International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 5 ǁ May 2023 ǁ PP. 468-472

A Survey Paper on a Model for Stroke Risk Prediction System

Dr. KUMARASWAMY S
Professor and Head of Department, Department of Computer Science and Engineering

JEEVAN A, ROHITH S, KISHAN P, PRASHANTH T N
Global Academy of Technology Students, Department of Computer Science and Engineering, Global Academy of Technology

Abstract—*Stroke has emerged as a leading cause of mortality and long-term impairment globally, yet effective treatments remain elusive. Deep learning techniques show promise in surpassing existing stroke risk prediction models; however, their success hinges upon access to substantial, well-annotated datasets. Unfortunately, due to stringent privacy regulations in healthcare systems, stroke data is typically fragmented across various hospitals in small fragments. Moreover, the distribution of positive and negative instances within this data is highly imbalanced. To address the issue of limited data, transfer learning presents a viable solution by leveraging knowledge from related domains, particularly when multiple sources of data are available. In this study, with the help of latest approach called the Hybrid Deep Transfer Learning based on Stroke Risk Prediction (HDTL SRP) framework, which harnesses the structural knowledge obtained from multiple interconnected sources, including external stroke data and chronic diseases data such as hypertension and diabetes. Keywords- Stroke Risk Prediction, Hybrid Deep Transfer learning*

Date of Submission: 04-05-2023 Date of acceptance: 15-05-2023

I. INTRODUCTION

Heart disease is a widespread health issue on a global scale, persisting as a primary contributor to mortality rates. As stated by the World Health Organization (WHO), Cardio Vascular Disease (CVD) accounts for about 17.9 million fatalities annually, encompassing approximately thirty one percentage of total worldwide deaths. Early diagnosis and accurate prediction of the severity of heart diseases are critical for effective treatment and improved patient outcomes. Therefore, there is a growing need for efficient and accurate diagnostic methods to predict the severity of heart disease.

Traditionally, doctors use several diagnostic tools, such as physical exams, blood tests, electrocardiograms (ECG), echocardiograms, and stress tests, to diagnose (diag) heart disease .These methods can provide important information about the heart's structure and function, but they may not always be accurate in predicting the severity of heart disease. Furthermore, these methods can be time-consuming, expensive, and require highly specialized personnel to perform and interpret the results.

Recent advancements in technology and Machine Learning (ML) algorithms have made it possible to analyze large volumes of medical data and predict the severity of heart disease with high accuracy. ML algorithms have the capacity to acquire knowledge from medical data, enabling them to detect concealed patterns and correlations that might not be readily discernible to human specialists. By training on a dataset of medical records and diagnostic test results, ML algorithms possess the capability to achieve a remarkable level of precision in predicting the severity of heart disease.

Consequently, the primary objective of this study is to explore and evaluate the efficacy of diverse ML algorithms in accurately predicting the severity of heart diseases. The study will analyze a sample of medical data from patients with heart disease comparison of the performance exhibited by distinct ML algorithms, namely Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), with regards to their predictive capabilities in determining the severity of heart disease. The findings of this research can help healthcare professionals in diagnosing and treating heart diseases more effectively, leading to improved patient outcomes.

II. RESEARCH METHODOLOGY

The primary objective of this investigation was to employ ML algorithms for predicting the severity of heart disease. To accomplish this, a dataset comprising information from 303 patients diag with heart disease was gathered by the researchers from the UCI ML Repository. The dataset contained 13 attributes that were considered relevant to the prediction of the severity of heart disease.

The collected dataset consisted of a comprehensive range of attributes, encompassing factors such as age, sex, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, and thallium heart scan results. These variables were meticulously included to facilitate a thorough analysis of their impact on predicting the severity of heart disease using ML algorithms.

To ensure that the dataset was suitable for use in ML algorithms, it was preprocessed by removing missing values and encoding categorical variables. Subsequently, the dataset was divided into training and testing sets, employing a ratio of 70:30 respectively. This approach ensured that the algorithms could learn from the training data and then test their accuracy on new data.

A variety of ML algorithms, namely LR, Support Vector Machine, and RF, were employed to forecast the severity of heart disease. These algorithms were chosen to leverage their distinct strengths and capabilities in capturing patterns and relationships within the dataset, thus enabling comprehensive and accurate predictions.

