A Machine Learning-Based Framework for Automated Stroke Detection Using Neuroimaging Data

Aseena Shaik Babu[1], Vinay Kumar Yarraguntla[2], Haridasu Sujana[3], Chigurambotla Anjali[4], Vardhineni Jitender[5]

[1] Assistant Professor, Department of Computer Science and Engineering(AI & ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

[2,3,4,5]BTech Students ,Department of Computer Science and Engineering(AI & ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

***Abstract*—Stroke, a common cerebrovascular disease, is one of the leading causes of death and long-term disability worldwide, yet it can often be managed effectively if detected early. Timely diagnosis plays a crucial role in enabling rapid medical intervention and improving patient outcomes. In this study, an intelligent diagnostic system is proposed for early stroke detection using brain CT images, leveraging the capabilities of machine learning techniques. The model integrates a genetic algorithm with a Bidirectional Long Short-Term Memory (BiLSTM) network to enhance prediction accuracy. The genetic algorithm is employed to select the most relevant features from the imaging data, reducing unnecessary information and improving the efficiency of the classification process. These optimized features are then fed into the BiLSTM model, which captures complex patterns and dependencies to identify stroke conditions at an early stage. The system's performance was evaluated using cross-validation along with key metrics such as accuracy, precision, recall, and F1-score. The proposed model achieved an accuracy of 96.6%, outperforming traditional algorithms such as K-Nearest Neighbors, Naive Bayes, and Support Vector Machines. Overall, the system demonstrates strong potential in supporting healthcare professionals with accurate and timely stroke diagnosis.**

***Keywords*—Stroke Detection, Neuroimaging, Computed Tomography (CT), Machine Learning, Genetic Algorithm, Bidirectional LSTM (BiLSTM), Feature Selection, Medical Image Classification, Early Diagnosis, Healthcare AI**

I. INTRODUCTION

Stroke is a serious medical condition that occurs when the blood supply to the brain is disrupted due to blockage or rupture of blood vessels. It remains one of the leading causes of death and long-term disability worldwide, affecting millions of individuals every year, as highlighted by Centers for Disease Control and Prevention [2]. Survivors often face significant physical and cognitive challenges that reduce their quality of life. However, early diagnosis and preventive care can greatly reduce its severity and long-term impact. Research has shown that understanding biological mechanisms such as ferroptosis can further improve stroke management strategies [3]. In recent years, integrating data-driven approaches like machine learning has gained attention for improving early prediction and prevention of stroke-related risks [1].

Stroke is generally classified into two major types: ischemic and hemorrhagic. Ischemic stroke is the most common, accounting for nearly 87% of all cases, and occurs when blood flow to the brain is obstructed due to clots or fatty deposits [2]. In contrast, hemorrhagic stroke occurs when a blood vessel ruptures, leading to internal bleeding within the brain. Although less frequent, hemorrhagic strokes are often more severe and associated with higher mortality rates. Recent studies have explored the complex interactions between the central nervous system and immune responses in hemorrhagic stroke, providing deeper insights into its severity and outcomes [4]. Understanding these distinctions is essential for accurate diagnosis and effective treatment planning, as well as for developing advanced predictive models that can differentiate between stroke types and support clinical decision-making [7].

In recent years, the increasing availability of healthcare data has encouraged the use of machine learning techniques in disease prediction and diagnosis. Several studies have demonstrated that machine learning models can effectively analyze large-scale patient data to predict the likelihood of stroke occurrence [5], [6]. These models utilize information such as electronic health records, genetic profiles, and lifestyle factors to identify individuals at higher risk. Feature selection techniques are particularly important in this context, as they help extract the most relevant attributes from complex datasets, improving both accuracy and computational efficiency. Additionally, deep learning approaches have shown promising results in analyzing neuroimaging data, enabling more precise detection of stroke-related abnormalities and enhancing the overall reliability of diagnostic systems [12].

Medical imaging, especially CT and MRI scans, plays a vital role in stroke diagnosis and assessment. Recent advancements in image processing and deep learning have significantly improved the ability to detect stroke-related conditions at an early stage. Techniques such as automated feature extraction and image classification have been applied to identify intracranial hemorrhages and other abnormalities with high accuracy [17], [18]. Furthermore, deep learning-based systems have been developed to enhance the detection of hemorrhagic lesions in CT images, providing more reliable support for clinical evaluation [16]. These innovations not only improve diagnostic accuracy but also reduce the time required for analysis, which is crucial in emergency situations where rapid intervention can save lives and minimize brain damage.

