Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 6s (2024)

An Approach to Develop Cerebra-Vascular-Haemato-Cardiac Detector using Machine Learning Techniques

Shambhu Nath Saha[1], Soumya Bhattacharyya[2], Mayank Bala[3], Tirthankar Ray[4], Sambit Chakraborty[5], Parthib Kumar Deb[6], Pritam Kapat[7]

[1,2,3,4,5,6,7]Narula Institute of Technology, Kolkata

Article History:

Received: *28-05-2024*

Revised: *30-06-2024*

Accepted: *19-07-2024*

Abstract:
The "Design of Cerebra-Vascular-Haemato-Cardiac Detector using Machine Learning Approaches" project is a groundbreaking initiative that seeks to revolutionize medical diagnostics by integrating advanced machine learning techniques. This multifaceted project encompasses the development of a comprehensive detector capable of analysing cerebrovascular, haematological, and cardiac data to provide accurate and timely health assessments. The integration of diverse data modalities and the creation of a real-time detection system will contribute to early disease detection and improved patient outcomes. [1][2][3]

Cerebrovascular stroke and heart attack stand as prominent causes of disability and global mortality, underscoring the critical need for early detection to facilitate effective treatment. Traditional diagnostic systems for Electroencephalograph (EEG) and Electrocardiograph (ECG) are non-wearable due to their unwieldy instruments, posing a significant challenge. In response, we propose a Cerebra-Vascular-Haemato-Cardiac (CVHC) detector designed to detect strokes and cardiac arrests. The key elements of the CVHC detector include EEG and ECG, with additional integration of an oxygen sensor and blood pressure sensor. This comprehensive system aims to identify anomalies in the human body associated with strokes, cardiac arrests, respiratory issues, and heat strokes. Leveraging Machine Learning, the CVHC device provides immediate emergency alerts.

The signals originating from ECG, EEG, blood pressure sensor, and oxygen sensor undergo processing through a simulator, Discrete Wavelet Transformation (DWT), and Convolutional Neural Network (CNN). This processing involves convolution, noise reduction, and wavelet differentiation of EEG and ECG signals, ultimately amplifying them to attain usable voltage. The proposed system is adept at detecting Ischemic Stroke, Haemorrhagic Stroke, and Cardiopulmonary Arrest by comparing the acquired signals with predefined patterns of normal brain and heart activities. This innovative CVHC solution thus addresses the limitations of traditional diagnostic instruments, offering a wearable and efficient alternative for early detection of life-threatening conditions.

Keywords: Cerebra-Vascular-Haemato-Cardiac, CNN, EEG, ECG, Machine Learning

1. Introduction

The rapid advancements in machine learning and artificial intelligence have opened new horizons in the field of medical diagnostics. One critical area that stands to benefit immensely from these technologies is the early detection and diagnosis of complex, multi-system conditions involving cerebrovascular, hematological, and cardiac systems. These conditions, often interlinked, present significant challenges in clinical settings due to their overlapping symptoms and the intricate interplay between the brain, blood, and heart.

Cardiovascular diseases (CVDs) remain the leading cause of mortality globally, while cerebrovascular diseases, such as strokes, contribute significantly to long-term disability and death. Additionally, hematological disorders, including anemia and coagulopathies, frequently exacerbate the complications associated with CVDs and cerebrovascular conditions. The interplay between these systems underscores the necessity for a holistic diagnostic approach that can simultaneously assess and monitor the health of the cerebrovascular, hematological, and cardiac systems.

Traditional diagnostic methods often rely on isolated tests and examinations, which may not capture the comprehensive health status of a patient. This segmented approach can lead to delayed diagnoses and suboptimal treatment outcomes. Therefore, integrating machine learning techniques to develop a unified detector that can analyze and interpret data from all three systems offers a promising solution to these challenges.

