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 4, March 2024

Brain Stroke Prediction

Siddharth Sonawane
AISSMS Polytecnic, Pune, India

Abstract: *A study focused on predicting the likelihood of a stroke occurring at an early stage using both deep learning and machine learning techniques. Stroke is highlighted as a critical medical emergency, with the potential for severe consequences such as long-term neurological damage and death. Early detection of stroke warning symptoms is emphasized as crucial for reducing its severity.*

To evaluate the effectiveness of the prediction algorithm, the study utilized a dataset sourced from Kaggle, a platform for data science competitions. Several classification models were employed, including popular machine learning algorithms such as Extreme Gradient Boosting (XGBoost), Ada Boost, Light Gradient Boosting Machine, Random Forest, Decision Tree, Logistic Regression, K Neighbours, SVM with a Linear Kernel, and Naive Bayes. Additionally, deep neural networks, specifically three-layer and four-layer artificial neural networks (ANN), were employed for classification tasks.

Among the machine learning classifiers, the Random Forest model achieved the highest classification accuracy of 99%. This indicates that Random Forest was exceptionally effective in predicting the likelihood of stroke based on the dataset features. Furthermore, the four-layer deep neural network (4-Layer ANN) surpassed the performance of the three-layer ANN, achieving an accuracy of 92.39% when utilizing selected features as input.

Interestingly, despite the success of both machine learning and deep learning techniques, the research findings suggested that machine learning methods generally outperformed deep neural networks in this specific study. This insight highlights the importance of carefully selecting the appropriate modeling approach based on the nature of the data and the task at hand. In the context of predicting stroke occurrence, machine learning algorithms like Random Forest demonstrated superior performance compared to deep neural networks. However, further investigation and experimentation may be necessary to fully understand the reasons behind this performance discrepancy and to refine the predictive models for stroke risk assessment.

Keywords: Deep Neural Networks, Extreme Gradient Boosting, Machine Learning for Stroke Prediction, deep learning

