JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Stroke Identification Using Machine Learning-Based Diagnostic Model

[1]Mala K, [2]Vinay T P,[3] Faizan Razak S ,[4]Deepika B

[1] *Assistant professor,Department Of ISE,CIT,Gubbi,Tumakuru*
[2] *Assistant professor,Department Of ISE,CIT,Gubbi,Tumakuru*
[3] *Student, Department Of ISE,CIT,Gubbi,Tumakuru*
[4] *Student, Department Of ISE,CIT,Gubbi,Tumakuru*

Abstract : In order to lessen the impact of cerebrovascular disorders on public health, early detection and intervention are essential. These diseases, which include stroke, are important causes of death and disability globally. Advances in machine learning (ML) have demonstrated potential in the identification of strokes, with the goal of supporting healthcare providers in making well-informed decisions about treatment and prophylactic measures. This study presents two approaches: one involved using brain CT images in conjunction with a genetic algorithm and a bidirectional long short-term memory (BiLSTM) todevelop an early stroke detection system that achieved 96.5% accuracy. The other investigated the use ofrobust machine learning (ML) algorithms, such as logistic regression (LR), random forest (RF), and K- nearest neighbor (KNN), for the exceptional performance of 99% detection accuracy using ML.the ADASYN oversampling method combined with the RF algorithm. In the end, these methods improve patient outcomes by providing useful instruments for boosting clinical diagnosis and decision-making in stroke care.

IndexTerms - BiLSTM, Logistic regression (LR), Random forest (RF), Knearest neighbor (KNN)

I. INTRODUCTION

Stroke is a worldwide health emergency that results in millions of fatalities each year and severely incapacitates survivors. Since hemorrhagic strokes are more severe and account for a larger percentage of stroke cases, prompt detection and intervention are critical. Ischemic strokes are mostly caused by blocked or restricted blood arteries. This work attempts to create systematic methods for stroke diagnosis by applying machine learning techniques to a variety of datasets, including genetic, lifestyle, and electronic health record data.

The project aims to support doctors in identifying patients at risk of stroke and promoting informed decision-making for prevention and treatment through feature selection algorithms and machine learning models. Additionally, this study aims to offer significant insights for clinical practice by assessing the efficacy of suggested algorithms and contrasting them with proven clinical risk detection techniques, thereby transforming stroke prevention and enhancing treatment outcomes.

The creation of diagnostic systems, the extraction of features from CT scans, the comparison of different machine learning models, and the provision of early detection to assist medical practitioners are some of the major achievements. Additionally, by using machine learning models on publically accessible datasets to predict stroke risk and improve clinical decision-making, this study addresses the pressing need for efficient stroke detection and treatment strategies. By employing several machine learning models and oversampling strategies, this study seeks to provide a new framework for effective stroke identification, thereby making progress in stroke prevention and treatment strategies

II. RELATED WORK

Researchers have explored several machine learning (ML) techniques to predict strokes, aiming to improve diagnosis precision and speed up treatment decisions. These techniques include Random Forest, K-Nearest Neighbors, Support Vector Machine, Naïve Bayes, and Deep Neural Networks. However, inadequate physiological data and class imbalance in patient datasets have presented some challenges. To address these issues, researchers have employed hybrid ML methodologies, artificial neural networks (ANNs), and advanced optimization techniques like Randomized-Hyper opt to improve predictive accuracy.

Integrating clinical indicators with ML models also helps to represent the dynamic nature of stroke risk factors, which include variables such as the prevalence of hypertension, body mass index, heart disease, and smoking status. Additionally, the ability to differentiate between ischemic and hemorrhagic strokes through the appropriate preparation of CT scan data for stroke subtype classification has been made possible by developments in image processing techniques combined with machine learning algorithms. These initiatives demonstrate how crucial it is to improve image quality to improve diagnostic precision. Researchers have also looked into the possibility of using electronic health records to identify strokes, but there are still issues with the data's complete coverage of risk variables. Additionally, efforts have been made to predict the recurrence of stroke over long periods

Current research projects are utilizing novel approaches and machine learning techniques to improve stroke prediction models. The aim is to optimize the effectiveness of stroke prediction and management strategies by improving diagnostic accuracy, addressing concerns related to class imbalance, and integrating thorough risk factor assessments. Researchers are promoting interdisciplinary collaborations and utilizing advanced machine learning algorithms and healthcare data to develop more efficient approaches for stroke prevention and treatment, customized to the specific requirements of each patient and the local healthcare environment.