In this paper, we propose an innovative approach to develop a Cerebra-Vascular-Haemato-Cardiac (CVHC) detector using advanced machine learning techniques. Our objective is to create a robust and reliable diagnostic tool capable of early detection and risk stratification of conditions involving these critical systems. By leveraging the power of machine learning algorithms, we aim to enhance diagnostic accuracy, improve patient outcomes, and reduce the burden on healthcare systems.

We will discuss the methodologies employed in developing the CVHC detector, including data collection, feature extraction, model training, and validation. Furthermore, we will explore the potential applications of this detector in clinical practice and its implications for personalized medicine. Our approach underscores the importance of interdisciplinary collaboration and the integration of cutting-edge technology in advancing healthcare diagnostics.

2. Objectives

1. **Dataset Creation and Expansion:**

- Curate an expansive and diverse dataset from various medical sources, including hospitals, research institutions, and public health databases.
- Incorporate longitudinal data to capture the dynamic nature of health conditions.

2. **Advanced Machine Learning Model Development:**

- Explore and implement cutting-edge machine learning algorithms, including deep learning architectures such as transformers, attention mechanisms, and generative adversarial networks.
- Investigate transfer learning techniques to leverage pre-trained models for improved performance.

3. **Multimodal Data Fusion:**

- Develop innovative strategies for fusing and harmonizing multimodal data, including medical imaging, genetic information, electronic health records (EHR), and wearable device data.

4. **Explainability and Interpretability:**

- Implement methods to enhance model explainability, ensuring transparency and trust in the diagnostic outcomes.

- Investigate interpretable machine learning techniques for providing insights into the decision-making process.

5. **Real-time Detection System Enhancement:**

- Enhance the real-time detection system by optimizing computational efficiency and scalability through parallel processing and cloud-based solutions.
- Implement adaptive learning mechanisms to continuously update the models with emerging medical knowledge.

6. **Collaboration with Healthcare Professionals:**

- Collaborate with healthcare professionals to understand their needs, incorporate domain knowledge into the models, and validate the system's performance in real-world clinical settings.

3. Methods

1. **Data Collection and Preprocessing:**

- Collaborate with medical institutions to access diverse datasets, ensuring compliance with ethical standards and data privacy regulations.
- Implement advanced preprocessing techniques to handle missing data, outliers, and ensure data quality.

2. **Model Development and Optimization:**

- Utilize a wide range of machine learning frameworks, including TensorFlow and PyTorch, for model development.
- Optimize hyperparameters, conduct cross-validation, and implement ensemble learning to enhance model robustness.

3. **Interdisciplinary Research:**

- Collaborate with experts in neurology, cardiology, hematology, and data science to develop a holistic understanding of the diseases and conditions targeted by the detector.

4. **Ethical Considerations:**

- Develop ethical guidelines for the responsible use of AI in healthcare, addressing issues such as bias, fairness, and the impact on marginalized populations.

4. Deliverables

1. **Comprehensive and Diverse Dataset:**

- A curated dataset containing cerebrovascular, hematological, and cardiac health data with appropriate documentation.

2. **State-of-the-Art Machine Learning Models:**

- Trained models demonstrating high accuracy, sensitivity, and specificity across multiple health domains.

3. **Advanced Real-time Detection System:**

- A scalable and adaptive real-time detection system capable of handling diverse data streams.

4. **User Interface and Integration:**

- An intuitive and user-friendly interface integrated with existing healthcare systems for seamless adoption.

5. **Documentation and Guidelines:**

- Comprehensive documentation, including user manuals, technical specifications, and ethical guidelines for the use of the system.

