RESEARCH PAPER

A Transparent And Accurate Machine Learning Approach for Early Heart Stroke Prediction and Intervention

Dr. S. Venkata Achuta Rao[1*], **Arshiya**[1]

[1]Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana, India

[1]Professor - CSE & Dean-Academics (Dr. S. Venkata Achuta Rao) | Email: sreedatthaachyuth@gmail.com

[1]PG Student (Arshiya) | Email: arshiya1225@gmail.com

***Corresponding Author:** Dr. S. Venkata Achuta Rao, Professor - CSE & Dean-Academics, Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana, India | Email: sreedatthaachyuth@gmail.com

Abstract— Stroke is a major threat to health and the economy around the world. It occurs when there is an interruption of blood flow to the brain, which results in cognitive damage. As the population is becoming older, the numbers at risk for stroke are increasing. This makes it even more important to have accurate prediction tools right away. The problem of stroke is being worked on by this project, which is making automatic prediction algorithms. The point of these algorithms is to enable early intervention, which could save lives by a correct stroke prediction. With the increase in the number of people at risk, it becomes increasingly important that they work precisely and efficiently. The project contains an extensive analysis of the comparison of effectiveness of a proposed machine learning approach to six well known classifiers. To see how well the new algorithm worked at predicting strokes, metrics were looked at that measured both its ability to generalize and its accuracy in making predictions. The study uses methods which can be explained, specifically SHAP (Shapley Additive Explanations) to make the black-box nature of Machine learning models more clear. This method has been applied in the medical field for quite some time and can help you to understand how models make decisions. The results of the test show that models with more details worked better than models with less details, thus, more accurate. Using both global explainable methods and the proposed framework, increasing the consistency of complex models is the goal. This standardization can lead to better care and treatment for strokes because we will have useful information about how the programs make decisions. It employs ensemble techniques such as Categorical Boosting, Stacking Classifier to enhance the accuracy of the predictions by taking the results of more than one model in its predictions. In particular the Stacking Classifier has done amazingly well with an astounding 99% accuracy.

"Keywords— Stroke prediction, data leakage, explainable machine learning."

How to cite this article: Venkata Achuta Rao S, Arshiya. A transparent and accurate machine learning approach for early heart stroke prediction and intervention. Int J Drug Deliv Technol. 2026;16(75s): 1744-1749. DOI: 10.25258/ijddt.16.75s.184

I. INTRODUCTION

Stroke is now one of the leading causes of death and disability worldwide and the number is increasing. Long-term disability and death after a stroke is very important, early treatment. But traditional ways of figuring out the risk of having a stroke take a long time and often make mistakes. Recently, ML algorithms has shown a lot of promise in being able to correctly predict the risk of stroke based on different clinical risk factors. Using these algorithms, doctors can identify those patients who are at high risk and assist them immediately, which could reduce the number of complications associated with strokes and improve patient results. Also, there is a growing need for ML models in the healthcare to be understandable and easy to read. Using a ML model that can be interpreted can help doctors figure out what makes a patient more likely to have a stroke, which can help them decide how to treat them. According to the World Stroke Organization, 13 million people suffer from a stroke annually and 5.5 of them die from it [1]. Stroke is among the major causes of mortality and disability globally [1, 2]. It changes every aspect of a person's life including the family, friends, and job. There is a common misconception that stokes only occur in certain types of people, such as the old, or in people who already have health problems. In reality, it can affect anyone, regardless of their age, gender and health [1, 2]. When blood flow to the brain becomes suddenly deprived by the sudden and serious cutting off of blood flow, brain cells become oxygen deprived. This is called a stroke. There are two types of it: cerebral or hemorrhagic. Depending on the severity of strokes, moderate to serious strokes can damage you permanently or temporarily. Not many people experience hemorrhagic strokes, but it occurs when the blood vessel of the brain is broken. Most strokes occur when an artery becomes blocked or narrowed and blood no longer flows to the brain [3, 4]. A person has a higher probability of suffering a stroke if the patient is above age 55, has suffered a stroke or a TIA before, is suffering from arrhythmia, has high blood pressure, has atherosclerosis (causes carotid stenosis), smokes, has diabetes, is overweight, not active, takes estrogen, has blood clotting problems, uses cocaine or amphetamines, or has heart problems such as infarction, or cardiac arrest [5, 6, 7]. Strokes can come on fast, and the signs are also different and at strange times. Some of the main signs of a stroke include paralysis on one side of the body, numbness on the face, arms, legs, trouble speaking or walking, dizziness, blurred eyesight, headache, vomiting, a mouth that hangs open, and in the worst situation, a loss of awareness and entering a coma. You may feel these things all of a sudden, or over time and in very rare cases, may make you aware [8, 9, 10].