The primary goal of this study is to develop an efficient and scalable machine learning-based system for early stroke detection and prediction. The proposed model focuses on improving accuracy while maintaining interpretability, making it suitable for real-world clinical applications. Various machine learning techniques are evaluated and compared using real-world datasets to assess their effectiveness in predicting stroke risk [14]. In addition, recent research has shown that enhancing deep learning algorithms can significantly improve the detection of intracranial hemorrhages from CT images [19]. By combining advanced machine learning methods with medical imaging analysis, this study aims to support healthcare professionals in making timely and informed decisions, ultimately contributing to better patient outcomes and more effective stroke prevention strategies.

II. LITERATURE SURVEY

Khosla et al., [2010] [1] proposed an integrated machine learning framework for predicting stroke risk using clinical and demographic data. Their study focused on combining multiple algorithms to improve prediction accuracy and reliability in healthcare applications. The authors utilized data mining techniques to analyze patient information and identify key risk factors associated with stroke. They emphasized the importance of preprocessing and feature engineering in enhancing model performance. Various classification models were evaluated, and the results showed that combining different approaches could significantly improve prediction outcomes compared to individual models. The study highlighted how machine learning can support early diagnosis by identifying high-risk patients before the onset of stroke. This work laid an important foundation for future research in stroke prediction using data-driven methods. It also demonstrated the potential of integrating multiple machine learning techniques to develop more robust and accurate clinical decision support systems.

Yang et al., [2021] [5] conducted a retrospective study to predict stroke risk among hypertensive patients using machine learning techniques and large-scale medical data. Their research focused on analyzing clinical records to identify patterns that contribute to stroke occurrence. The authors applied various machine learning algorithms to improve prediction accuracy and compared their performance using different evaluation metrics. They found that models trained on comprehensive datasets performed better in identifying high-risk individuals. The study also emphasized the role of big data in healthcare, highlighting how large datasets can enhance the reliability of predictive models. Feature selection played a key role in improving model efficiency by removing irrelevant data. The results demonstrated that machine learning-based approaches can significantly aid in early stroke detection and prevention. This work provides valuable insights into the use of data-driven methods for personalized healthcare and supports the integration of AI systems in clinical practice.

Chun et al., [2021] [6] presented a large-scale study on stroke risk prediction using machine learning, based on data collected from over 500,000 individuals. Their research aimed to develop a reliable predictive model by analyzing a wide range of health-related factors. The study utilized advanced machine learning techniques to identify patterns and correlations within the dataset. The authors highlighted that large population-based data improves the generalizability and robustness of predictive models. They evaluated multiple algorithms and found that machine learning approaches outperformed traditional statistical methods in predicting stroke risk. The study also emphasized the importance of incorporating demographic, lifestyle, and medical variables for accurate prediction. Their findings suggest that machine learning can be effectively used for large-scale public health analysis and early risk assessment. This research contributes to the development of scalable and data-driven solutions for stroke prevention and healthcare planning.

Bacchi et al., [2020] [12] explored the application of deep learning techniques in predicting functional outcomes for patients undergoing treatment for ischemic stroke. Their study focused on using medical imaging data to train deep neural networks capable of analyzing complex patterns in brain scans. The authors demonstrated that deep learning models could provide accurate predictions regarding patient recovery and treatment effectiveness. They highlighted the advantage of automated feature extraction, which eliminates the need for manual intervention and improves efficiency. The study also compared deep learning methods with traditional approaches, showing superior performance in terms of accuracy and reliability. Their work emphasized the growing importance of AI in medical imaging and clinical decision-making. It also illustrated how deep learning can assist healthcare professionals in evaluating treatment options and predicting patient outcomes more effectively, thereby improving the overall quality of stroke care.

