HEART STROKE RISK DETECTION USING CONVOLUTIONAL NEURAL NETWORK

Komireddy Manikanta[1], Dr. G.N.R. PRASAD[2]

[1]Student, CBIT.

[2]Sr Asst Professor, CBIT.

ABSTRACT: Stroke, with no proven cure, is one of the leading causes of death and long-term disability worldwide. Deep learning-based methods have the potential to outperform current models in predicting stroke risk, but they require large amounts of well-labelled data. Due to strict privacy policies in the healthcare system, stroke data is usually distributed in small chunks across multiple institutions. The World Health Organization (WHO) claims stroke is the leading cause of death and disability worldwide. Various stroke warning indicators can be detected early, thus reducing the severity of stroke. Multiple machine learning (ML) and deep learning (DL) models have been implemented to predict stroke likelihood. Multiple physiological markers, machine learning, and deep learning techniques are used in this study, Hybrid Deep Transfer Learning, Super Vector Machines (SVM), Decision Tree (DT) Classification, Random Forest (RF) Classification, CNN+LSTM, Voting Classifiers, and many more to train the above models to make accurate predictions method was used. The algorithm that showed the highest accuracy rate on this task was CNN+LSTM. The open-access stroke prediction dataset was the dataset used to develop this method. The accuracy percentages of the models used in this study were significantly higher than in previous studies, indicating an improved level of confidence in these models. Its robustness has been demonstrated by comparing multiple models and the scheme can be derived from research analyses.

Keywords *–stroke risk detection, convolutional neural network, long-short time memory, machinelearning, voting classifier.*

1. INTRODUCTION

Stroke is one of the most common diseases that can cause death and long-term disability in older people worldwide. A recent study found that an estimated 795,000 people in the United States experience a new or recurrent stroke each year, with a stroke occurring every 40 seconds of her life[1]. One of her five stroke patients died within her year [2]. The cost of care and rehabilitation for survivors places a significant burden on their families and the health system, with direct and indirect costs due to stroke cases costing her $45.5 billion between 2014 and 2015. Estimated in USD[3]. Accurate stroke prediction is therefore of great importance to delay the onset of stroke and reduce the cost of early treatment to minimize its risk. A stroke risk prediction (SRP) model has been developed using medical data from multiple studies, including electronic health records and retinal scans. Deep learning-based approaches [6]-[11] and traditional machine learning methods [4], [5] such as support vector machines (SVM), decision trees, and logistic regression can be broadly categorized. Deep neural networks (DNN) have been shown to be excellent at predicting stroke [8]. However, a well-known drawback of such models is their reliance on the availability of large amounts of accurately classified data. The required amount of reliable information may not be readily available in real life [12]. Strict data protection regulations in the health system make it difficult to exchange stroke data between medical institutions. As a result, different institutions often share individual pieces of the overall collection of stroke data. The frequency of positive and negative strokes also varies greatly. As a result, the DNN-based real-world implementation of his SRP model may not work [13].

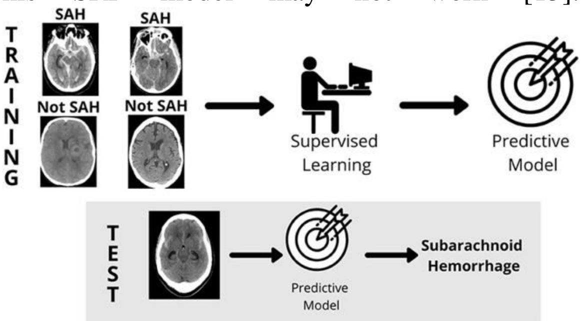

Fig.1: Example figure

Despite the paucity of data on stroke, clinical studies have shown strong associations between stroke risk and several common chronic conditions, such as diabetes and hypertension. 14], [15]. Long-term relationships and patterns found in continuous data were very well captured by the CNN+LSTM model. As stroke risk variables change over time, the model's ability to discern these long-term patterns contributed significantly to more accurate risk prediction. A CNN+LSTM model-based heart rate risk prediction experiment showed promising results and provided insightful information for the future development of medical AI applications. In the medical field, the ability of models to combine imaging and time-series data for accurate risk assessment will greatly improve stroke prevention and patient care.