Each algorithm was trained on the training set and then tested on the testing set To assess the effectiveness of the employed algorithms in predicting the severity of heart disease, their performance was evaluated utilizing a range of robust metrics, including accuracy, precision, recall, and F1-score. These metrics provided a comprehensive evaluation of the models' predictive capabilities, ensuring a thorough understanding of their performance across multiple dimensions.

The findings of the study revealed that among the utilized ML algorithms, the RF approach exhibited the highest performance, showcasing an impressive accuracy rate of eighty four percentage, precision of eighty four percentage, recall of eighty five percentage, and F1-score of eighty four percentage. These noteworthy results effectively underscore the considerable potential of ML algorithms in accurately predicting the severity of heart disease.

The precision, recall, and F1-score metrics also demonstrated that RF was the most effective algorithm. The precision metric, a fundamental evaluation measure, quantifies the proportion of correctly identified positive predictions relative to all positive predictions made by the algorithm. On the other hand, the recall metric assesses the percentage of accurately identified positive predictions compared to all the actual positive cases present in the dataset. The F1-score, a significant performance indicator, represents the harmonic mean of precision and recall, providing a balanced assessment of the algorithm's predictive accuracy. These metrics collectively offer valuable insights into the algorithm's ability to effectively predict the severity of heart diseases based on its precision, recall, and overall performance.

The operational mechanism of the RF algorithm involves constructing numerous decision trees and amalgamating their outputs to generate a final prediction. Each decision tree is constructed using a subset of the data, and the ultimate prediction is determined by aggregating the majority vote from the individual trees. This ensemble approach empowers the RF algorithm to harness the collective knowledge and expertise of multiple decision trees, resulting in robust and reliable predictions for the severity of heart diseases. This approach can help reduce overfitting, a common issue in ML, and improve the accuracy of the predictions.

The high accuracy achieved by the RF algorithm strongly suggests that ML can serve as a valuable and indispensable tool for predicting the severity of heart diseases. By analyzing large volumes of medical data, ML algorithms can help identify patterns and relationships that may not be immediately apparent to humans. This can lead to more accurate and timely diag, which can improve patient outcomes and reduce healthcare costs.

In conclusion, the study successfully demonstrated that ML algorithms can be used to predict the severity of heart disease. In this study, a dataset comprising 13 carefully preprocessed attributes was employed. The dataset was subsequently divided into training and testing sets, enabling a comprehensive evaluation of the performance exhibited by various ML algorithms. To assess the effectiveness of these algorithms in predicting the severity of heart diseases, a range of evaluation metrics, including accuracy, precision, recall, and F1-score, were meticulously applied. These metrics served as valuable indicators, providing a robust assessment of the algorithms' predictive capabilities and their potential to contribute to accurate and reliable predictions in the context of heart disease severity. The outcomes of the study demonstrated that among the evaluated ML algorithms, the RF approach exhibited superior performance, attaining an impressive accuracy rate of eighty four percentage, precision of eighty four percentage, recall of eighty five percentage, and F1-score of eighty four percentage These exceptional results validate the RF algorithm as a highly effective tool for accurately predicting the severity of heart disease. The findings highlight the potential of ML methodologies in revolutionizing the field of cardiovascular healthcare, offering substantial advancements in diagnosis and treatment strategies for improved patient care. The results obtained from this study strongly indicate that the utilization of ML algorithms holds immense promise in enhancing the precision and efficacy of heart disease diagnosis and treatment. By harnessing the power of these algorithms, healthcare professionals can potentially achieve higher levels of accuracy and efficiency in identifying and managing heart disease cases. This breakthrough in leveraging ML technologies offers great potential for advancing the field of cardiovascular medicine, ultimately leading to improved patient outcomes and more effective healthcare intervention.

III. Literature Survey

In the paper titled "EMR-based phenotyping of ischemic stroke using supervised ML and text mining techniques" (2020), the authors employed various supervised ML algorithms to classify ischemic stroke cases. The advantage of this methodology lies in the comparison of the performance of different ML algorithms. However, it is noted that this approach is suitable only for binary classification and not for multiclass classification.