Yeo et al., [2023] [19] investigated techniques to enhance the performance of deep learning algorithms for detecting intracranial hemorrhage using CT brain images. Their study focused on improving the accuracy and robustness of automated diagnostic systems through model optimization and data enhancement strategies. The authors evaluated different deep learning architectures and preprocessing methods to determine their impact on detection performance. They found that refining model parameters and improving data quality significantly enhanced the system's ability to identify hemorrhagic regions in brain scans. The study also highlighted the importance of early detection in reducing mortality and improving treatment outcomes for stroke patients. Their findings demonstrate that advanced deep learning techniques can play a crucial role in medical image analysis. This work contributes to the development of more reliable and efficient AI-based diagnostic tools, supporting faster and more accurate clinical decision-making in stroke management.

III. DATASET DETAILS

The dataset used in this project consists of brain CT scan images collected for the purpose of detecting stroke conditions at an early stage. These images are organized into two main categories: normal and stroke, allowing the model to learn and differentiate between healthy brain structures and abnormal patterns caused by stroke. The dataset is structured in a well-organized format, making it suitable for machine learning and deep learning applications. Each image represents important visual features such as tissue density, bleeding regions, or blocked blood flow areas, which are critical indicators for stroke diagnosis. The dataset serves as the foundation for training and evaluating multiple models including KNN, Naive Bayes, SVM, and advanced architectures like GA-BiLSTM and CNN-BiLSTM-Attention. Proper understanding of these medical images is essential, as even minor variations in patterns can significantly impact classification results and overall model performance .

Before applying machine learning algorithms, several preprocessing steps are performed to enhance the dataset quality and ensure better model performance. The images are resized and normalized to maintain consistency and improve computational efficiency. Feature extraction techniques are applied to capture important patterns from the CT scans, and a genetic algorithm is used to select the most relevant features for classification. The dataset is then divided into training and testing sets, typically using an 80:20 ratio, to evaluate the model's ability on unseen data. This helps in reducing overfitting and ensures reliable predictions. Additionally, data is shuffled to maintain a balanced distribution between normal and stroke classes. These preprocessing steps play a crucial role in improving accuracy and stability. The prepared dataset enables the system to effectively train models and achieve high performance in stroke detection tasks.

IV. PROPOSED METHODOLOGY

The proposed system is designed in a step-by-step manner to detect stroke at an early stage using brain CT scan images. Initially, the dataset consisting of categorized images (normal and stroke) is collected and carefully examined. To prepare the data for model training, preprocessing steps such as image resizing, normalization, and feature extraction are carried out to ensure consistency and improve data quality. A genetic algorithm is then used to identify the most important features from the dataset, helping to reduce unnecessary information and enhance model performance. The dataset is shuffled to maintain a balanced distribution of classes and is split into training and testing sets, typically using an 80:20 ratio. This division allows the system to learn from the training data and later evaluate its performance on new, unseen data.

Once the data preparation is complete, different classification algorithms including KNN, Naive Bayes, SVM, and advanced models like GA-BiLSTM and CNN-BiLSTM-Attention are applied. Each model is trained using the prepared dataset and tested to measure its effectiveness. The GA-BiLSTM model uses optimized features to capture deeper patterns within the data, resulting in better prediction accuracy. The performance of all models is evaluated using metrics such as accuracy, precision, recall, and F1-score. Confusion matrices and graphical outputs are also used to understand the prediction results clearly. Among all the methods, GA-BiLSTM and CNN-BiLSTM-Attention provide the best results with higher accuracy. Finally, the trained system is used to classify new CT scan images as either normal or stroke, supporting faster and more reliable diagnosis.

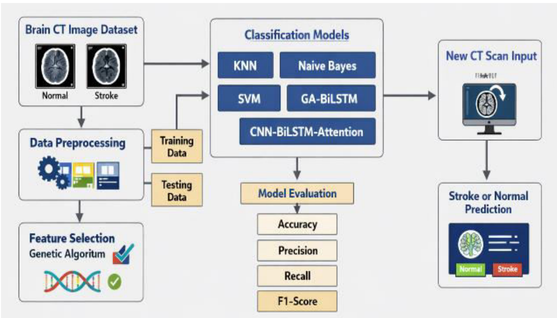

Figure [1]: Stroke Detection System Using Machine Learning and Deep Learning

Figure [1] shows the workflow of the proposed stroke detection system using brain CT images. The dataset is first preprocessed through resizing, normalization, feature extraction, and shuffling, followed by splitting into training and testing sets. Various models such as KNN, SVM, Naive Bayes, GA-BiLSTM, and CNN-BiLSTM-Attention are then trained and evaluated using accuracy, precision, recall, and F1-score. Finally, the system predicts whether a new CT scan image is normal or indicates a stroke.