II. RELATED WORK

[7] This study, "Stroke risk factors, genetics, and prevention," talks about how stroke is a complex syndrome and how to figure out risk factors and treatment depends on how the stroke starts. There are two types of risk factors for stroke: those that can be changed, and those that can't. Age, sex and race or ethnicity are risk factors for both ischemic and

hemorrhagic stroke that can't be changed. On the other hand, high blood pressure, smoking, diet, and physical exercise lack are some of the common risk factors that can be changed. Inflammatory disorders, infections, pollution, and heart atrial disorders other than caused by atrial fibrillation have been referred to as new risk factors and causes of stroke. Stroke is often the first symptom of rare genetic diseases which are caused by a single gene. New study also shows that both common and rare genetic polymorphisms can have an effect on risk of more common causes of stroke. This is because they can influence both other risk factors, as well as specific stroke mechanisms, such as atrial fibrillation. Genetic factors, particularly those of connection with the environment, may be more easily modified than was believed before. Stroke prevention has largely been directed towards risk factors that can be modified. Making changes to your lifestyle and habits, such as stopping smoking or watching what you eat, can reduce your risk of stroke and other heart diseases, as well. Finding and treating medical problems such as high blood pressure and diabetes that increase the risks for stroke is another way to stop them. New study into the genetics and risk factors of stroke has not only found people who are likely to have a stroke, but also ways to prevent strokes in those groups.

[12] Stroke is one of the major clinical outcomes in cardiovascular research and this study, Natural language processing and ML for identifying incident stroke from electronic health records: Algorithm development and validation, addresses it. Finding out about an incident stroke, on the other hand, is usually done by manual abstracting of charts which takes a lot of time. The way electronic health records are used now for phenotyping strokes is mainly for finding cases not incident diseases which need knowing the order of events in time. The purpose of this study was to develop a phenotyping algorithm for determining the type of stroke that occurred using the diagnosis codes, procedure codes and clinical ideas abstracted from clinical notes using natural language processing and ML. For training and testing, there was an existing epidemiology group of 4914 patients with atrial fibrillation (AF) who experienced strokes that was carefully curated. Different mixes of feature sets, machine learning models were looked at side by side. We used a rule derived from the composition of ideas and codes to determine the type of stroke (ischemic stroke, transient ischemic attack or hemorrhagic stroke) for each one that we found. A cohort (n=150), stratified sample from a group in Olmsted County, Minnesota (N=74,314) was used to test the algorithm even more.

[13] "Multi-frequency symmetry difference electrical impedance tomography with ML for human stroke diagnosis" is the title of a paper which talks about how Multi-frequency symmetry difference electrical impedance tomography (MFSD-EIT) can reliably find and name one-sided changes in symmetric scenes. Here, an investigation is conducted in order to understand whether it is possible the use of the algorithm to properly discover the reason why someone has stroke with the help of ML. We made four-layer finite element method models of the head that were anatomically accurate based on pictures of stroke patients. These models are then used to make EIT data for a frequency range of 5 Hz and 100 Hz with and without bleed and clot injuries. Conductivity maps of each head at each frequency by reconstruction are made. Using a quantitative measure to look at changes in symmetry over the sagittal plane of the reconstructed image, as well as over the frequency range, makes it possible to find and name lesions. The method is applied to both fake and 34 real data. The metric value is put through a classification program to tell the difference between normal, hemorrhage, and clot values. MFSD-EIT with SVM classification can distinguish between a bleed and a clot in the case of human data, with an average 85% accuracy. It can also distinguish between a normal blood flow and a stroke in data of human subjects with 77% accuracy. Classification algorithm utilizing metrics obtained from MFSD-EIT images is a new and interesting approach to discovering and naming changes in scenes which are otherwise still. In combination with machine learning, the MFSD-EIT method is promising for lesion detection and identification in challenging scenarios such as stroke. The results prove that it is possible to apply the method to "real" cases.