5. Literature Survey

1. In contemporary times, heart disease has emerged as a paramount health concern, often leading to fatal outcomes for a significant number of patients. The diagnostic process for heart disease poses considerable challenges due to its complexity, necessitating both precision and efficiency. Early detection of heart disease plays a pivotal role in reducing the likelihood of fatalities associated with cardiac issues. Given the increasing prevalence of cardiac ailments, predicting heart disease has evolved into one of the most formidable tasks in recent medical practice. Researchers have explored various closely linked attributes to ascertain the most dependable predictors for these conditions. This study employs Machine Learning (ML) techniques to discern the presence of cardiac abnormalities. The proposed methodology aims to forecast the likelihood of heart disease and categorize patients' risk levels by employing diverse ML algorithms such as Decision Tree (DT), Ada-Boost Classifier (AB), Extra Trees Classifier (ET), Support Vector Machine (SVM), Gradient Boost, MLP, Extreme Gradient Boost (XGB), Random Forest (RF), KNN, and LR. The system is trained and tested using a amalgamation of three distinct datasets. The experimental outcomes reveal that, in comparison to other ML algorithms, the Decision Tree algorithm demonstrates the highest accuracy, reaching 99.16%. [4]

2. The societal repercussions of stroke have prompted coordinated endeavors to enhance stroke management and diagnosis. With an increasingly harmonized integration of technology and medical practices, caregivers are creating opportunities for improved patient care by systematically scrutinizing and cataloging patients' medical histories. Consequently, it becomes imperative to investigate the interdependencies among various risk factors within patients' health records and comprehend their respective contributions to predicting stroke. This paper systematically examines diverse factors within electronic health records to enhance the efficacy of stroke prediction. Employing statistical techniques and principal component analysis, we pinpoint the pivotal factors for accurate stroke prediction. Our findings indicate that age, heart disease, average glucose level, and hypertension emerge as the most critical factors in identifying stroke in patients. Furthermore, a perceptron neural network, utilizing these four attributes, achieves the highest accuracy and the lowest miss rate when compared to employing all available input features and other benchmarking algorithms. Given the highly imbalanced nature of the dataset concerning stroke occurrences, we present our results based on a balanced dataset created through sub-sampling techniques.

This project provides an exhaustive examination of patient attributes within electronic health records for the purpose of stroke prediction. We systematically scrutinized various features, conducting feature correlation analysis and step-wise analysis to identify an optimal set of features. Our findings revealed that the features exhibit low correlation, and a combination of only four features (A, HD, HT, and AG) significantly contributes to stroke prediction. Additionally, we conducted principal component analysis, indicating that nearly all principal components are necessary to explain higher variance. Variable loadings highlighted that the first principal component, with the highest variance, might elucidate the underlying phenomenon of stroke prediction.

Subsequently, three machine learning algorithms were applied to different feature and principal component configurations. The neural network demonstrated superior performance with the feature combination of A, HD, HT, and AG, achieving an accuracy of 78% and a miss rate of 19%. While the results are promising with this limited feature set, the accuracy of the perceptron model may not improve further due to a lack of additional discriminatory features and a limited dataset.

Observations revealed high correlation among existing features in the electronic health record dataset, offering minimal additional information. A larger dataset would facilitate more efficient training of deep neural networks. Future work involves collecting institutional data to enhance the model's performance. The systematic analysis of electronic health record features is anticipated to aid clinicians in the efficient archival of records. The integration of the electronic records dataset with background knowledge on diseases and drugs, using Semantic Web technologies, is planned. Knowledge graph technologies can contribute to publishing electronic health records in an interoperable manner to the research community. The incorporation of background knowledge from other datasets has the potential to enhance the accuracy of stroke prediction models. [5][6]

3. Stroke ranks as the third leading cause of global mortality, posing a significant health threat attributed to the disruption of blood flow to the brain, resulting in severe illness, disability, or fatality. Timely and precise stroke prediction is crucial for early-stage intervention and reducing mortality rates. This study advocates for a machine learning approach to enhance the accuracy of stroke diagnosis, particularly when dealing with imbalanced data. To address data imbalances, the Random Over Sampling (ROS) technique is employed.