The collective efforts in stroke prediction have seen extensive exploration of machine learning (ML) algorithms and methodologies. Researchers have strived to improve prediction accuracy and mitigate computational complexity. Various ML models, including Random Forest, Support Vector Machine, and Naïve Bayes, have been employed, showcasing promising results in predicting stroke risk and outcomes. Studies have addressed challenges such as class imbalance and incomplete datasets, utilizing techniques like hyper parameter optimization and hybrid ML approaches. Additionally, there's a focus on integrating diverse physiological and clinical factors, including hypertension, body mass index, and prior stroke history, to enhance prediction models. Furthermore, advancements in image processing techniques and the application of ML algorithms on medical imaging data have contributed to effective stroke subtype classification and recurrence prediction. However, challenges persist, including the need for larger, more comprehensive datasets encompassing a broader range of risk factors to improve the robustness and generalizability of stroke prediction models. Efforts in this direction signify ongoing endeavors to refine stroke diagnosis and management through innovative ML-based approaches.

The collective efforts in stroke prediction have seen extensive exploration of machine learning (ML) algorithms and methodologies. Researchers have strived to improve prediction accuracy and mitigate computational complexity. Various ML models, including Random Forest, Support Vector Machine, and Naïve Bayes, have been employed, showcasing promising results in predicting stroke risk and outcomes. Studies have addressed challenges such as class imbalance and incomplete datasets, utilizing techniques like hyper parameter optimization and hybrid ML approaches. Additionally, there's a focus on integrating diverse physiological and clinical factors, including hypertension, body mass index, and prior stroke history, to enhance prediction models. Furthermore, advancements in image processing techniques and the application of ML algorithms on medical imaging data have contributed to effective stroke subtype classification and recurrence prediction. However, challenges persist, including the need for larger, more comprehensive datasets encompassing a broader range of risk factors to improve the robustness and generalizability of stroke prediction models. Efforts in this direction signify ongoing endeavors to refine stroke diagnosis and management through innovative ML-based approaches.

Researchers in the field of stroke prediction have extensively explored a multitude of machine learning (ML) techniques to enhance diagnostic accuracy and streamline decision-making processes. These endeavors encompass a wide range of methodologies, including ensemble models, deep learning architectures, and optimization algorithms. Despite facing challenges such as class imbalance and incomplete datasets, recent studies have demonstrated promising advancements in stroke prediction by leveraging ML approaches. By integrating diverse physiological parameters and clinical variables, researchers aim to develop more robust and adaptable prediction models. Moreover, efforts to improve image processing techniques and optimize model performance contribute to the ongoing evolution of stroke diagnosis and management strategies. Nevertheless, the pursuit of larger and more comprehensive datasets remains crucial to further refine the efficacy and applicability of ML-based stroke prediction systems.

In the realm of stroke prediction, researchers have diligently explored a plethora of machine learning (ML) methodologies to refine diagnostic capabilities and streamline prognostic outcomes. This collective effort has seen the deployment of diverse ML algorithms, including Random Forest, Support Vector Machines, and Neural Networks, each offering unique insights into stroke risk assessment and patient outcomes. Despite grappling with challenges like data incompleteness and class imbalances, recent studies have showcased notable advancements in stroke prediction through innovative ML-based approaches. By incorporating a comprehensive array of clinical variables and physiological markers, researchers endeavor to construct more robust and adaptable prediction models. Furthermore, ongoing endeavors to optimize model performance and enhance image processing techniques underscore the relentless pursuit of precision in stroke diagnosis and management strategies. However, the quest for larger, more representative datasets remains pivotal to fostering the continued evolution and refinement of ML-driven stroke prediction systems.

