of Innovative Research in Engineering Management, vol. 8, no. 4, pp. 6-9, 2021.

[27] T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, "Challenges of deep learning in medical image analysis—Improving explainability and trust," IEEE Transactions on Technology and Society, vol. 4, no. 1, pp. 68–75, 2023.

[28] J. Chen, Y. Chen, J. Li, J. Wang, Z. Lin, and A. K. Nandi, “Stroke risk prediction with hybrid deep transfer learning framework,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 1, pp. 411–422, 2021.

[29] S. Penafiel, N. Baloian, H. Sanson, and J. A. Pino, “Predicting stroke risk with an interpretable classifier,” IEEE Access, vol. 9, pp. 1154–1166, 2020.

[30] J. Yu, S. Park, S.-H. Kwon, K.-H. Cho, and H. Lee, “AI-based stroke disease prediction system using ECG and PPG bio-signals,” Ieee Access, vol. 10, pp. 43623–43638, 2022.

[31] Q. Lu et al., “Machine learning models for stroke detection by observing the eye-movement features under five-color visual stimuli in traditional Chinese medicine,” Journal of Traditional Chinese Medical Sciences, vol. 10, no. 3, pp. 321–330, 2023.

[32] K. Mridha, S. Ghimire, J. Shin, A. Aran, M. M. Uddin, and M.

[33] F. Mridha, “Automated stroke prediction using machine learning: An explainable and exploratory study with a web application for early intervention,” IEEE Access, 2023.

[34] A. Srinivas and J. P. Mosiganti, “A brain stroke detection model using soft voting based ensemble machine learning classifier,” Measurement: Sensors, vol. 29, p. 100871, 2023.

[35] P. Bathla and R. Kumar, “A hybrid system to predict brain stroke using a combined feature selection and classifier,” Intelligent Medicine, 2023.

[36] Y. Wang et al., “Assessment of stroke risk using MRI-VPD with automatic segmentation of carotid plaques and classification of plaque properties based on deep learning,” Journal of Radiation Research and Applied Sciences, vol. 16, no. 3, p. 100630, 2023.

[37] J. Mak et al., “Detection of Stroke-Induced Visual Neglect and Target Response Prediction Using Augmented Reality and Electroencephalography,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1840–1850, 2022.

[38] M. Alruily, S. A. El-Ghany, A. M. Mostafa, M. Ezz, and A. A. El- Aziz, “A-tuning ensemble machine learning technique for cerebral stroke prediction,” Applied Sciences, vol. 13, no. 8, p. 5047, 2023.

[39] Y. Freund, R. E. Schapire, and others, “Experiments with a new boosting algorithm,” in icml, 1996, vol. 96, pp. 148–156.

[40] G. Ke et al., “Lightgbm: A highly efficient gradient boosting decision tree,” Advances in neural information processing systems, vol. 30, 2017.

[41] Stroke Prediction Dataset. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. Accessed: May 25, 2022.

[42] K. M. Rexrode, T. E. Madsen, A. Y. Yu, C. Carcel, J. H. Lichtman, and E. C. Miller, “The impact of sex and gender on stroke,” Circulation research, vol. 130, no. 4, pp. 512–528, 2022.

[43] J. Dubow and M. E. Fink, “Impact of hypertension on stroke,” Current atherosclerosis reports, vol. 13, pp. 298–305, 2011.

[44] C. W. Tsao et al., “Heart disease and stroke statistics—2022 update: a report from the American Heart Association,” Circulation, vol. 145, no. 8, pp. e153–e639, 2022.

[45] K. Andersen and T. Olsen, “Stroke case-fatality and marital status,” Acta Neurologica Scandinavica, vol. 138, no. 4, pp. 377–383, 2018.

[46] A. M. Cox, C. McKevitt, A. G. Rudd, and C. D. Wolfe, ”Socioeconomic status and stroke,” Lancet Neurol., vol. 5, no. 2, pp. 181-188, Feb. 2006.

[47] G. Howard, ”Rural-urban differences in stroke risk,” Preventive Medicine, vol. 152, p. 106661, 2021.

[48] Y. Cai, C. Wang, W. Di, W. Li, J. Liu, and S. Zhou, ”Correlation between blood glucose variability and the risk of death in patients with severe acute stroke,” Revue Neurologique, vol. 176, no. 7-8, pp. 582-586, 2020.

[49] S. Elsayed and M. Othman, ”The effect of body mass index (BMI) on the mortality among patients with stroke,” Eur. J. Mol. Clin. Med, vol. 8, pp. 181-187, 2021.

[50] R. S. Shah and J. W. Cole, ”Smoking and stroke: the more you smoke the more you stroke,” Expert Review of Cardiovascular Therapy, vol. 8, no. 7, pp. 917-932, 2010.

[51] V. Novak, K. Hu, L. Desrochers, P. Novak, L. Caplan, L. Lipsitz, and M. Selim, ”Cerebral flow velocities during daily activities depend on blood pressure in patients with chronic ischemic infarctions,” Stroke, vol. 41, no. 1, pp. 61-66, 2010.

[52] TIMESOFINDIA.COM, Oct. 29, 2021. ”Brain stroke symptoms: 6 simple signs to know if someone is having a brain stroke,” The Times of India. [Online]. Available: https://timesofindia.indiatimes.com/lifestyle/health-fitness/health-news/6-simple-signs-to-know-if-someone-ishaving-a-brain-stroke/articleshow/87343111.cms. Accessed: May 25, 2022.