Another study titled "Feature isolation for hypothesis testing in retinal imaging: An ischemic stroke prediction case study" (2019) utilized a deep learning algorithm for detecting ischemic stroke prediction. Despite achieving relatively low accuracy of less than seventy percentage, the advantage lies in the application of deep learning in the prediction process. However, the time-consuming nature of the approach is a disadvantage.

The paper "Metapred: Meta-learning for clinical risk prediction with limited patient electronic health records" (2019) employed CNN and RNN models to develop a unique model-creation technique for solving real-world problems. While the advantage lies in the ability to handle binary classification tasks, the limitation is its unsuitability for multiclass classification.

In "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets" (2021), the authors utilized PCA with Variational Autoencoders to achieve a ninty percentage accuracy in predicting coronary heart disease risk. However, similar to previous studies, this method is only suitable for binary classification and cannot classify stages of stroke-related information.

previous studies, this method is only suitable for binary classification and cannot classify stages of stroke-related information.

In 2019, a paper was published with the title "The use of deep learning to predict stroke patient mortality," which introduced a novel approach to predicting stroke patient mortality using a scaled principal component analysis (PCA) coupled with a deep neural network (DNN). This approach combined the strengths of a DNN forstudying relevant variables and scaled PCA for generating improved continuous inputs. The advantage of this method is its potential application to predict diseases other than stroke. However, training time consumption and the lack of medication recommendations are notable limitations.

The authors of a 2019 paper titled "A hybrid ML approach to cerebral stroke prediction based on imbalanced medical dataset" proposed a unique two-step process for predicting cerebral stroke. This approach involved using a RF regression algorithm to handle missing data values, followed by an automated hyperparameter optimization method based on a deep neural network to predict the occurrence of stroke. By effectively reducing the false negative rate and achieving a high level of accuracy, this approach has made significant strides towards addressing the issue of misdiagnosis in stroke prediction. Despite its success, however, it is important to note that this method still has a binary classification limitation, as seen in previous studies.

Taken together, these studies showcase the diverse range of ML and deep learning techniques that have been employed for the prediction and phenotyping of ischemic stroke. While they offeradvantages in terms of classification and prediction, such as high accuracy or reduced misdiagnosis rates, limitations existregarding multiclass classification, training time, and suitability for specific medical datasets.

IV. RESULTS

According to the study's findings, the RF algorithm proved to be the most effective approach for accurately predicting the severity of heart disease. Out of the three ML algorithms evaluated, the RF method demonstrated the highest accuracy of eighty six point one four percentage in predicting the severity of heart disease. The SVM model was a close second with an accuracy rate of eighty four point nine three percentage, while the LR approach achieved an accuracy of eighty three point two two percentage.

The evaluation metrics including precision, recall, and F1-score further confirmed that the RF algorithm outperformed the other methods and was the most effective in predicting the severity of heart disease. In order to evaluate the performance of the ML algorithms in predicting the severity of heart disease, several metrics were used. Precision was calculated as the percentage of true positive predictions out of all positive predictions made by the algorithm. The recall metric, on the other hand, measured the percentage of true positive predictions out of all the actual positive cases in the dataset. Finally, the F1-score, which is the harmonic mean of the precision and recall metrics, was used as an overall measure of the effectiveness of the algorithms.

The RF algorithm functions by constructing numerous decision trees and subsequently consolidating them to arrive at a final prediction. This approach is a form of ensemble learning, where the algorithm generates a diverse range of decision trees and aggregates them to enhance the accuracy of the prediction. To prevent overfitting, which is a prevalent challenge in ML, each decision tree in the RF algorithm is created using a different subset of the data. The final prediction is then made by taking the majority vote of the predictions generated by the individual trees. This ensemble approach helps to reduce overfitting and enhances the accuracy of the predictions, making RF a highly effective algorithm for a wide range of applications.

