IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

Brain Stroke Prediction

Prof. Vishakha Dilpak[1] and Siddharth Sonawane[2]
Professor, AIML, AISSMS Polytechnic, Pune, India[1]
Student, AIML, AISSMS Polytechnic, Pune, India[2]

Abstract: *The passage describes a study aimed at predicting the possibility of a stroke using various machine learning and deep learning techniques. It starts by highlighting the significance of strokes as a medical emergency, emphasizing their potential to cause severe damage and even death. The World Health Organization's declaration that stroke is a leading cause of mortality and disability globally underscores the urgency of early detection.*

I. INTRODUCTION

To address this, the study utilizes a dataset from Kaggle and employs a range of classification models, including traditional machine learning algorithms like Random Forest, Decision Tree, Logistic Regression, SVM, Naive Bayes, as well as ensemble methods like XGBoost, Ada Boost, and Light Gradient Boosting Machine. Additionally, deep neural networks, specifically three-layer and four-layer artificial neural networks (ANN), are utilized for classification tasks.

The results indicate that the Random Forest classifier achieves the highest classification accuracy at 99%, among all the machine learning classifiers. Furthermore, the four-layer deep neural network (4-Layer ANN) outperforms the three-layer ANN, achieving an accuracy of 92.39% when using selected features as input.

Despite the success of both machine learning and deep learning approaches, the study concludes that machine learning techniques overall outperformed deep neural networks in this particular task of stroke prediction.

In summary, the study underscores the importance of early stroke detection and demonstrates the efficacy of employing various machine learning and deep learning techniques for this purpose. While both approaches show promise, the results suggest that machine learning models, particularly Random Forest, are more effective in predicting the likelihood of strokes based on the given dataset

Deep Neural Network (DNN):

- A Deep Neural Network is a type of artificial neural network (ANN) that consists of multiple layers of interconnected nodes, known as neurons. Each neuron in one layer is connected to the neurons in the next layer, forming a complex network capable of learning intricate patterns from data.
- In the context of stroke prediction, a DNN would be trained on a dataset containing various features related to patients' health, lifestyle, medical history, etc. The network learns to recognize patterns in these features that are indicative of a person's likelihood of experiencing a stroke.
- DNNs are particularly effective at capturing nonlinear relationships in data and can automatically extract relevant features from raw input, making them suitable for tasks where the relationships between variables are complex and not easily characterized by traditional methods.

Extreme Gradient Boosting (XGBoost)

- XGBoost is a machine learning algorithm that belongs to the ensemble learning family, specifically the gradient boosting method.
- It works by sequentially adding decision trees to an ensemble, with each new tree correcting the errors made by the previous ones. This iterative process allows XGBoost to learn complex patterns in the data and improve predictive performance.

- In the context of stroke prediction, XGBoost would be trained on a dataset similar to that used for DNNs. It would iteratively build a collection of decision trees, each focusing on different aspects of the data, to collectively make accurate predictions about stroke risk.

Machine Learning (ML):

- Machine learning is a broader field encompassing various algorithms and techniques that enable computers to learn from data without being explicitly programmed.
- In the context of stroke prediction, machine learning techniques include not only ensemble methods like XGBoost but also other algorithms like Random Forest, Logistic Regression, Support Vector Machines (SVM), Decision Trees, Naive Bayes, and K Nearest Neighbors (KNN).
- Each of these algorithms has its strengths and weaknesses, and their effectiveness depends on the nature of the data and the specific task at hand. For example, Random Forest is known for handling high-dimensional data well and being robust to overfitting, while SVMs are effective in handling complex relationships in data.

Stroke Prediction:

- Stroke prediction refers to the task of using various features or risk factors associated with individuals to predict the likelihood of them experiencing a stroke in the future.
- Features used for stroke prediction can include demographic information (age, gender), medical history (hypertension, diabetes), lifestyle factors (smoking, physical activity), and physiological measurements (blood pressure, cholesterol levels), among others.
- By analyzing these features and their relationships with stroke occurrence, machine learning and deep learning models can help identify individuals at higher risk of experiencing a stroke, allowing for early intervention and preventive measures to be taken.

II. CONCLUSION

Stroke is a potentially fatal medical condition that needs to be treated right away to prevent future consequences. The creation of a machine learning (ML) and Deep Learning model could help with stroke early diagnosis and subsequent reduction of its severe consequences. This study examines how well different machine learning (ML) as well as Boosting algorithms predict stroke based on various biological factors. With a classification accuracy of 99%, and AUC of 1, random forest classification exceeds the other investigated techniques. According to the study, the random forest method performs better than other methods when forecasting brain strokes using cross-validation measures.

ACKNOWLEDGMENT

Data pre-processing is necessary prior to model construction in order to eliminate a dataset's undesirable noise and outliers, which could cause the model to deviate from its intended training. This phase deals with all the issues that keep the model from operating more effectively. Data must be cleansed and processed for model development after the pertinent dataset has been collected. Twelve attributes make up the dataset, as was previously said. The column id is firstly ignored because its inclusion has no impact on model creation. After that, the dataset is checked for null values and filled if any are found. In this instance, the data column's "most frequent" value is used to fill in the null values in the BMI column. The string literals in the dataset are changed by label encoding into integer values that the computer can understand. It is necessary to transform the strings to integers because the computer is typically educated on numerical data.