V. RESULT AND DISCUSSION

The experimental results of this project demonstrate the effectiveness of machine learning and deep learning models in detecting stroke using brain CT images. After preprocessing the dataset and splitting it into training and testing sets, multiple models including KNN, Naive Bayes, SVM, GA-BiLSTM, and CNN-BiLSTM-Attention were trained and evaluated. Among all models, the GA-BiLSTM and CNN-BiLSTM-Attention achieved the highest accuracy of 96%, indicating strong performance in classifying stroke and normal cases. Traditional models such as KNN and SVM showed moderate performance, while Naive Bayes produced comparatively lower accuracy. Evaluation metrics like precision, recall, and F1-score further confirmed the reliability of the proposed models. Confusion matrices showed that most predictions were correctly classified with very few misclassifications. Graphical comparisons clearly highlighted the superior performance of deep learning-based approaches over conventional methods.

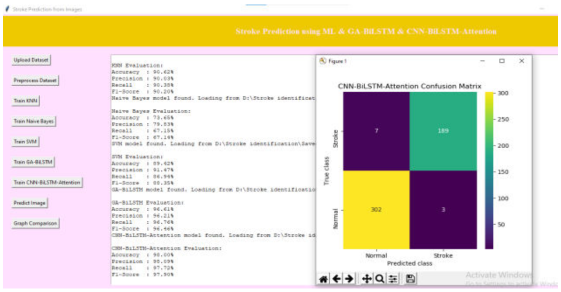

Figure [2]: Proposed CNN-BiLSTM-Attention Model

Figure [2] presents the performance of the proposed CNN-BiLSTM-Attention model for stroke detection. It includes evaluation metrics such as accuracy, precision, recall, and F1-score along with the confusion matrix. The results show that the model achieves high accuracy with most predictions correctly classified, as indicated by the diagonal values in the matrix. Overall, the model demonstrates strong performance in identifying both normal and stroke cases.

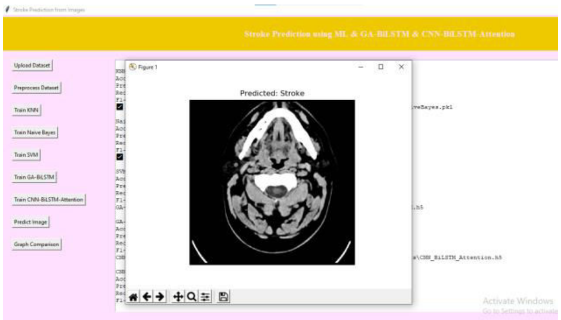

Figure [3]: Stroke Prediction Output Using CNN-BiLSTM-Attention Model

Figure [3] shows the prediction output of the proposed system for a given brain CT image. The model processes the input image and classifies it as “Stroke,” as displayed above the image. This demonstrates the system’s ability to accurately analyze CT scans and provide clear prediction results, helping in early stroke detection and diagnosis.

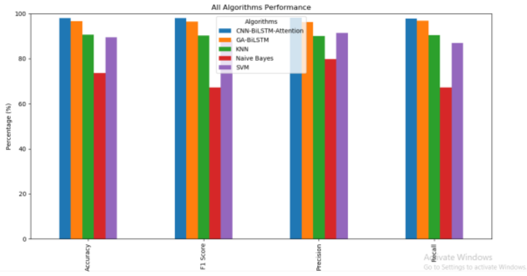

Figure [4]: Performance Comparison of Models

Figure [4] shows the comparison of different models including CNN-BiLSTM-Attention, GA-BiLSTM, KNN, Naive Bayes, and SVM based on evaluation metrics such as accuracy, precision, recall, and F1-score. The graph clearly indicates that CNN-BiLSTM-Attention and GA-BiLSTM achieve the highest performance among all models, while Naive Bayes shows comparatively lower results. This comparison highlights the effectiveness of deep learning-based approaches for stroke detection.