Given the high accuracy obtained by the RF algorithm in predicting the severity of heart disease, it appears that this method could serve as a useful tool in healthcare settings. With its ability to accurately predict the severity of heart disease, healthcare professionals may be able to intervene earlier and provide better care to patients, potentially leading to better health outcomes. Through the analysis of vast amounts of medical data, ML algorithms have the potential to identify patterns and correlations that may not be readily apparent to human analysis. This can ultimately result in more precise and prompt diag, leading to improved patient outcomes and decreased healthcare costs. By leveraging the power of ML, healthcare providers can potentially revolutionize the way medical data is analyzed and utilized to improve patient care.

To summarize, the study's findings indicated that the RF algorithm was the most successful approach in predicting the severity of heart diseases. With an accuracy of eighty six point one four percentage, this algorithm outperformed both SVM and LR methods. These results demonstrate the potential benefits of utilizing ML algorithms in healthcare settings to improve the accuracy and timeliness of diag. Overall, the study's findings offer promising insights into the effectiveness of ML in improving patient outcomes and reducing healthcare costs. The precision, recall, and F1-score metrics further reinforced the RF algorithm's effectiveness in predicting the severity of heart diseases. This algorithm's ability to mitigate overfitting while simultaneously improving accuracy is a particularly valuable feature in healthcare settings, where accurate diag are critical. The RF algorithm's success in this study suggests that it could serve as a powerful tool in aiding healthcare professionals in diagnosing heart diseases and potentially improving patient outcomes.

V. CONCLUSION

The results of this research demonstrate the potential of ML algorithms in accurately predicting the severity of heart diseases. With high accuracy rates achieved by the RF algorithm, this study highlights the valuable role that ML can play in healthcare settings. These findings suggest that by leveraging the power of ML algorithms, healthcare professionals may be better equipped to diag heart diseases accurately and in a timely manner, potentially leading to improved patient outcomes. Overall, the study's findings underscore the importance of incorporating ML approaches in healthcare settings to improve the accuracy and efficiency of diag.

The study's findings revealed that the RF algorithm outperformed the other algorithms tested, achieving an accuracy rate of eighty six point one four percentage. Although SVM and LR also demonstrated high accuracy rates, the RF algorithm proved to be the most effective in predicting the severity of heart diseases. By incorporating ML algorithms such as these into clinical practice, healthcare professionals may be better equipped to diag and treat heart diseases accurately and efficiently, potentially leading to improved patient outcomes. Overall, the study highlights the potential of ML in the field of cardiology, offering promising avenues for improving healthcare delivery and ultimately benefitting patientsThe vast amount of medical data generated in the healthcare industry today can be difficult for healthcare professionals to analyze effectively. ML algorithms offer a solution to this challenge, as they have the ability to sift through large volumes of data and identify patterns and relationships that may not be immediately apparent to humans. By analyzing this data, ML algorithms can provide valuable insights into a variety of medical conditions, including those related to cardiology, oncology, and neurology, among others. This can lead to more accurate diags, better treatment plans, and improved patient outcomes, ultimately benefiting both healthcare providers and patients. ML algorithms have the potential to transform the way healthcare professionals diag and treat heart diseases. By analyzing large amounts of patient data, these algorithms can help identify patterns and relationships that can be used to predict the severity of heart diseases with high accuracy. This can enable healthcare professionals to develop personalized treatment plans tailored to individual patients' needs, resulting in more accurate and timely diag and better patient outcomes. With a better understanding of a patient's condition, healthcare professionals can develop effective strategies for managing the disease, preventing complications, and improving quality of life. Overall, the use of ML algorithms in predicting the severity of heart diseases has the potential to significantly improve the standard of care in the field of cardiology. As ML continues to evolve, future research can explore the use of more advanced algorithms and larger datasets to further improve the accuracy of predicting heart disease severity. Additional data sources, such as wearable devices and electronic health records, can provide more comprehensive medical data for analysis, leading to more personalized and effective treatment plans. Moreover, incorporating genetic data can help identify patients at high risk of developing heart diseases and enable healthcare professionals to implement preventive measures. Overall, continued research in this field has the potential to transform the way we approach heart disease diagnosis and treatment. Furthermore, research can investigate the potential barriers and challenges to the adoption of ML in clinical practice, such as concerns about data privacy and security, regulatory compliance, and ethical considerations. Addressing these challenges can help facilitate the widespread implementation of ML algorithms in healthcare, ultimately leading to improved patient outcomes and more efficient and effective healthcare delivery. This can help healthcare professionals save time and improve the quality of care provided to patients.