III. METHODOLOGY

The methodology successfully predicts cerebral strokes by combining sophisticated machine learning techniques and algorithms with careful data preprocessing. the dataset is obtained and goes through data preprocessing to fix imbalances and missing values. Robust characteristics are extracted from RGB photos using five convolutional neural network (CNN) architectures: AlexNet, Inception V3, VGG 19, NASNet-Large, and ShuffleNet. When selecting features, genetic algorithms are used the features according to their ideal relevance. Furthermore, temporal dependencies are captured by a traditional LSTM model for sequential data analysis. Effective information propagation and processing are made possible by the LSTM architecture, which has gates for input, output, and forget mechanisms. For temporal modeling, BiLST combines forward and backward information flow next step in machine learning involves converting quantitative features into numerical representations by encoding the dataset. To ensure accurate representation of minority groups and to balance class imbalance , These techniques include oversampling using algorithms such as SMOTE, ADASYN, and ROSE. Feature selection is also to improve predicted accuracy and reduce computing complexity. The Select KBest method is an effective way to perform feature selection. It uses chi-square tests to produce a condensed dataset for model training. most pertinent features when feature selection. Effective feature selection using the Select KBest method uses chi-square tests to produce a condensed dataset for model training.

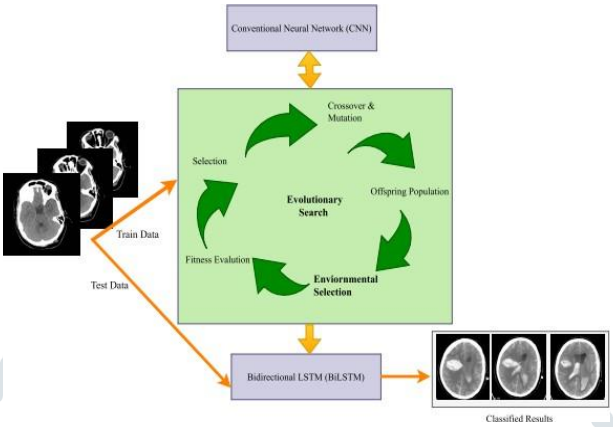

Fig 1 : Framework of the proposed model

Three different machine learning techniques, namely Logistic Regression (LR), K-Nearest Neighbor (KNN), and Random Forest (RF), were utilized to classify strokes. KNN measures similarity using casebased learning to categorize records, while RF uses ensemble learning with decision trees to improve the accuracy of the predictions. On the other hand, LR predicts the likelihood of a stroke occurrence and is suitable for categorical dependent variables. Performance evaluation is conducted using measures suchas accuracy,

precision, recall, and F1-score to ensure a comprehensive model assessment. This methodology combines meticulous data preparation with cutting-edge machine-learning techniques to create an accurate and efficient stroke prediction system.

IV. .RESULTS AND DISCUSSION

The study aimed to address the issue of imbalanced data by utilizing machine learning techniques to predict strokes. The researchers balanced the dataset for training RF, KNN, and LR models by utilizing oversampling methods such as SMOTE, ADASYN, and ROSE. The class distribution balancing results obtained from ADASYN were the most favorable. In correlation analysis, age and hypertension were found to be significantly associated with stroke risk, while BMI showed a weak connection. The ML models showed a significant increase in accuracy when evaluated on oversampled data, with ADASYN consistently outperforming alternative techniques.

Models trained on oversampled data, especially with ADASYN, performed better in the confusion matrix analysis. Among the models tested, ADASYN_RF continuously had the best accuracy. Additional assessment criteria, including F1-score, precision, and recall, validated the efficacy of ADASYN oversampling in improving model performance.