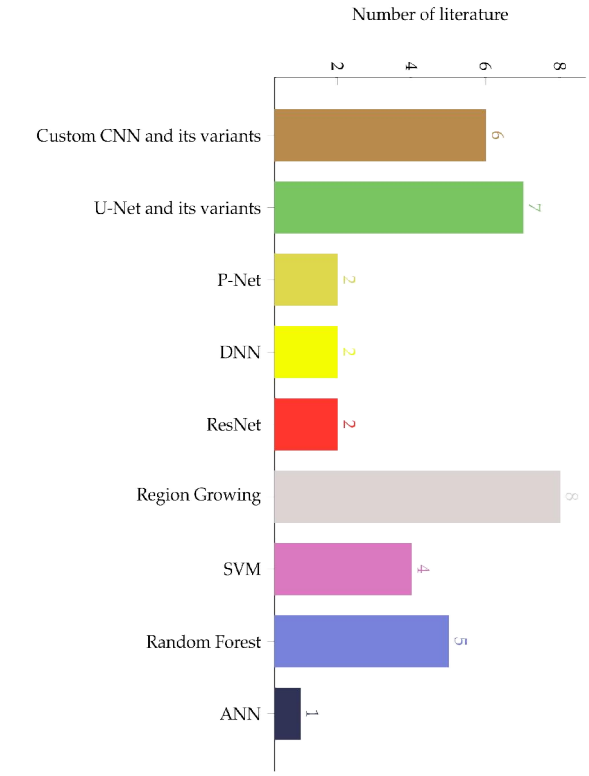
I. INTRODUCTION

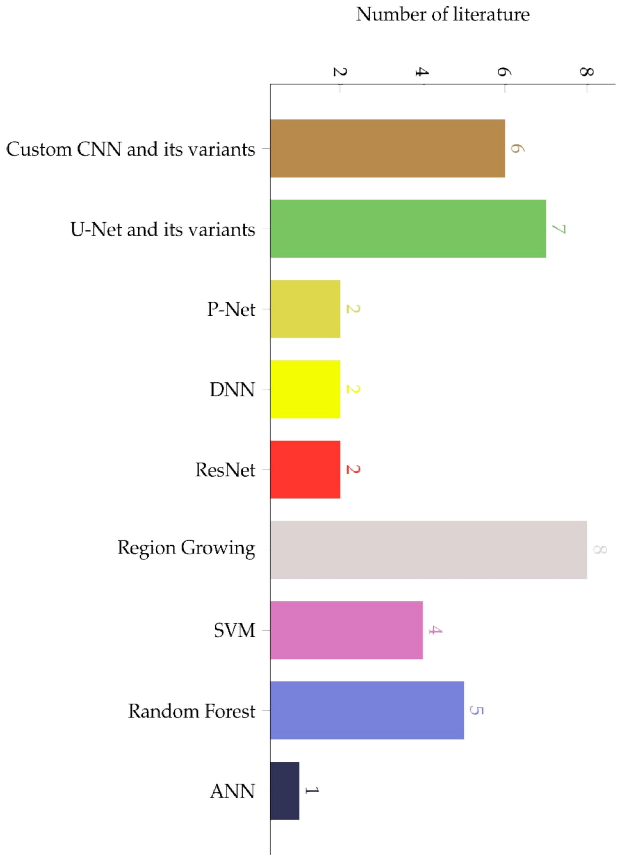
The intricate functioning of different body parts forms the cornerstone of human existence. Among the perilous conditions that can abruptly terminate lives is stroke, a condition commonly detected after the age of 65. Just as heart attacks disrupt the functioning of the heart, strokes similarly impact the brain. Whether due to a blockage in blood supply or the rupture and bleeding of brain blood vessels, strokes impede the flow of essential blood and oxygen to the brain's tissues. This dire circumstance ranks as the fifth leading cause of death in both industrialized and developing nations.

The key to a stroke victim's potential recovery lies in the swift administration of medical care. Urgent attention from healthcare professionals can significantly enhance the chances of a full recovery, as delaying treatment may lead to fatal outcomes, permanent disabilities, and enduring brain damage. Stroke may manifest for various reasons, with contributing factors including dietary habits, sedentary lifestyles, alcohol and tobacco consumption, personal medical history, and associated complications, as outlined by the National Heart, Lung, and Blood Institute.

In essence, the prevention, recognition, and timely treatment of strokes are paramount in safeguarding individuals from its debilitating effects. Understanding the interconnected factors influencing stroke occurrence underscores the necessity for comprehensive approaches to mitigate risk factors and ensure prompt medical intervention, thereby preserving human health and longevity.

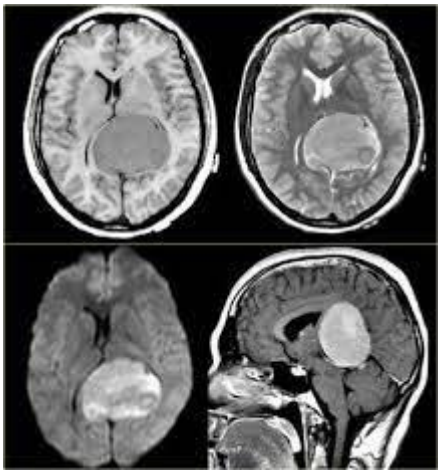
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, March 2024





DOI: 10.48175/568

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 4, March 2024



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 a crucial step before constructing a model, aimed at preparing the dataset by addressing issues like noise, outliers, and missing values that could hinder the model's effectiveness during training. This phase ensures that the data is cleaned and processed to optimize model development. In the context of the study, which focuses on predicting strokes, the dataset comprises twelve attributes, with the 'id' column deemed irrelevant for model creation and thus ignored.

The pre-processing begins with checking for and addressing null values in the dataset. For instance, if any null values are found in the 'BMI' column, they are filled using the "most frequent" value from the data column. Additionally, to facilitate model training, string literals in the dataset, such as those in the 'gender,' 'ever married,' 'work type,' 'Residence type,' and 'smoking status' columns, are encoded into integer values using label encoding. This transformation ensures that the computer can understand the data, as it typically works with numerical inputs.

One significant challenge encountered during pre-processing is the severe imbalance in the stroke prediction dataset. With 5110 rows, only 249 hint at the likelihood of a stroke, while 4861 indicate its absence. While using such imbalanced data may initially inflate accuracy metrics, it could compromise other crucial metrics like recall and precision. To address this issue, the Random Oversample (ROS) approach is employed to balance the dataset, ensuring that both classes (stroke and no-stroke) have the same number of instances.

By implementing the ROS approach, the imbalanced data is effectively addressed, paving the way for the development of a more efficient and accurate model. This step is essential for ensuring that the model is trained on balanced data, thereby reducing the risk of biased predictions and ensuring more reliable stroke predictions in practice. Overall, data pre-processing plays a critical role in optimizing the dataset for model training and enhancing the overall performance and reliability of the predictive model.

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.

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 4, March 2024

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