DISCUSSION

The results of this study highlight the importance of advanced machine learning and deep learning techniques in medical image analysis. The GA-BiLSTM and CNN-BiLSTM-Attention models performed better due to their ability to capture complex spatial and temporal features from CT images. Feature selection using a genetic algorithm further improved performance by removing irrelevant data and focusing on important patterns. Traditional models like KNN and SVM showed decent performance but were limited in handling complex image features. Proper preprocessing steps such as normalization and feature extraction played a crucial role in achieving higher accuracy. Visualization tools like confusion matrices and comparison graphs helped in understanding model performance clearly. Overall, the system demonstrates that combining machine learning with deep learning can provide accurate and efficient solutions for early stroke detection, supporting better clinical decisions and improving patient outcomes.

VI. CONCLUSION

This project successfully demonstrates the application of machine learning and deep learning techniques for early stroke detection using brain CT images. Through effective preprocessing steps such as image normalization, resizing, feature extraction, and data shuffling, the dataset was prepared to ensure reliable model training. Multiple algorithms including KNN, Naive Bayes, SVM, GA-BiLSTM, and CNN-BiLSTM-Attention were implemented and evaluated. Among them, the GA-BiLSTM and CNN-BiLSTM-Attention models achieved the best performance with high accuracy and minimal misclassification. The results confirm that advanced hybrid models are more effective in capturing complex patterns in medical imaging data compared to traditional methods. Evaluation metrics such as accuracy, precision, recall, and F1-score, along with confusion matrices and comparison graphs, provided clear insights into model performance. The system is capable of accurately classifying CT images as normal or stroke, supporting faster and more reliable diagnosis. Overall, this work highlights the potential of intelligent systems in healthcare and provides a strong foundation for developing real-time, high-accuracy stroke detection solutions to assist medical professionals in clinical decision-making.

REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] G. Ntaios, D. Sagris, A. Kallipolitis, E. Karagkiozi, E. Korompoki, E. Manios, V. Plagianakos, K. Vemmos, and I. Maglogiannis, ''Machine learning-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.

[8] 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]

[9]H.Z.U.Rehman,H.Hwang,andS.Lee,''C onventional and deeplearning methods for skull stripping in brain MRI,'' Appl. Sci., vol. 10, no. 5, p. 1773, Mar. 2020.

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

[11] 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.

[12] 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.

[13] M. Saritha, K. Paul Joseph, and A. T. Mathew, ''Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network,'' Pattern Recognit. Lett., vol. 34, no. 16, pp. 2151–2156, Dec. 2013.

[14] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, ''A predictive analytics approach for stroke prediction using machine learning and neural networks,'' Healthcare Anal., vol. 2, Nov. 2022, Art. no. 100032.

[15] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, ''An automated diagnostic system for heart disease prediction based $\chi 2$ on chi sequare statistical model and optimally configured deep neural network,'' IEEE Access, vol. 7, pp. 34938–34945, 2019.

[16] A. Phaphuangwittayakul, Y. Guo, F. Ying, A. Y. Dawod, S. Angkurawara non, and C. Angkurawaranon, ''An optimal deep learning framework for multi-type hemorrhagic lesions detection and quantification in head CT images for traumatic brain injury,'' Int. J. Speech Technol., vol. 52, no. 7, pp. 7320–7338, May 2022.

[17] S. Gudadhe and A. Thakare, ''Classification of intracranial hemorrhage CT images for stroke analysis with transformed and image-based GLCM features,'' Tech. Rep., 2022.

[18] S. Gudadhe, A. Thakare, and A. M. Anter, ''A novel machine learning based feature extraction method for classifying

intracranial hemorrhage computed tomography images,'' Healthcare Anal., vol. 3, Nov. 2023, Art. no. 100196.

[19] M. Yeo, B. Tahayori, H. K. Kok, J. Maingard, N. Kutaiba, J. Russell, V. Thijs, A. Jhamb, R. V. Chandra, M. Brooks, C. D. Barras, and H. Asadi, ``Evaluation of techniques to improve a deep learning algorithm for the automatic detection of intracranial haemorrhage on CT head imaging,'' Eur. Radiol. Experim., vol. 7, no. 1, p. 17, Apr. 2023.