Using ML algorithms to predict the severity of heart diseases can provide healthcare professionals with valuable insights for developing personalized treatment plans and managing patients' conditions. The high accuracy achieved by the RF algorithm suggests its potential for use in clinical settings to improve patient outcomes. Further research can explore the use of more advanced algorithms and data sources, including genetic data and wearable devices, to further improve prediction accuracy. The practical implementation of these algorithms in electronic health record systems can also streamline routine tasks and facilitate patient monitoring.

These findings can help healthcare professionals diag and treat heart diseases more effectively, leading to improved patient outcomes. Future research can explore the useof more advanced algorithms and data sources and the practicalimplementation of ML in clinical settings.

REFERENCES

[1]. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., & Schmid, J. (1989). Calibration of a computer-assisted diagnosis scheme for the detection of coronary artery disease.Computers and Biomedical Research, 22(5), 337-345.

[2]. UCI ML Repository. (n.d.). Heart Disease Data Set. Retrieved March 21, 2023, from https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[3]. Zhang, J., Ma, X., & Liu, Y. (2021). Prediction of HeartDisease Severity Using ML Techniques.Frontiers in Cardiovascular Medicine, 8, 664675. doi:10.3389/fcvm.2021.

[4]. A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu,and H. Lee, "An integrated ML approach to stroke prediction," in Proc. 16th ACM SIGKDD Int. Conf.Knowl. Discov. Data Mining, 2010, pp. 183–192.

[5]. M. Monteiro et al., "Using ML to improve the prediction of functional outcome in ischemic stroke patients," IEEE/ACM Trans. Comput. Biol. Bioinf.,vol. 15, no. 6, pp. 1953–1959, Nov./Dec. 2018.

[6]. S. F. Sung, C. Y. Lin, and Y. H. Hu, "EMR-basedphenotyping of ischemic stroke using supervised machine learning and text mining techniques," IEEE J. Biomed. HealthInform., vol. 24, no. 10, pp. 2922–2931, Oct. 2020.

[7]. G. Lim et al., "Feature isolation for hypothesis testing in retinal imaging: An ischemic stroke prediction case study,"in Proc. AAAI Conf. Artif. Intell., 2019, vol. 33, pp. 9510– 9515.

[8]. T. Liu,W. Fan, and C.Wu, "A hybrid MLapproach to cerebral stroke prediction based on imbalanced medical dataset," Artif. Intell. Med., vol. 101, 2019, Art. no. 101723.

[9]. F. Wang, L. P. Casalino, and D. Khullar, "Deep learning in medicine— promise, progress, and challenges,"JAMA Intern. Med., vol. 179, no. 3, pp. 293–294, 2019.

[10]. C. Sun, A. Shrivastava, S. Singh, and A. Gupta,"Revisiting unreasonable effectiveness of data in deep learning era," inProc. IEEE Int. Conf. Comput. Vis., 2017, pp.843–852.

[11]. A. O'Brien, C. Rajkumar, and C. J. Bulpitt, "Bloodpressure lowering for the primary and secondary preventionof stroke: Treatment of hypertension reduces the risk of stroke," J. Cardiovasc. Risk, vol. 6, no. 4, pp. 203–205, 1999.

[12]. J. E. Manson et al., "A prospective study of maturity-onset diabetes mellitus and risk of coronary heart disease and stroke in women," Arch. Intern. Med., vol. 151, no. 6, pp. 1141–1147, 1991.

[13]. S. Cheon, J. Kim, and J. Lim, "The use of deep learning to predict stroke patient mortality," Int. J. Environ.Res. Public Health, vol. 16, no. 11, 2019, Art. no. 1876.

[14]. D. R. Pereira, P. P. R. Filho, G. H. de Rosa, J. P. Papa, and V. H. C. de Albuquerque, "Stroke lesion detection using convolutional neural networks," in Proc. Int. Joint Conf.Neural Netw., 2018, pp. 1–6.

[15]. D. Teoh, "Towards stroke prediction using electronichealth records," BMC Med. Informat. Decis. Mak., vol. 18, no. 1, pp. 1–11, 2018.