REFERENCES

[1]. Pikula A, Howard BV, Seshadri S. Stroke and Diabetes. In: Cowie CC, Casagrande SS, Menke A, et al., editors. Diabetes in America. (3rd ed.). Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (US), 2018, ch.19.

[2] Gary H, Gibbons L. National Heart, Lung and Blood Institute. 2022 [updated 2022 March 24]. Available from: https://www.nhlbi.nih.gov/health/stroke.

[3] Jeena RS, Kumar S. Stroke prediction using SVM, International Conference on Control. Instrumentation, Communication and Computational Technologies (ICCICCT), 2016: 600–602.

[4] Hanifa SM, Raja SK. Stroke risk prediction through non-linear support vector classification models. Int. J. Adv. Res. Comput. Sci., 2010; 1(3).

[5] Chantamit-o P, Madhu G. Prediction of Stroke Using Deep Learning Model. International Conference on Neural Information Processing, 2017: 774-781.

[6] Khosla A, Cao Y, Lin CCY, Chiu HK, Hu J, Lee H. An integrated machine learning approach to stroke prediction, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010: 183–192.

[7] Hung CY, Lin CH, Lan TH, Peng GS, Lee CC. Development of an intelligent decision support system for ischemic stroke risk assessment in a population-based electronic health record database. PLOS ONE, 2019;14(3):e0213007. https://doi.org/10.1371/journal.pone.0213007.

[8] Adam SY, Yousif A, Bashir MB. Classification of ischemic stroke using machine learning algorithms. International Journal of Computer Application, 2016;149(10):26–31.

[9] Singh MS, Choudhary P. Stroke prediction using artificial intelligence. 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), 2017:158–161.

[10] Emon MU, Keya MS, Meghla TI, Rahman MA, Mamun SA, Kaiser MS. Performance Analysis of Machine Learning Approaches in Stroke Prediction, International Conference on Enumerative Combinatorics and Applications, Nov. 2021.

[11] Kansadub T, Ammaboosadee S, Kiattisin S, Jalayondeja C. Stroke risk prediction model based on de-mographic data, in Proceedings of the 2015 8th Biomedical Engineering International Conference (BMEiCON), Pattaya, - Thailand, November 2015: 1-3.

[12] Tazin T, Alam MN, Dola NN, Bari MS, Bourouis S, Khan M. Stroke Disease Detection and Prediction Using Robust Learning Approaches. Journal of Healthcare Engineering, 2021:1-12. doi: 10.1155/2021/7633381.

[13] Sharma C, Sharma S, Kumar M, Sodhi A. Early Stroke Prediction Using Machine Learning. International Conference on Decision Aid Sciences and Applications; Mar. 2022. [14] Teoh D. Towards stroke prediction using electronic health records. BMC Medical Informatics and Decision Making, 2018; Dec.(1): 1–11. doi: 10.1186/s12911-018-0702-y.

[15] Hung CY, Lin CH, Lan TH, Peng GS, Lee CC. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017: 3110–3113.

[16] Fang G, Huang Z, Wang Z. Predicting Ischemic Stroke Outcome Using Deep Learning Approaches. Front Genet. 2022 Jan 24;12:827522. doi: 10.3389/fgene.2021.827522.

[17] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 1991 May-June; 21(3): 660-674. doi: 10.1109/21.97458.

[18] Navada A, Ansari AN, Patil S, Sonkamble BA. Overview of use of decision tree algorithms in machine learning. IEEE Control and System Graduate Research Colloquium, ICSGRC, 2011: 37–42.

[19] Rahman MM, Rana MR, Alam NAA, Khan MSI. A web-based heart disease prediction system using machine learning algorithms; 2022 June; 12. 64-80.

[20] Dhillon S, Bansal C, Sidhu B. Machine Learning Based Approach Using XGboost for Heart Stroke Prediction. in International Conference on Emerging Technologies: AI, IoT, and CPS for Science & Technology Applications, September 06–07, 2021.

[21] Akash K, Shashank HN, Srikanth S, Thejas AM. Prediction of Stroke Using Machine Learning. June 2020.

[22] Aiello S, Cliff C, Roark H, Rehak L, Stetsenko P, and Bartz A. Machine Learning with Python and H2O. (5th Ed.). H2O. ai Inc. Nov. 2017.

[23] Sailasya G and Kumari G. L. A. Analyzing the performance of stroke prediction using ML classification algorithms. International Journal of Advanced Computer Science And Applications. 2021; 12(6): 539–545.

[24] Gurjar R, Sahana K, Sathish BS. Stroke Risk Prediction Using Machine Learning Algorithms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2022: 20-25. doi: 10.32628/CSEIT2283121.

[25] Tavares J-A. Stroke prediction through Data Science and Machine Learning Algorithms. 2021; doi: 10.13140/RG.2.2.33027.43040.

[26] Martín J.R, Ayala J.L, Roselló G.R and Camarasaltas J. M. Comparison of Different Machine Learning Approaches to Model Stroke Subtype Classification and Risk Prediction. Spring Simulation Conference