Cerebrovascular Accident Prediction Using Data Analysis Based on Machine Learning

Sri Partha Sarathi R., Kukatlapalli Pradeep Kumar and Cherujuri Ravindranath Chowdary
Computer Science and Engineering, Christ University, Bangalore, Karnataka, India

Keywords: Data Analysis, Machine Learning, Health Care Decision, Stroke Prediction, Clinical Decision Support System.

Abstract: Cerebrovascular accident is a condition that causes damage to blood vessels in the brain. It can also occur when blood flow and other nutrients to the brain stop. According to the World Health Organization (WHO), cerebrovascular accident is the leading cause of death worldwide. Most of the tasks were successful in predicting heart disease, but few of the tasks showed a stroke risk. With this concept of in mind, several machine learning models have been developed to predict the probability of having a stroke in the brain. This research selects various physiological factors, uses machine learning algorithms such as logistic regression, decision tree distribution, and random forest classification, and trains five different models for accuracy. The best algorithm for this task is Logistic Regression which is about 80% accurate.

1 INTRODUCTION

Cerebrovascular accidents (CVAs), commonly known as strokes, are a leading cause of mortality and disability worldwide. Early prediction and timely intervention are crucial for minimizing the impact of strokes on patients' lives. In recent years, the integration of analytics and machine learning techniques has emerged as a promising approach to enhance stroke prediction accuracy and improve patient outcomes. This introduction provides an overview of how analytics and machine learning can be leveraged for CVA prediction.

It highlights the significance of early detection, the limitation of traditional risk explores the application of analytics and machine learning algorithms in predicting CVAs, and discusses the advantages and challenges associated with this approach.

With the advancement of technology in medicine, they can use machine learning to predict the probability of a stroke. Algorithms in machine learning are designed in the correct order and ensure correct analysis. Previous studies of beats mainly involved studies of predicting heart palsy. There is very little research on stroke.

The introduction is followed by a literature review; it discusses articles related to CVA prediction. The literature review is followed by methodology; where it discusses the dataset, data cleaning procedure, processed data, and machine learning algorithm used in the model. The methodology is followed by results and discussion; where it is showcasing data visualization and machine learning models accuracy, precision, sensitivity, and other results. At the last conclusion of the CVA, prediction is provided.

2 LITERATURE REVIEW

This section explains the literature aspect of cerebrovascular stroke described in connection with the AI decision support system.

2.1 AI Enables Support System for Rheumatoid Arthritis

It is an AI-driven CDSS to predict RA (Rheumatoid Arthritis) exacerbations. Here, the AI must determine whether the patient has RA. Rheumatologists predict the onset of rheumatoid arthritis using several factors, including time of release, anticitrullinated protein antibody status, and multibiomarker disease activity. Therefore, many problems complicate medical decision making for patients and rheumatologists.

R., S., Kumar, K. and Chowdary, C.
Cerebrovascular Accident Prediction Using Data Analysis Based on Machine Learning.
DOI: 10.5220/0012517600003739
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 461-466
ISBN: 978-989-758-661-3
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

They developed RA flare estimators in ML. Machine learning tools are built from modern clinical data.

Risk estimation of the model using 10 clinical variables DAS28-ESR; the duration of the disease; the mode of administration of the biological DMARD; against the CCP; gender; HAQ (Health Assessment Questionnaire); To count. The prevalence of seizures in patients at high risk of seizures is 23%. To be binary, <23% must be marked "no glow" and >=23 must be marked "glare". The sensitivity of the first RA disease predictor was 72%, specificity 76%, positive predictive value 37%, negative predictive value 93%, and AUROC 80%.

The overall accuracy of the hazard assessment tool is 75%. Small sample sizes, lack of comparative studies, exclusion of patients, bias, reliability issues, and uncertainty are limitations of previous studies. Disagreement between doctors and forecasting tools can also lead to more emotion for patients. This will be beneficial for patient care.