By comparing the proposed ADASYN_RF model with previous research, it was found that it performed better, with an astounding accuracy rate of 99%. The large dataset and strong oversampling methods enhanced the generalizability and dependability of the model. To further increase the accuracy of stroke prediction, future study may investigate expanding to multi-class classification and integrating visual data.The study carried out an in-depth analysis of a proposed model, comparing it with other classifiers and deep learning approaches. The results showed that the model was effective in identifying significant indicators for stroke risk. The model combined neural network-based genetic algorithms with LSTM and BiLSTM architectures to address issues related to data imbalance and small sample sizes, demonstrating its potential to diagnose stroke efficiently.

The GA_BiLSTM model proposed in this study exhibited a significant improvement in accuracy compared to earlier stroke prediction models, showcasing the ability of advanced machine learning techniques in stroke prediction tasks. Overall, this work contributes valuable insights to the field of stroke prediction by adopting a rigorous evaluation methodology and leveraging cutting-edge machine learning breakthroughs.

V. CONCLUSION

In recent years, machine learning methods have been integrated with medical data to improve stroke diagnosis and prediction. Scientists have made significant progress toward increasing prediction accuracy and model efficiency through algorithmic modification and methodical experimentation. With the application of ensemble approaches and creative oversampling strategies, more trustworthy stroke prediction models are made possible. These methods tackle intrinsic problems such as imbalances in class composition and incomplete datasets. The combination of deep learning frameworks with image-derived datasets signifies a noteworthy advancement in stroke detection techniques. Researchers have made impressive progress in identifying pertinent aspects of brain pictures through the application of sophisticated algorithms such as BiLSTM networks and genetic algorithms. These advancements in stroke detection techniques could lead to precise stroke prediction.

However, to ensure the practical use of these cutting-edge solutions in clinical settings and to build trust among healthcare professionals, it is crucial to increase dataset sizes, improve data quality, and apply explainable AI approaches.

VI. REFERENCES

[1] World Stroke Organization. (2023). Vision Impairment and Blindness. Accessed: Feb. 3, 2023. [Online]. Available: https://www.worldstroke.org/world-stroke-day-campaign/whystroke-matters/learn-aboutstroke

[2] Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, ‘‘An integrated machine learning approach to stroke prediction,’’ in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2010, pp. 183–192.

[3] Centers for Disease Control Prevention. (2023). Stroke Facts. Accessed: Mar. 25, 2023. [Online]. Available: https://www.cdc.gov/stroke/facts.htm

[4] Y. Xu, K. Li, Y. Zhao, L. Zhou, Y. Liu, and J. Zhao, ‘‘Role of ferroptosis in stroke,’’ Cellular Mol. Neurobiol., vol. 43, no. 1, pp. 205–222, Jan. 2023.

[5] X. Li and G. Chen, ‘‘CNS-peripheral immune interactions in hemorrhagic stroke,’’ J. Cerebral Blood Flow Metabolism, vol. 43, no. 2, pp. 185–197, Feb. 2023.

[6] Y. Yang, J. Zheng, Z. Du, Y. Li, and Y. Cai, ‘‘Accurate prediction of stroke for hypertensive patients based on medical big data and machine learning algorithms: Retrospective study,’’ JMIR Med. Informat., vol. 9, no. 11, Nov. 2021, Art. no. e30277.

[7] M. Chun, R. Clarke, B. J. Cairns, D. Clifton, D. Bennett, Y. Chen, Y. Guo, P. Pei, J. Lv, C. Yu, L. Yang, L. Li, Z. Chen, and T. Zhu, ‘‘Stroke risk prediction using machine learning: A prospective cohort study of 0.5 million Chinese adults,’’ J. Amer. Med. Inform. Assoc., vol. 28, no. 8, pp. 1719– 1727, Jul. 2021.

[8] G. Ntaios, D. Sagris, A. Kallipolitis, E. Karagkiozi, E. Korompoki, E. Manios, V. Plagianakos, K. Vemmos, and I. Maglogiannis, ‘‘Machinelearning-derived model for the stratification of cardiovascular risk in patients with ischemic stroke,’’ J. Stroke Cerebrovascular Diseases, vol. 30, no. 10, Oct. 2021, Art. no. 106018.