2. LITERATURE REVIEW

Heart disease and stroke statistics—2017 update a report from the American Heart Association:

The American Heart Association (AHA) annually compiles the most recent statistics on heart disease, stroke, and the AHA's Life's Simple 7 (Figure 1), which include fundamental health behavior's (smoking, physical activity [PA], diet, and weight) and health factors (cholesterol, blood pressure). This is done in collaboration with the Centre's for Disease Control and Prevention, the National Institutes of Health, and other governmental organizations. For the most recent information on these causes and diseases, The Statistical Update is a vital resource for members of the public, politicians, media professionals, physicians, healthcare administrators, researchers, health advocates, and everyone else. In the United States and across the world, cardiovascular disease (CVD) and stroke have a significant negative impact on health and the economy. The Update also contains the most recent information on a number of significant clinical heart and circulatory disease conditions and outcomes, including stroke, congenital heart disease, rhythm problems, subclinical atherosclerosis, coronary heart disease, heart failure (HF), valvular disease, venous disease, and peripheral arterial disease (PAD), as well as the cost, procedure quality, and economic impact of these conditions and outcomes. Since 2006, Statistical Update's annual report has been referenced in the literature more than 20,000 times. In 2015, various statistical updates were discussed more than 4,000 times.

An integrated machine learning approach to stroke prediction

Stroke is the leading cause of serious long-term disability and the third leading cause of death in the United States. Accurately predicting stroke is critical for early intervention and treatment. Compare the Cox proportional hazards model to machine learning techniques for predicting stroke using the Cardiovascular Health Study (CHS) dataset. We focus on typical problems of feature selection, data imputation, and prediction in medical datasets. We provide a new automated method for feature selection based on our recommended heuristic and conservative average. Our proposed feature selection method outperforms the Cox proportional hazards model and the L1-normalized Cox feature selection approach in terms of area under the ROC curve when combined with a support vector machine (SVM). As an alternative to the Cox model, we also provide a margin-based censored regression method that combines the idea of margin-based classifiers with censored regression. Overall, both his AUC and Concordance Index measurements of our technique are superior to the current state-of-the-art. In addition, our study also identified potential risk factors that could not be found by conventional approaches. Our method can be used to predict clinical outcome in a variety of diseases, often with missing data and unknown risk variables.

Using machine learning to improve the prediction of functional outcome in ischemic stroke patients

Ischemic stroke is the leading cause of human disability and death worldwide. A patient's prognosis after stroke is greatly influenced by the therapeutic decisions doctors make during the acute phase. The ASTRAL, DRAGON, and THRIVE scores were all provided as tools to help physicians assess the functional prognosis of patients after stroke within the past five years. These rule-based classifiers use information available at the time a patient is admitted to the emergency department. This study addresses the problem of predicting the functional outcome of ischemic stroke patients 3 months after hospitalization using machine learning algorithms. We found that a pure machine learning method using features specified at admission only slightly exceeded the highest scores ($0.771 - $0.056) for area under the ROC curve (AUC) ($0.808 - $0.085 PM). indicate. However, we found that the AUC can be significantly increased beyond 0.90 by incrementally adding attributes that become available at a later point in time. We conclude that this result supports the use of admissions assessment, but in fact also highlights the need to use other variables and, where feasible, more sophisticated methods. is required.

EMR-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques

Ischemic stroke is the leading cause of adult death and disability worldwide. Because ischemic stroke is highly variable in morphology, phenotypic analysis is of great importance for medical research and clinical prognosis. However, this task is difficult when the study population is large. Most human annotation of medical data has been used in previous studies on ischemic stroke phenotyping. This study explores multiple approaches for automating ischemic stroke phenotypes into his four subtypes of the Oxfordshire Regional Stroke Project Classification (EMR) based on structured and unstructured electronic medical record data. Considered. A total of 4,640 adult patients admitted to a teaching hospital for acute ischemic stroke participated in this study. In addition to the structured components of the National Institutes of Health Stroke Scale, unstructured clinical narratives were pre-processed with Meta Map to discover medical concepts and encoded into feature vectors. A number of supervised machine learning techniques were used to generate the classifiers. The results of this study suggest that combining textual and structured data from EMR may be useful for phenotypic analysis of ischemic stroke. Furthermore, splitting this multiclass problem into binary classification jobs and aggregating the classification results improves performance.

Feature isolation for hypothesis testing in retinal imaging: An ischemic stroke prediction case study

Ischemic stroke is a significant and unpredictable cause of death and disability. Retinal fundus photography is non-invasive and is recommended for stroke risk assessment because of the similarities between retinal and cerebral microcirculation. Previous studies have shown an association between vein caliber and stroke risk. However, other retinal features may be more appropriate. This article presents an extensive deep learning investigation on six retina datasets. The effectiveness of vessel diameter and vessel shape alone in stroke classification was assessed using segmented vessel tree images for feature separation and generalizability of the model to unidentified sources using dataset ablation. is inspected. This result indicates that source-specific features can affect model performance, and that vessel width and shape can be markers of ischemic stroke.