"AI for decision support in acute stroke-Current roles and potential" is the title of one such paper, which discusses what this means. There are an increasing number of treatment options for stroke patients, and new relationships between disease characteristics and treatment are always being discovered. This presents a more difficult diagnosis and treatment of stroke patients. That's why doctors have to get on top of learning new things, such as how to do clinical evaluations or how to read images, the latest research, and apply these new things in their everyday work. Using AI to help doctors make decisions could cut down on differences between raters in everyday clinical practice and make it easier to get important data that could help find stroke patients, predict how they will respond to treatment, and improve patient outcomes. These types of support systems would be effective in centers that only see a few stroke patients or support centers throughout the area. They could help patients and their families to have more informed conversations. Also, using AI to process and understand the images of the stroke could provide any doctor an imaging report that's the same quality as one by an expert. Any AI-based decision support system, though, should allow for an expert clinician to work with it so that mistakes could be found (such as automated image processing). This review talks about how imaging is becoming more and more important in managing strokes. It then focuses on the advantages and disadvantages of AI to assist in making treatment decisions for individuals who have recently experienced a stroke.

[15] A paper entitled at "A systematic review of ML models to predict outcomes of stroke using structured data" discusses ML in a paper. ML has garnered a lot of attention because people believe that it can use big and regularly collected datasets to provide accurate and personalized prognoses. The purpose of this systematic study is to find and evaluate the reporting and development of ML models to predict what will happen after a stroke. We searched PubMed and Web of Science database from 1990 to March 2019 with the previously published search terms of stroke, ML and prediction models. Here we only looked at organized clinical data and we didn't look at images or text. This review was filed All studies looked at discrimination thirteen times using area under the ROC curve, and three times looked at calibration. External validation was completed in two instances. None of them provided sufficient information of the end model to make a copy of the end model. ML is being used more and

more to predict how a stroke will turn out. But few of them adhered to the basic rules for reporting clinical prediction tools and none of them made their models available to the public in a form that is usable and can be assessed. Before ML can be seriously considered for use, there needs to be a lot of work done on how studies are done and how they are reported.

III. Materials And Methods

A novel approach is adopted to test the stroke prediction models and how it would compare to six classifiers in the proposed system. By using SHAP, the group learns more about the way decisions are made. The accuracy is better when the information is preprocessed and balanced using SMote. In a unique way, the project produces and experiments with different methods of machine learning at predicting strokes early on. The other algorithm implemented in this study was ensemble methods which uses the predictions of several separate models used to make the general prediction more accurate. Some of the methods are Categorical Boosting, Stacking Classifier. In particular the Stacking Classifier performed amazingly well getting an impressive 99% accuracy. As a real-world example of this new technology, a frontend created through the Flask framework to make it easy to test the users. Adding user authentication also ensures that authorized users only can access the system, creating a safe and easy-to-use interface for using the automated stroke prediction methods.