2.2 Explainable AI for Decision Support System

This article discusses the use of AI in CDSS (Clinical Decision Support Systems). Explain the advantages and disadvantages of using AI in CDSS. Their analysis focuses on three levels: emotional intelligence, human factors, and the role of role in decision-making. They present, with further interpretation, two cases of social-technical relations. In the case of the definition described above, they use the heart attack predictor. The tool is a black box (ANN-artificial network node) Research studies show better understanding and shorter research times. In terms of decision-making, it should be ensured that the system performance is not limited to certain situations, but can be generalized to many situations, in order to achieve the best performance and achieve the best performance. In general, technical excellence, robust validation, and generalizability are required when translation is neglected in medical systems. In the next layer of human factors, users must believe that there is also a process in the audit verification process of the system.

The recommended way to solve this is to involve the client in the design and analysis. The design phase of research should allow for user feedback and criticism. which should be done. This increases their confidence in the system. Level 3 Responsibilities established in decision making, critical conditions for emergency dispatch, miscalculation can be fatal. Therefore, the lack of explanation significantly changes the performance of the system, because the decision-making process must support the algorithm-driven system.

In summary, CDSS depends on many factors. These factors include efficiency, the effectiveness of descriptive algorithms, specific content features, assigned roles, and user groups for decision-making. According to the authors, a clear answer to the role of translation cannot be determined at the theoretical level. Context specific consideration of interpretation is required in CDSS and supports collaboration to address the ethical and practical implications of AI in healthcare (Amann, 2022).

2.3 AI in Depression Treatment on Doctor-Patient Communication

CDSS in the context of depression treatment. CDSS is designed to support the hospital by taking into account many patient changes. It emphasizes the trial-and-error process, which is often used in treatment selection, and aims to reduce the number of failed treatments. The tool is designed for use in patient-centered discussions and facilitates collaborative decision-making between doctors and patients. Their aim is to connect patients with their healthcare providers as they explore decision-making, patient preferences, and best practices for shared decision-making. use it 75% of the participants reported using the hospital app for 5 minutes or less during the consultation. 40% of the participants stated that the application saves them time, and 30% stated that it is possible to do it in a real clinical setting, considering that the application does not save them time or waste their time. In most cases, participants can check the application at the designated time, indicating that the application is complete. These results provide preliminary evidence of participant satisfaction, perceived efficacy, and potential clinical implications. It should be noted that these are preliminary results and full research results will be published separately (Benrimoh, 2021).

Aifred is an AI-powered clinical decision support system for the treatment of major depression. So, find the presence of Aifred. In particular, it affects the doctor-patient interaction. 20 psychiatrists and family physicians attended, and residents completed 2.5 hours of training in a standard clinical trial. All psychiatric disorders have options to use CDSS and the patient sample can be defined as mild, moderate, and severe depression. Data were collected through patient questionnaires, case studies, and interviews. Therefore, the results for all 20 participants. Where it is acceptable and possible to use the device during treatment. Physicians are willing to rely on artificial intelligence predictions to assist in their treatment choices, and their tools are reported to help increase patient understanding and confidence in treatment. The simulated environment allowed evaluation of the device's impact on doctor-patient dialogue. Participants' transition to clinical decision-making was higher over time. The assumption is that most physicians want a tool they can use in five minutes or less, supported by the finding that 75% of respondents initially thought they could use the tool gained in that time. The integration of new technologies into treatment is an important step in the appropriate development and use of new medical devices and may be important for research and dissemination of this doctor-patient relationship tool (Benrimoh, 2021).

2.4 Role of AI in Radiology Education

Technical training is based on the interpretation of patients during the operation. They randomly divided the radiologists into two groups. One group received a simulated brain MRI with CDS and the other group received a brain MRI without CDS. Compare the time it takes to explain between the two groups. Students received a response from CDS. They use this insight to build AI-driven medical decision support. An AI-driven medical decision support system facilitates translation learning for staff. CDS helps trainees improve their skills and improve their diagnostic abilities. CDS should be considered as an additional tool to support students. Thus, combining AI with CDS opens up the possibility of improving the education of electronic students. They analyzed real-time clinical data with and without CDS and TF interpretation data with CDS for 75 MRI examinations. Students showed less confidence in TF patients compared to patients using CDS. According to the observation, the lead time with and without CDS is the same. A place with no time difference.