3.METHODOLOGY

When it comes to stroke prediction, hybrid deep transfer learning frameworks may perform poorly. The fact that such models rely on the availability of well-characterized large-scale data is a well-known drawback. In real-world scenarios, you may not have easy access to the required amount of reliable data. Cooperation with medical professionals and adherence to privacy and ethical standards are essential throughout the development process. Strict privacy policies in the healthcare system usually make it difficult to share stroke data between hospitals. As a result, small subsets of the overall stroke data collection are often spread across multiple institutions. In addition, extreme positive and negative imbalance cases can occur in stroke data. As a result, hybrid deep transfer learning frameworks may not perform well when used in the real world.

Cons:

1. Strict data protection guidelines in the healthcare system can make it difficult to share stroke data between institutions.

2. Hybrid deep transfer learning systems may perform poorly in the real world.

We provide a stroke risk prediction model based on CNN+LSTM and voting classifiers to leverage knowledge structures from multiple linked sources (external stroke data, data on chronic diseases such as hypertension and diabetes). Masu. The proposed model has undergone rigorous testing in both simulated and real settings and outperforms state-of-the-art stroke risk prediction models. Furthermore, in real hospitals he shows the potential for widespread deployment of 5G/B5G infrastructure.

Pros:

1. The proposed model offers higher accuracy and improved performance.

2. CNNs are excellent feature extraction tools that can automatically select relevant features from the raw form of the input data.

3. By combining analysis of medical images (such as heart scans) with time-aware data, CNN+LSTM can provide a more comprehensive picture of heart attack and stroke risk factors.

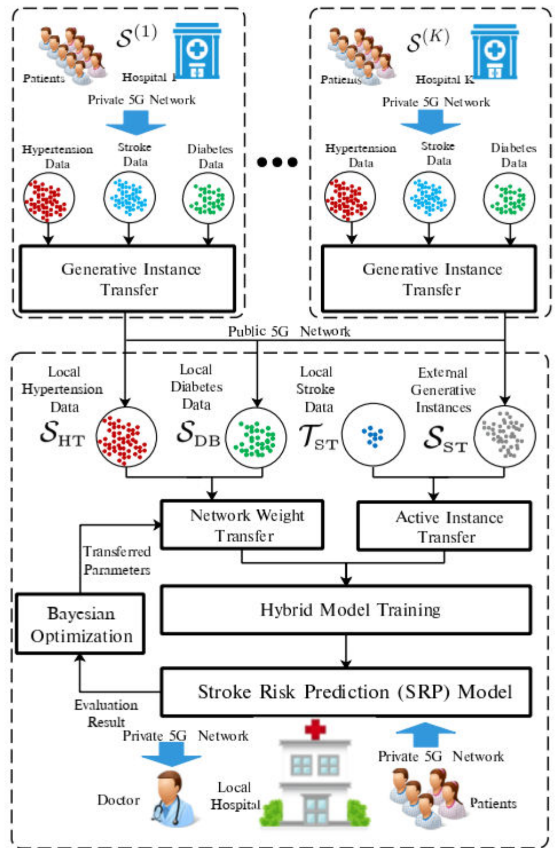

Fig.2: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Enter data: Using this module we will import data into the system.
- The treatment: Using this module, we will read the data to be processed.
- The data will be divided into training and testing using this module.
- Model building: CNN + LSTM + DT + RF + DT - Deep hybrid TL - DNN - CNN + LSTM - SVM - DT - RF - Voting classifier. accuracy of the calculation method.
- User registration and login are required to use this module.

- This module can be used to provide data for predictions.
- Prognosis: The final prognosis is displayed

4. IMPLEMENTATION

ALGORITHMS:

Hybrid Deep TL: Using Hybrid Deep Transfer Learning (HDTL), the information structure of multiple source domains scattered in many hospitals is transferred to the target domain of the stroke. The HDTL-SRP architecture has been developed that works in a distributed manner without the need for organizations to share patient data directly. It is divided into three parts: (1) Creative Versioning (GIT), synthetic versioning to train models using GANs in external data; (2) Network Weight Transfer (NWT), using data from highly correlated diseases (such as hypertension or diabetes); (3) Bayesian Optimization (BO), determining the best passed parameters; and (4) Active Versioning (AIT), which selects more deeply aggregated trait versions to create a balanced trait dataset, which is then used to develop the model.

DNN: Large neural networks with complex input-output transformations are trained using DNN:

Study carefully. Mapping an image to the name of the person or people in the image, the same way they do on social media, is a modern use of DL. One creative use of DL is to describe an image in one word.

CNN + LSTM: The CNN-LSTM model consists of the LSTM layer to predict the sequence and the CNN layer to extract features from the input data. CNN-LSTM is commonly used for activity recognition, image tagging, and video tagging.

SVM: Both regression and classification problems can be solved using SVM, a supervised machine learning method. Although we call them regression problems, classification is the more appropriate term. The goal of the SVM method is to find a hyperplane in N-dimensional space that categorizes the entry points unambiguously.

DT: Decision tree (DT) is a non-parametric supervised learning method and can be applied to classification and regression problems. The root node, branches, inner nodes, and leaf nodes form its hierarchical tree structure.

RF:The random forest (RF) method is a popular supervised machine learning method for classification and regression problems. We know that a forest is made up of many types of trees, and the more active a forest is, the more trees there are.

Voting Classifier: Kagglers regularly use Voting Classifier, a machine learning technique, to improve their model performance and rank up. The voting classifier has major limitations, but it can be used to improve performance on real-world datasets.

6. CONCLUSION

This study compared the effectiveness of deep learning and machine learning methods in accurately predicting stroke risk. Research to predict stroke risk based on the CNN+LSTM model has proven to be extremely promising and promising in the medical field. The main objective of this study was to reliably predict stroke risk variables over time by combining the advantages of long-term short-term memory (LSTM) and convolutional neural networks (LSTMs). CNN). The project's studies and results highlight a number of important key factors. First, the CNN+LSTM model effectively combined and evaluated patient monitoring time series data with medical imaging data such as MRI or CT scans. The accuracy of risk predictions has been improved by this synergistic approach, providing a deeper understanding of the many components that influence stroke risk. The ability of the CNN+LSTM model to automatically extract features is a major advantage. Although the LSTM component successfully identified important temporal patterns from the time series data, the CNN component skillfully learned the relevant spatial features from the medical images. The model development process has been simplified and made more efficient by eliminating the need for manual feature engineering. The CNN+LSTM model also works very well for identifying long-term relationships and patterns seen in sequence data. The model's ability to identify these long-term patterns has significantly contributed to more accurate risk predictions since stroke risk variables change over time.

In conclusion, the study of stroke risk prediction based on the CNN+LSTM model has shown encouraging results and provides useful insights for the future development of medical AI applications. . In the medical field, the model's ability to combine time series data and imaging for accurate risk estimates represents a significant improvement in stroke prevention and patient care.

7. FUTURE SCOPE

Predicting stroke risk has a bright future and offers many innovative opportunities in the medical field. We can predict increasingly accurate and personalized risk assessment models as AI, data analytics, and health technology continue to evolve. Combining modern data sources, such as genetic data and wearable technology, will provide comprehensive real-time information to improve predictions. Individualized prevention and intervention tactics will be enabled through longitudinal data analysis and risk assessment. Creating interpretable AI models will improve the understanding and transparency of risk factors. Remote monitoring and quick action will be enabled by the real-time risk monitoring system and mobile health app. These predictive models will be used more widely through collaborative data sharing and addressing ethical issues, which will improve patient care and stroke prevention.

REFERENCES

[1] E. J. Benjamin, M. J. Blaha, and S. E. Chiuve, “Heartdisease and stroke statistics—2017 update a report from the American Heart Association,” Circulation, vol. 135, no. 10, pp. e146—e603, 2017.

[2] S. Koton et al., “Stroke incidence and mortality trends in US communities, 1987 to 2011,” JAMA, vol. 312, no. 3, pp. 259–268, 2014.

[3] E. J. Benjamin, P.Muntner, andM. S. Bittencourt, “Heart disease and stroke statistics—2019 update: A report from the American Heart Association,” Circulation, vol. 139, no. 10, pp. e56–e528, 2019.

[4] A.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. Discov. Data Mining, 2010, pp. 183–192.

[5] M. Monteiro et al., “Using machine learning 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-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques,” IEEE J. Biomed. Health Inform., 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] 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.

[9] 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.

[10] D. Teoh, “Towards stroke prediction using electronic health records,” BMC Med. Informat. Decis. Mak., vol. 18, no. 1, pp. 1–11, 2018.

[11] T. Liu,W. Fan, and C.Wu, “A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset,” Artif. Intell. Med., vol. 101, 2019, Art. no. 101723.

[12] 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.

[13] 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.

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

[15] 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.