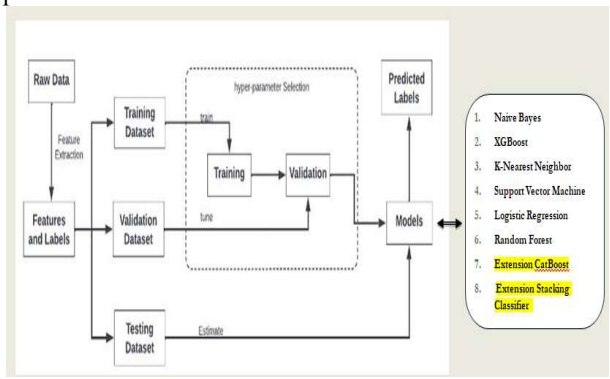


Fig. 1. System Architecture

ML is being used more and more in medical diagnostics, like putting skin cancer into groups, because it is good at processing huge amounts of medical data, like shots of skin lesions. Mostly using machine learning models to identify strokes is useful to improve the accuracy of diagnoses and the speed at which patients are grouped. A system that automatically predicts strokes is often made using a number of different machine learning models. This system is then tested using accuracy, memory, and F1 Score to determine the best model for the job. Making a "Yes" or "No" guess is the automated way to group the predictions of stroke according to this study. Figure 1 shows the five of the steps that are used to make the model. The first step towards this is to obtain a collection of electronic health records. The first three steps prepare the dataset including rescaling and normalizing the data. The fourth step is extraction of features. The procedure of the fifth step is to construct a classifier algorithm given the extracted feature vectors. The sixth step is to use the SHAP and LIME methods to demonstrate how the model is making decisions. This better way is trying to better the way doctors make better decisions on treatment and making a more accurate predictions on strokes.

A) Dataset Collection

The "Stroke Dataset" has health-related information such as gender, age, high blood pressure, heart disease, marital status, work, type of home, average glucose level, BMI, smoking status, and number of strokes that occurred. Each post has its own "id" that gives it a unique appearance. With binary indicators for hypertension, heart disease, marital and residence status, the collection gives us useful information about the things that are linked to stroke. This information is loaded and looked into. This includes looking at the structure of the dataset, making sure there are no missing values and learning more about how the features are distributed and what their traits are.

	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	0.0	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
...	...	...	...	...	...	...	...	...	...	...	...	...
5105	18234	Female	80.0	1	0	Yes	Private	Urban	83.75	0.0	never smoked	0
5106	44873	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0
5107	19723	Female	35.0	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
5108	37544	Male	51.0	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5109	44679	Female	44.0	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

Fig. 2. Stroke Dataset

B) Pre-Processing

Data processing is the process of transforming unstructured data into knowledge that can be utilized by businesses. In general, data scientists work with data, which means that they collect it, organize it, clean it, check it, analyze it, and transform it into something that can be read, such as graphs or documents. There are three ways in which data can be handled: by hand, mechanically, or electronically. The aim is to make knowledge more useful and decision-making easier. This assists companies to run better and make smart strategy decisions faster. This is made possible in large part thanks to automated data handling tools, such as computer programming. It can help convert big data and other types of data into useful information for decision-making and quality control purposes.

Feature selection: Feature selection is the process by which to select the most consistent, useful and non-redundant traits to use in developing a model. As the number and types of records become larger, it is important to decrease their sizes in a planned manner. One of the major goals of feature selection is to get a prediction model to operate better and with less computing power.

One of the most important parts of feature engineering is feature selection which is the process of selecting the most important features to feed it into algorithms for ML. Feature selection methods involve getting rid of unwanted or unwanted features and at the same time reserving only those features that are most important to the ML model. This reduces the number of input variables. If you specify which features are most important as opposed to letting the ML model figure it out, here are the major benefits.

D) Algorithms

XGBoost is a complicated ML technique that falls into the category of the collection of gradient boosting frameworks. It's very good both at regression and classification task. This is done by using an ensemble method to build decision trees in such a way as to correct mistakes as it goes. Its "extreme"

skills come on the way it works, how well it can scale and how well it can regularize data. XGBoost is used in the project because it is very good in making prediction and capable of handling complicated links between the clinical risk factors for strokes very well. Its ensemble method ensures to ensure that the prediction is correct, which is in accordance with the project's goal to develop an accurate tool for early detecting high-risk stroke patients and getting them help.

$y^i=\sum k=1Kfk(xi)$ (1)

The last predicted value for the ith data point is y^iy^i, K which is the number of trees in the ensemble and fk(xi)fk(xi) is the estimate for the ith data point in the ensemble from the Kth tree.

Naive Bayes is a probabilistic ML method that is based on Bayes theorem and assume that features is not related each other. Naive Bayes is picked because it is easy to use, and good at dealing with large amounts of data that might be related. Naive Bayes, essentially, is a very rapid method of determining the risk for a stroke based on a wide range of clinical factors. It fits with the goal of the project of making accurate risk predictions. It's a good choice for healthcare apps because it's easy to set up and understand.

When we use expanded notation we write:

p(A,B|A) = p(A) * p(B|A) (2)

"The chance of A and B given A is equal to the chance of A times the chance of B given A is known or has happened." This type of probability is called conditional probability or better put, a shared conditional probability in the sense that it is based on a condition or an event that took place in the past.

KNN is a kind of flexible ML method which uses the rules of proximity to do both classification and regression. In stroke prediction, where the links between clinical risk factors are complicated, because KNN is simple we can find patterns by looking how close two data points are to each other. The goal of this project is to accurately predict the risk of stroke which is non-linear relationship and changing depending on various data distributions than it can be handled by this model. KNN works well in situations with complex or non-linear data structures because of its ability to assist in cases where decision limits are not clear.

The Euclidean distance between the data points will be found. We already learnt about the Euclidean distance in mathematics. It is the length of a distance between two points. Here's how to figure it out:

dist((x, y), (a, b)) = $\sqrt{(x — a)^2 + (y — b)^2}$ (3)

SVM is a powerful classification and regression algorithm, which particularly works well in a space of many dimensions. SVM was picked for the project because it is good at dealing with complicated relationships between clinical risk factors for stroke. By finding the best hyperplanes, SVM helps by improving the accuracy of determining the risk of stroke. It is particularly effective when the data has a complex pattern. It is capable of handling multidimensional and nonlinear data, which is in line with the project's goal of making a good prediction model.

What is the equation of the straight line hyperplane?

wTx+b=0$wTx+b=0$ (4)

Where:

- To understand the direction it is perpendicular to the hyperplane, ww is
- bb is the offset or bias term which shows how far the hyperplane is from the origin along the normal vector ww.

Logistic Regression is a statistical technique for binary classification based on the logistic function which is used to model the probability that a given instance is a member of a particular class. Logistic Regression is used for binary classification for stroke prediction as it is easy to use and works well. The aim of the project is to develop a valid and comprehensible model that classifies the risk of having a stroke according to various clinical characteristics. This approach of the method is simple and fits to this goal.

The linear line of regression,

^y=a+bx y ^ = a + b x (5)

can be understood in this way:

^In the example of y = mx + c, assuming that an is the intercept and y is the y-axis then y is the number that is expected to be found. If x changes by a unit b is going to tell you how much y is going to change by.

Random Forest is a form of ensemble learning technique which uses the predictions of multiple decision trees and combines them in a single form for the task, such as classification or regression. Random Forest enhances the accuracy with the help of the combination of the estimates of several trees. It was chosen because it can manage complex relationships in the clinical risk factors. The goal of the project is to have a very accurate and usable model to predict the risk of having a stroke, and this tool works well at handling large amounts of information and avoiding overfitting.

Stacking Classifier is an ensemble method where a meta-classifier is used to combine various classifiers using a base model for better accuracy of predictions. Used to exploit the various strengths of algorithms. By combining the results of these models, the Stacking Classifier attempts to create a strong and accurate stroke risk prediction model that fixes the flaws in each method for a complete picture of a patient's risk.

CatBoost is a strong gradient boosting algorithm made for decision trees that is known for being good at working with category features without a lot of extra work. Due to this good performance at category features, CatBoost improves the speed of modeling and reduces the work to be done for the model to be used. Its speed and dependability assist in making accurate predictions on the risk of getting a stroke by recording complicated relationships between clinical risks. CatBoost enhances prediction of stroke accuracy of strokes and generalizability

$F(x)=F0(x)+\sum m=1M\sum i=1Nfm(xi)$ (6)

The equation says that if you want to get the overall forecast F(x) you add-up the first guess F 0(x) and the predictions made by each tree f_m(xi) for each training sample. All trees (m) and all training samples (i) are added up in this manner.

IV. Experimental Results

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} (3)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (4)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (5)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

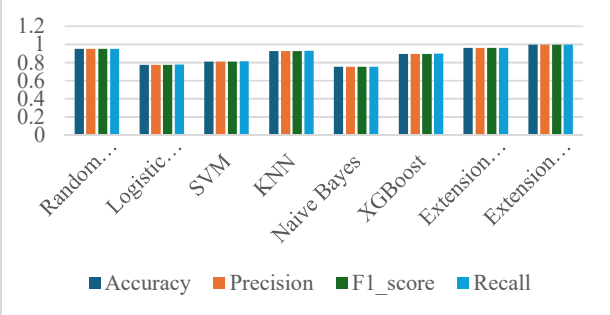
$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 (6)$$

Look at Table (1) to see how good each method performs in terms of accuracy, precision, recall and F1-score. The Stacking classifier always gives a better result than all the other methods. The tables also indicate how the measurements for the other algorithms compare to each other.

Table. 1. Performance Evaluation

ML Model	Accuracy	Precision	F1_score	Recall
Random Forest	0.952	0.952	0.952	0.952
Logistic Regression	0.777	0.777	0.777	0.778
SVM	0.809	0.808	0.808	0.812
KNN	0.927	0.926	0.927	0.932
Naive Bayes	0.752	0.751	0.751	0.752
XGBoost	0.895	0.895	0.895	0.897
Extension CatBoost	0.960	0.960	0.960	0.960
Extension Stacking Classifier	0.999	0.999	0.999	0.999

Fig. 3. Comparison Graph



The above formula shows that accuracy is blue, precision is orange, memory is green, and F1 - Score is in the sky blue graph (1). When compared to the rest of the models, the Stacking classifier performs better on all of them and achieves the maximum values. The above graphs demonstrate these results very well.

V. Conclusion

Using cutting-edge ML methods to find high risk people at an early stage, the project has made big steps to increasing the accuracy of stroke predictions. The project was successful in fixing problems with class imbalance in the dataset, which ensures that there is a more accurate and balanced picture of stroke and non-stroke cases to make the model work better. The extension method Stacking Classifier had the best job of predicting strokes, it got it right 99% of the time. It worked well with easy-to-use front end and processed feature values correctly, showing very good speed and usefulness to work with for real-life healthcare implementation. The project supports such consistency and effectiveness of stroke care by standardisation of complicated models of the global and local explainable methodologies. This provides for better treatment plans for more patient situations. Building AI systems that can be trusted as well as whose explanation is clear and concise is a big step ahead in the area of XAI in medicine. This ensures that predictive models for stroke risk assessment are accountable and accurate.

More study could look into more types of ML algorithms and methods to make stroke prediction models more accurate and useful in more situations. There is potential for development in the construction of the web application for early stroke intervention. The effect regarding stroke care and treatment could point in the future by adding more advanced features and functions. If you combine global and local methods for explaining, but especially SHAP, it opens up some new areas of study. If you dig into these methods, you may be able to learn more about the ways that the ML models used in stroke detection make decisions. Making a complete end-to-end smart stroke prediction system that can work with mobile apps for both Android and iOS is one way that could go in the future. This addition might make things easier to get to and use. A good idea for future study is to get more in-depth look at demographic factors such as age and gender. Finding out how these factors affect the risk of stroke in a more complex way could lead to the creation of more accurate and individualised prediction models.

References

[1] Learn About Stroke. Accessed: May 25, 2022. [Online]. Available: https://www.world-stroke.org/world-stroke-day-campaign/why-strokematters/learn-about-stroke

[2] T. Elloker and A. J. Rhoda, ‘‘The relationship between social support and participation in stroke: A systematic review,’’ Afr. J. Disability, vol. 7, pp. 1–9, Oct. 2018.

[3] M. Katan and A. Luft, ‘‘Global burden of stroke,’’ Seminar Neurol., vol. 38, no. 2, pp. 208–211, Apr. 2018.

[4] A. Bustamante, A. Penalba, C. Orset, L. Azurmendi, V. Llombart, A. Simats, E. Pecharroman, O. Ventura, M. Ribó, D. Vivien, J. C. Sanchez, and J. Montaner, ‘‘Blood biomarkers to differentiate ischemic and hemorrhagic strokes,’’ Neurology, vol. 96, no. 15, pp. e1928–e1939, Apr. 2021.

[5] X. Xia, W. Yue, B. Chao, M. Li, L. Cao, L. Wang, Y. Shen, and X. Li, ‘‘Prevalence and risk factors of stroke in the elderly in northern China: Data from the national stroke screening survey,’’ J. Neurol., vol. 266, no. 6, pp. 1449–1458, Jun. 2019.

[6] A. Alloubani, A. Saleh, and I. Abdelhafiz, ‘‘Hypertension and diabetes mellitus as a predictive risk

factors for stroke,'' Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 12, no. 4, pp. 577–584, Jul. 2018.

[7] A. K. Boehme, C. Esenwa, and M. S. V. Elkind, ''Stroke risk factors, genetics, and prevention,'' Circ. Res., vol. 120, no. 3, pp. 472–495, Feb. 2018.

[8] I. Mosley, M. Nicol, G. Donnan, I. Patrick, and H. Dewey, ''Stroke symptoms and the decision to call for an ambulance,'' Stroke, vol. 38, no. 2, pp. 361–366, Feb. 2007.

[9] J. Lecouturier, M. J. Murtagh, R. G. Thomson, G. A. Ford, M. White, M. Eccles, and H. Rodgers, ''Response to symptoms of stroke in the UK: A systematic review,'' BMC Health Services Res., vol. 10, no. 1, pp. 1–9, Dec. 2010.

[10] L. Gibson and W. Whiteley, ''The differential diagnosis of suspected stroke: A systematic review,'' J. Roy. College Physicians Edinburgh, vol. 43, no. 2, pp. 114–118, Jun. 2013.

[11] N. M. Murray, M. Unberath, G. D. Hager, and F. K. Hui, ''Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: A systematic review,'' J. NeuroInterventional Surgery, vol. 12, no. 2, pp. 156–164, Feb. 2020.

[12] Y. Zhao, S. Fu, S. J. Bielinski, P. A. Decker, A. M. Chamberlain, V. L. Roger, H. Liu, and N. B. Larson, ''Natural language processing and machine learning for identifying incident stroke from electronic health records: Algorithm development and validation,'' J. Med. Internet Res., vol. 23, no. 3, Mar. 2021, Art. no. e22951.

[13] B. McDermott, A. Elahi, A. Santorelli, M. O'Halloran, J. Avery, and E. Porter, ''Multi-frequency symmetry difference electrical impedance tomography with machine learning for human stroke diagnosis,'' Physiological Meas., vol. 41, no. 7, Aug. 2020, Art. no. 075010.

[14] A. Bivard, L. Churilov, and M. Parsons, ''Artificial intelligence for decision support in acute stroke—Current roles and potential,'' Nature Rev. Neurol., vol. 16, no. 10, pp. 575–585, Oct. 2020.

[15] W. Wang, M. Kiik, N. Peek, V. Curcin, I. J. Marshall, A. G. Rudd, Y. Wang, A. Douiri, C. D. Wolfe, and B. Bray, ''A systematic review of machine learning models for predicting outcomes of stroke with structured data,'' PLoS ONE, vol. 15, no. 6, Jun. 2020, Art. no. e0234722.