A comprehensive analysis is conducted on eleven classifiers, including Support Vector Machine, Random Forest, K-nearest Neighbor, Decision Tree, Naïve Bayes, Voting Classifier, AdaBoost, Gradient Boosting, Multi-Layer Perception, and Nearest Centroid. Prior to data balancing, ten classifiers exhibit results surpassing 90% accuracy, while post-balancing with oversampling, four classifiers demonstrate accuracy levels exceeding 96%. Each model undergoes Hyperparameter tuning and cross-validation to optimize results. Evaluation metrics such as Accuracy, F1-Measure, Precision, and Recall are employed to gauge the performance of the machine learning models.

The results highlight the Support Vector Machine as the most accurate, achieving 99.99% accuracy, recall, precision, and F1-measure. Following closely is the Random Forest, attaining 99.87% accuracy with a minimal 0.001% error. Additionally, to enhance user accessibility, a user-friendly web app and mobile app are developed based on the most accurate model. [7]

4. This work introduces a prototype for stroke classification that integrates text mining tools and machine learning algorithms. Machine learning plays a crucial role as a robust tracker in various domains, including surveillance, medicine, and data management, utilizing appropriately trained

machine learning algorithms. Data mining techniques, employed in this study, offer a comprehensive overview of information tracking from both semantic and syntactic perspectives.

The proposed approach involves extracting patients' symptoms from case sheets and training the system with the acquired data. During the data collection phase, case sheets from 507 patients at Sugam Multispecialty Hospital in Kumbakonam, Tamil Nadu, India, were gathered. Subsequently, the case sheets underwent mining using tagging and maximum entropy methodologies, and a proposed stemmer extracted a set of common and unique attributes for stroke classification.

The processed data were then input into various machine learning algorithms, including artificial neural networks, support vector machines, boosting and bagging, and random forests. Among these algorithms, artificial neural networks trained with a stochastic gradient descent algorithm demonstrated superior performance, achieving a classification accuracy of 95% with a smaller standard deviation of 14.69 compared to other algorithms. [8]

5. Ischemic stroke stands as a prevalent neurological disorder and remains the primary cause of significant, enduring disability worldwide. The identification of features pertinent to stroke prognosis holds immense value for facilitating effective interventions and treatments. In this investigation, we employed an integrated machine learning approach to discern prognosis factors related to stroke using The International Stroke Trial (IST) dataset. Addressing common challenges in feature selection and prediction within medical datasets, we initially ranked the importance of features utilizing the Shapiro-Wilk algorithm and analyzed Pearson correlations between features.

Subsequently, we employed Recursive Feature Elimination with Cross-Validation (RFECV), incorporating linear Support Vector Classifier (SVC), Random Forest Classifier, Extra Trees Classifier, AdaBoost Classifier, and Multinomial Naïve Bayes Classifier as estimators, to robustly select features. Further, the significance of the selected features was ascertained through Random Forest Classifier and the Shapiro-Wilk algorithm. Ultimately, a set of twenty-three chosen features was utilized by SVC, Multilayer Perceptron (MLP), Random Forest, and AdaBoost Classifier to predict RVISINF (Infarct visible on CT) in acute stroke cases within the IST dataset.

The findings suggest that the selected features exhibit high accuracy in inferring the long-term prognosis of acute stroke. Additionally, these features prove instrumental in extracting factors associated with RVISINF, particularly in cases linked to large artery occlusion (LAO) in ischemic stroke patients. [9]

6. A novel and promising approach to the clinical management of acute stroke is presented in this work. Accurate diagnosis and real-time prediction models have been developed from hemodynamic data through the application of machine learning techniques. Stroke subtypes can be diagnosed within 30 minutes of monitoring, exitus can be predicted during the first 3 hours of monitoring, and stroke recurrence can be predicted within just 15 minutes of monitoring. These positive results benefit patients with limited access to a CT scan and all individuals arriving at the stroke unit of a specialized hospital. The outcomes from the real-time developed models include stroke diagnosis precision of around 98% (97.8% sensitivity, 99.5% specificity), exitus prediction precision of 99.8% (99.8% sensitivity, 99.9% specificity), and stroke recurrence prediction precision of 98% (98% sensitivity, 99% specificity). [10]