Studies, Optimizing CDS Algorithms, and Simulation Cases with AI-Based Clinical Decision Support (CDS) Guidelines can increase the educational value of Radiology Education. This approach focuses on areas of weakness, providing greater insight and a more focused learning experience for students. AI-based consulting has the potential to revolutionize radiology education and improve the skills and diagnostics of future radiologists (Shah,2023).

2.5 AI for Short Term System Operations

The evolution of European energy systems and their need for real decision support. The limitations of traditional control centers and the potential of artificial intelligence (AI) and deep learning (RL) to increase operational efficiency. The authors highlight the ability of AI, especially deep learning AI, to process large amounts of data, learn from historical data, and make effective decisions with time horizons and predictive uncertainty. They demonstrate the success of RL in many fields and its practical applications in electrical systems. This article presents an energy education community about the potential and challenges of AI-based decision support for energy efficiency and the importance of making decisions on a regular basis.

It includes an example supported by real-world case studies and presents the latest research in decision-making in disease control. It begins by explaining the decision-making process of the operator's power and presents the mathematical methods, including intelligence and control theory, to individually answer the phone's decision sequence. The goal is to create a system that simulates the decision-making process of humans using AI.

This document highlights the importance of incorporating AI into energy efficiency and decision support tools. It discusses operator decision-making in the electronics industry, presents real-world case studies, outlines research challenges, and presents collaborative approaches to effective AI innovation. (Viebahn, Jan, 2022).

3 METHODOLOGY

This section explains the dataset, data cleaning process, and machine algorithm that fits this data. It has used a stroke prediction dataset that has an attribute of id, age, gender, hypertension, heart disease, smoking status, residence, average glucose level, body mass index, work type, and marital status if one had a stroke or not. Age, body mass index, the average glucose level is scalar. Gender, hypertension, heart diseases, marital status, and smoking status are nominal. If the patient had hypertension is 1 else 0. If the patient had heart disease 1. else it is 0. Smoking status has never smoked, formerly smoked, smoked. The residences are categorized into rural and urban. Gender is categorized into male, Female, and others. Work type is categorized into private, government job, and self-employed and it had 49348 records.

After dropping the rows that have null values in the column. then the dataset has 29834 records with no null values. Of those records It had 548 records have had a stroke. So, 548 had no strokes are selected at random. The processed dataset has 1097 records of a stroke and not a stroke.

Using analytics, find the correlation of gender, age, age, gender, hypertension, heart disease, smoking status, residence, average glucose level, body mass index, work type, and marital status. After calculating the correlation of the attribute. We got the value; residences type and id do have a poor correlation. So, we have dropped the type and id of the residence. Using jupyter notebook, importing train and test split from the sci-kit learn-model selection. The provided dataset is segregated into two subsets, namely the training dataset and the test dataset, with a ratio of 8:2. The training dataset consists of 877 records, whereas the test dataset comprises 220 records.

The training set is fit into a machine learning algorithm. Logistic regression, random forest classifier, and decision tree machine learning algorithm are used. Logistic regression is a supervised learning machine-learning technique. The output of the categorical dependent variable is predicted. As a result, the outcome of the categorical dependent variable through the random forest algorithm, which is a supervised learning method. This algorithm is utilized to tackle both classification and regression problems, wherein a classifier is built using numerous decision trees that use varying datasets subsets to improve the prediction accuracy. Decision trees, with their decision and leaf nodes, are an effective tool for solving supervised learning problems. The output of these trees is categorical or discrete in nature, such as yes or no, true or false, and 0 or 1. To evaluate the model's performance, accuracy, precision, sensitivity, and F1 score are computed.