[9] K. Uchida, J. Kouno, S. Yoshimura, N. Kinjo, F. Sakakibara, H. Araki, and T. Morimoto, ‘‘Development of machine learning models to predict probabilities and types of stroke at prehospital stage: The Japan urgent stroke triage score using machine learning (JUST-ML),’’ Transl. Stroke Res., vol. 13, no. 3, pp. 370–381, Jun. 2022.

[10] H. Z. U. Rehman, H. Hwang, and S. Lee, ‘‘Conventional and deep learning methods for skull stripping in brain MRI,’’ Appl. Sci., vol. 10, no. 5, p. 1773, Mar. 2020.

[11] M. Lim, A. Abdullah, N. Z. Jhanjhi, M. K. Khan, and M. Supramaniam, ‘‘Link prediction in time-evolving criminal network with deep reinforcement learning technique,’’ IEEE Access, vol. 7, pp. 184797–184807, 2019.

[12] M. Lim, A. Abdullah, N. Jhanjhi, and M. Supramaniam, ‘‘Hidden link prediction in criminal networks using the deep reinforcement learning technique,’’ Computers, vol. 8, no. 1, p. 8, Jan. 2019.

[13] S. Bacchi, T. Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, and J. Jannes, ‘‘Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes,’’ Academic Radiol., vol. 27, no. 2, pp. e19–e23, Feb. 2020.

[14] Fang, G.; Liu, W.; Wang, L. A machine learning approach to select features important to stroke prognosis. Comput. Biol. Chem. 2020, 88, 107316. [CrossRef] [PubMed]

[15] Park, S.J.; Hussain, I.; Hong, S.; Kim, D.; Park, H.; Benjamin, H.C.M. Real-time gait monitoring system for consumer stroke prediction service. In Proceedings of the 2020 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 4–6 January 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.

[16] Rajora, M.; Rathod, M.; Naik, N.S. Stroke prediction using machine learning in a distributed environment. In Proceedings of the Distributed Computing and Internet Technology: 17th International Conference, ICDCIT 2021, Bhubaneswar, India, 7–10 January 2021; Proceedings 17; Springer: Berlin/Heidelberg, Germany, 2021; pp. 238–252.

[17] Tazin, T.; Alam, M.N.; Dola, N.N.; Bari, M.S.; Bourouis, S.; Monirujjaman Khan, M. Stroke disease detection and prediction using robust learning approaches. J. Healthc. Eng. 2021, 2021, 7633381. [CrossRef] [PubMed]

[18] Govindarajan, P.; Soundarapandian, R.K.; Gandomi, A.H.; Patan, R.; Jayaraman, P.; Manikandan, R. Classification of stroke disease using machine learning algorithms. Neural Comput. Appl. 2020, 32, 817–828. [CrossRef]

[19] Dritsas, E.; Trigka, M. Stroke risk prediction with machine learning techniques. Sensors 2022, 22, 4670. [CrossRef] 7. Liu, T.; Fan, W.; Wu, C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. Artif. Intell. Med. 2019, 101, 101723. [CrossRef]

[20] Liu, Y.; Yin, B.; Cong, Y. The probability of ischaemic stroke prediction with a multi-neuralnetwork model. Sensors 2020, 20, 4995. [CrossRef]

[21] Ali, A.A. Stroke prediction using distributed machine learning based on Apache spark. Stroke 2019, 28, 89–97

[22] Wu, Y.; Fang, Y. Stroke prediction with machine learning methods among older Chinese. Int. J. Environ. Res. Public Health 2020, 17, 1828. [CrossRef]

[23] Xie, Y.; Yang, H.; Yuan, X.; He, Q.; Zhang, R.; Zhu, Q.; Chu, Z.; Yang, C.; Qin, P.; Yan, C. Stroke prediction from electrocardiograms by deep neural network. Multimed. Tools Appl. 2021, 80, 17291–17297. [CrossRef]

[24] Kim, J.K.; Choo, Y.J.; Chang, M.C. Prediction of motor function in stroke patients using machine learning algorithm: Development of practical models. J. Stroke Cerebrovasc. Dis. 2021, 30, 105856. [CrossRef] [PubMed]