7. Stroke is a medical condition characterized by the rupture of blood arteries in the brain, leading to damage. Interruption of blood and nutrient supply to the brain can result in the manifestation of symptoms. Globally recognized as the primary cause of death and disability, early identification of stroke warning signs is crucial in mitigating its severity. Various machine learning (ML) models have been developed to predict the likelihood of stroke occurrence. In this research, a diverse set of physiological parameters and machine learning algorithms, including Logistic Regression (LR), Decision Tree (DT) Classification, Random Forest (RF) Classification, and Voting Classifier, were utilized to train four distinct models for reliable prediction. Notably, Random Forest emerged as the most effective algorithm, achieving an accuracy of approximately 96 percent.

The open-access Stroke Prediction dataset was employed in developing this method. The accuracy percentages of the models employed in this study surpass those reported in previous research, underscoring the enhanced reliability of the models in this investigation. Rigorous model comparisons have substantiated their robustness, and the details of the approach are elucidated in the study's analysis. [11]

8. The advent of the latest technologies has facilitated the utilization of noninvasive techniques to support healthcare systems. Among the four major cardiovascular diseases, stroke is particularly perilous and life-threatening; however, early detection can significantly enhance a patient's chances of survival. Existing literature indicates that patients often undergo ministrokes, also known as transient ischemic attacks (TIA), before experiencing a full-fledged stroke. Current research predominantly relies on costly approaches such as MRI and CT scan images for classifying cardiovascular diseases, including stroke, making early diagnosis an expensive endeavor. With a rising incidence of strokes in India, there is a pressing need to explore cost-effective noninvasive methods for early stroke diagnosis. Consequently, this concern has motivated the study presented in this paper, proposing a noninvasive approach for early stroke diagnosis.

Cascaded prediction algorithms, while effective, are time-consuming and unable to work directly on raw data without leveraging the properties of EEG. Therefore, the primary objective of this paper is to devise mechanisms for forecasting strokes based on processed EEG data. The proposed time series-based approaches, including LSTM, biLSTM, GRU, and FFNN, are tailored for handling time series-based predictions to inform critical decisions. Experimental research outcomes reveal the commendable performance of all algorithms in the early stroke detection prediction task. Notably, GRU stands out with a remarkable 95.6% accuracy, followed by biLSTM with 91%, LSTM with 87%, and FFNN with 83% accuracy. The experimental findings demonstrate the efficacy of measuring brain waves to predict stroke signs, potentially aiding physicians in the early detection of strokes to save patients' lives. [12]

9. Heart disease (HD) currently stands as the leading global cause of mortality. Recognized as the second-most crucial organ after the brain, early diagnosis plays a pivotal role in improving treatment outcomes and significantly reducing fatality risks in healthcare. This paper introduces a method for heart disease prediction utilizing various machine learning algorithms (MLA), including logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), Naive Bayes (NB), random forest (RF), and decision tree (DT). Evaluation of model accuracy in heart disease prediction was conducted using a testing dataset. Among the six models, both the k-nearest neighbor algorithm

and random forest outperform others, achieving a remarkable accuracy rate of 99.04%. Six feature selection algorithms were employed for performance evaluation, utilizing MCC parameters for accuracy, precision, recall, and F-measure to assess model effectiveness. [13]

10. Most strokes result from an abrupt blockage in the blood vessels supplying the brain and heart. Recognizing the early warning signs of a stroke can mitigate its impact. This paper proposes the early prediction of stroke diseases using diverse machine learning approaches, considering factors such as hypertension, body mass index, heart disease, average glucose level, smoking status, previous stroke, and age. Ten classifiers, including Logistics Regression, Stochastic Gradient Descent, Decision Tree Classifier, AdaBoost Classifier, Gaussian Classifier, Quadratic Discriminant Analysis, Multi-layer Perceptron Classifier, KNeighbors Classifier, Gradient Boosting Classifier, and XGBoost Classifier, were trained using these high-feature attributes.