4 RESULTS AND DISCUSSION

This section provides a comprehensive data analysis considered from a particular dataset explaining stroke prediction. Visualization such as box whisker plots and histograms were observed and described in this regard. Machine learning model results are described in tabular forms.

4.1 Data Visualization

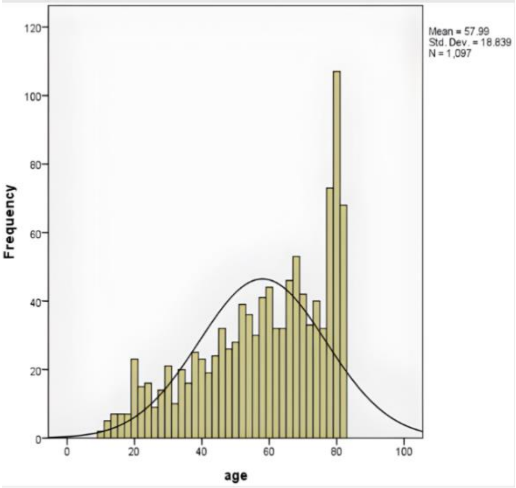

Figure 1: Histogram for age and frequency.

The above histogram explains it has a large population the 50 – 70 years. This tells the age between 50 – 70 record accuracy is higher than compared to others.

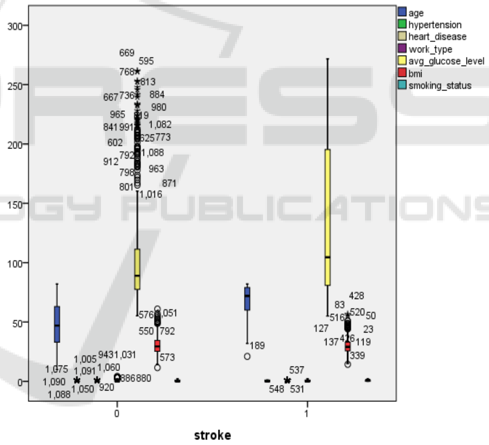

Figure 2: Box whisker plot for stroke prediction.

In the above box whisker, the value is clustered into who had a stroke ad who had not had a stroke. box whisker plotted for age, hypertension, heart disease, work type, glucose level, BMI (body mass index), and smoking status.

4.2 Correlation

Table 1: Correlation of age, hypertension, heart disease, glucose level, stroke.

		stroke
age	Pearson Correlation	.557
	Sig. (2-tailed)	.000
	N	1097
hypertension	Pearson Correlation	.205
	Sig. (2-tailed)	.000
	N	1097
Heart Disease	Pearson Correlation	.248
	Sig. (2-tailed)	.000
	N	1097
Glucose level	Pearson Correlation	.253
	Sig. (2-tailed)	.000
	N	1097
stroke	Pearson Correlation	1
	Sig. (2-tailed)	.000
	N	1097

This above Table1 is a bivariate correlation between age, hypertension, heart disease, glucose level, and stroke. age, hypertension, heart disease, and glucose level have a relation toward stroke.

4.3 Machine Learning Model

Table 2: accuracy, precision, sensitivity, specificity, and F1 score of ML models.

	Logistic regression	Random forest	Decision tree
Accuracy	0.8	0.76	0.69
Precision	0.76	0.73	0.69
Sensitivity	0.87	0.82	0.69
Specificity	0.72	0.71	0.7
F1 score	0.81	0.77	0.69

In Table 2, we can see the accuracy, precision, sensitivity, specificity, and F1-score of three machine learning models. Accuracy is measurement of the true value. Precision is measurement of repeated observation errors. Sensitivity is evaluation of the ability to accurately identify patients with the stroke. Specificity is evaluation of the ability to not accurately identify patients with stroke. The F1 score equals accuracy and goes back to the quality class, while accuracy looks at the correct distinction between good and bad.