The study achieved an impressive accuracy of 97%, employing a weighted voting approach to aggregate results from the base classifiers, outperforming them. This model proves to be the most accurate for stroke prediction, with the weighted voting classifier exhibiting a high area under the curve value. Additionally, the false positive rate and false negative rate of the weighted classifier are the lowest compared to others. Consequently, the weighted voting classifier emerges as an almost perfect tool for predicting strokes, offering physicians and patients a valuable resource for prescribing and early detection of potential strokes. [14]

6. Conclusion

The "Design of Cerebra-Vascular-Haemato-Cardiac Detector using Machine Learning Approaches" project is poised to make significant strides in the field of medical diagnostics. This comprehensive proposal outlines the project's ambitious objectives, methodology, deliverables, timeline, and budget. We look forward to the opportunity to discuss and collaborate on this transformative initiative.

It also emphasizes the development of a machine learning model for precise stroke risk prediction using ECG and EEG data. Creation of a user-friendly interface for risk assessment accessible to both healthcare professionals and individuals. Validation of the predictive model's accuracy and reliability through clinical testing. Establishment of a scalable and secure data infrastructure for future research and healthcare applications. Promotion of increased awareness and adoption of the stroke risk prediction tool.

References

[1] Rajpurkar, P., et al. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." arXiv preprint arXiv:1711.05225.

[2] Esteva, A., et al. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature, 542(7639), 115-118.

[3] Cho, K., et al. (2014). "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation." arXiv preprint arXiv:1406.1078.

[4] Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN COMPUT. SCI. 2020;1:345. https://doi.org/10.1007/s42979-020-00365-y.

[5] Johnson, A. E., et al. (2016). "MIMIC-III, a freely accessible critical care database." Scientific Data, 3, 160035.

[6] Dev, S., Wang, H., Nwosu, C.S., Jain, N., Veeravalli, B. and John, D., 2022. A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, *2*, p.100032.

[7] Biswas, N., Uddin, K.M.M., Rikta, S.T. and Dey, S.K., 2022. A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics*, *2*, p.100116.

[8] Govindarajan, P., Soundarapandian, R.K., Gandomi, A.H., Patan, R., Jayaraman, P. and Manikandan, R., 2020. Classification of stroke disease using machine learning algorithms. *Neural Computing and Applications*, *32*, pp.817-828.

[9] Fang, G., Liu, W. and Wang, L., 2020. A machine learning approach to select features important to stroke prognosis. *Computational Biology and Chemistry*, *88*, p.107316.

[10] García-Terriza, L., Risco-Martín, J.L., Roselló, G.R. and Ayala, J.L., 2021. Predictive and diagnosis models of stroke from hemodynamic signal monitoring. *Medical & Biological Engineering & Computing*, *59*(6), pp.1325-1337.

[11] Tazin, T., Alam, M.N., Dola, N.N., Bari, M.S., Bourouis, S. and Monirujjaman Khan, M., 2021. Stroke disease detection and prediction using robust learning approaches. *Journal of healthcare engineering*, *2021*.

[12] Kaur, M., Sakhare, S.R., Wanjale, K. and Akter, F., 2022. Early stroke prediction methods for prevention of strokes. *Behavioural Neurology*, *2022*.

[13] Ansari, G.A., Bhat, S.S., Ansari, M.D., Ahmad, S., Nazeer, J. and Eljialy, A.E.M., 2023. Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction. *Computational and Mathematical Methods in Medicine*, *2023*.

[14] Emon, M.U., Keya, M.S., Meghla, T.I., Rahman, M.M., Al Mamun, M.S. and Kaiser, M.S., 2020, November. Performance analysis of machine learning approaches in stroke prediction. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1464-1469). IEEE.