The accuracy rate of Logistic regression in 0.8, random forest is 0.76 and Decision tree is 0.69. So, the accuracy of the logistics regression is higher than others. The logistics regression model classify dataset in better way compared to other machine learning models. In logistic regression the accuracy rate is higher than other machine learning models. Among all three-machine learning algorithm, logistic regression has highest rate of accuracy comparing to the other machine learning algorithm.

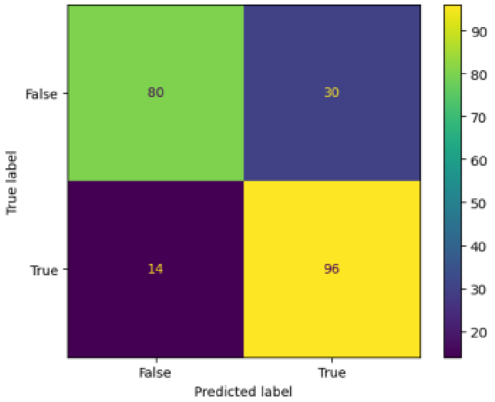

Figure 3: Confusion matrix of logistics regression.

In the above Figure 3 is a confusion matrix it explains the performance of a classification algorithm. In confusion matrix one is predicted label and another is true label. As per the confusion matrix of logistics regression 80 are predicted false it is false, 96 are predicted true it is true, 14 are predicted false but it is true and 30 are predicted true but it is false.

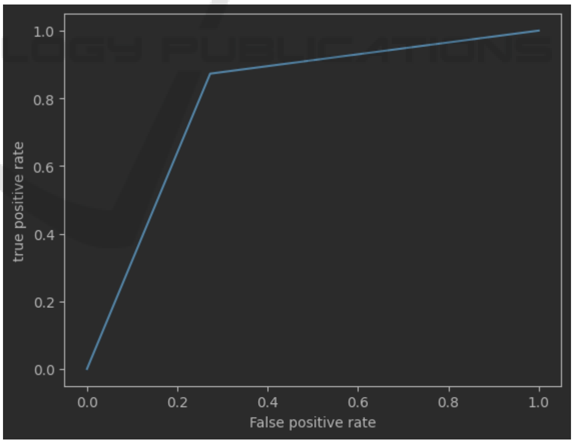

Figure 4: ROC of logistics regression.

The ROC (Receiver operating characteristic) graph above describes the performance of the classification model of each classification. The curve draws two parameters: positive value and negative value.

5 CONCLUSIONS

In conclusion, the prediction of cerebrovascular accidents using data analytics and machine learning. Using the dataset, the machine learning model is trained. In health care, the cerebrovascular accident prediction machine learning model helps in diagnosing stroke. This model reduces the time and cost consumed processed. These ml models support the clinical. While decision-making for standardized patients. It will also be useful for trainees and education perhaps. This clinical support system makes the diagnosis in a short time. These decision support systems help the future. Where it is a decision-making support system. To get greater results should have more records and a better machine learning algorithm. The accuracy of a stroke prediction model can vary depending on the quality and quantity of available data. Access to comprehensive and diverse datasets, including demographic information, medical history, lifestyle factors, and genetic markers, can significantly enhance the performance of the predictive models.

REFERENCES

Shah, Chintan, et al. "Artificial Intelligence-Powered Clinical Decision Support and Simulation Platform for Radiology Trainee Education." Journal of Digital Imaging 36.1 (2023): 11-16.

Benrimoh, David, et al. "Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician–patient interaction." BJPsych open 7.1 (2021): e22.

Viebahn, Jan, et al. "Potential and challenges of AI powered decision support for short-term system operations." CIGRE Session 2022 (2022).

Benrimoh, David, et al. "Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician–patient interaction." BJPsych open 7.1 (2021): e22.

Amann, Julia, et al. "To explain or not to explain? Artificial intelligence explainability in clinical decision support systems." PLOS Digital Health 1.2 (2022): e0000016.

Labinsky, Hannah, et al. "An AI-Powered Clinical Decision Support System to Predict Flares in Rheumatoid Arthritis: A Pilot Study." Diagnostics 13.1 (2023): 148.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS