A Seizure Detection Using Natural Language Processing on Clinical Text Data

Diana Julie D[1], Sri Vaishnavi V[2], Bhavadharini S[3], Harikesh S[4]

[1] Dept. of Artificial Intelligence and Data Science, KIT-KalaignarKarunanidhi Institute of Technology, Coimbatore, India. Email: dianajuliekit@gmail.com

[2] Dept. of Artificial Intelligence and Data Science, KIT-KalaignarKarunanidhi Institute of Technology, Coimbatore, India. Email: kit27.ad53@gmail.com

[3] Dept. of Artificial Intelligence and Data Science, KIT-KalaignarKarunanidhi Institute of Technology, Coimbatore, India. Email: Kit27.ad07@gmail.com

[4] Dept. of Artificial Intelligence and Data Science, KIT-KalaignarKarunanidhi Institute of Technology, Coimbatore, India. Email: kit27.ad21@gmail.com

How to cite this article: Diana Julie D, Sri Vaishnavi V, Bhavadharini S, Harikesh S. A Seizure Detection Using Natural Language Processing on Clinical Text Data. Int J Drug Deliv Technol. 2026;16(37s): 54-59. DOI: 10.25258/ijddt.16.37s.9

Abstract - **Catching seizures early really matters when it comes to diagnosing and treating things like epilepsy. Right now, most doctors look at EEG signals, but that means fancy equipment and an expert who knows how to read them. We're trying something different. Instead of wires and electrodes, we're using Natural Language Processing basically, teaching computers to "read" things like clinical notes, patient stories, and electronic health records for clues. It starts with cleaning up the text breaking sentences into words, getting rid of the fluff, and making everything consistent so our system can actually use it. From there, we run this cleaner data through machine learning and deep learning models designed to spot signs of seizures. These models pick up on the patterns faster and more accurately than older techniques. When we put it to the test, this NLP approach didn't just match what you'd get with traditional EEG-based systems. It made everything easier to scale and much less expensive. In the big picture, plugging this tool into real clinical decision-making means sharper diagnoses and faster, better care for patients.**

Keywords - Seizure Detection, Natural Language Processing, Epilepsy, Machine Learning, Deep Learning, Clinical Text Analysis, Healthcare Analytics

I INTRODUCTION

Seizures occur suddenly in the brain due to an uncontrollable electrical disturbance that alters how somebody acts, moves or is aware of their environment. They are one of the most significant indicators of epilepsy, a long-term neurological disorder that affects millions of individuals around the world. It is very important for clinicians to find seizures early and accurately so that they can provide their patients with good diagnoses, treatment plans, and ultimately help them to improve their quality-of-life. Traditionally, the main method used by clinicians to detect seizures is through measuring the brain's electrical activity via EEG. EEG-based methods can be effective; however, they also require specialized equipment, continuous monitoring, and intensive expertise in interpreting seizures. Thus, patients do not have access to this type of resource in many healthcare systems. In addition, there is often valuable information related to seizure events documented within clinical notes, patient histories, and other medical documents that are typically unstructured text-based formats and often not sufficiently used by clinicians. Recent development of NLP technology has greatly improved our ability to extract value from unstructured text data in new ways. With NLP, it's possible to automatically process clinical narratives in order to find patterns and to identify symptoms and contextual information concerning seizure events. NLP can also be used with machine learning and deep learning models (ML/DL) to convert documents into actionable insights for clinical decision-making concerning seizure events.

II LITERATURE STUDY

The portable closed-loop brain-computer interface (BCI) system for seizure control in individuals with epilepsy. The authors stressed that almost 25% of people with epilepsy do not respond well to current treatments, which shows how important it is to find other ways to help. The suggested framework combines real-time seizure detection with immediate electrical stimulation to create a closed-loop control system that works well. The method was able to find things with a high accuracy rate of 92% to 99% while people were awake and asleep, and it had a low false detection rate of 1.2% to 2.5%. Also, the system responded quickly, with detection and stimulation happening within 0.6 seconds of the start of a seizure. The development of a small, wireless system is an important part of this work. It makes patients more comfortable and mobile by getting rid of the need for wired connections. Experiments with rats that were free to move around showed that the model worked and could monitor and control seizures in real time [1].

This work is in the field of epilepsy treatment Transactions on Neural Systems and Rehabilitation Engineering and explored the potential of Multi-Site Closed-Loop Stimulation (MSCLS) as an effective treatment modality for refractory Epilepsy. The research paper addressed the drawbacks associated with the single-site neurostimulation methods, which were relatively inefficient and had a poor understanding of the mechanisms of their effect on the brain's activities. The authors developed a dedicated closed-loop neurostimulator and tested its functionality through experiments on a rodent model. Specifically, CA1 and CA3 hippocampal areas, sub-thalamic nucleus, and motor cortex were chosen as stimulation sites and the cost-efficient multi-site seizure detection algorithm was developed for real-time operation of the device. Results of the experiments indicated the efficiency of MSCLS as a tool for controlling seizure activity: the number and duration of the seizures were considerably lower in rats without status epilepticus (SE). As for rats with SE, MSCLS proved efficient in seizure severity reduction but did not demonstrate much effect regarding duration and frequency of the attacks [2].

This paper digs into a multimodal way to spot seizures without using EEGs. One of the big headaches in this area is that false alarms crop up way too often, plus it's hard to tell normal movements apart from actual seizure activity for people with Epilepsy. The research uses the Open Seizure Database, which has 94 events everything from big tonic-colonic seizures to more subtle focal ones. Each event is packed with acceleration and heart rate data, all labelled as Normal, Pre-Ictal, or Ictal, broken down into neat time slices. To make sense of this mix of sensor data, the team created AMBER, which stands for Attention-guided multi-branching pipeline with Enhanced Residual Fusion. It's a deep learning model built just for spotting the ictal (seizure) phase. AMBER uses separate pipelines to pull features from each sensor type, then a residual fusion layer blends those features into one unified package. After that, it runs the fused data through a set of densely connected layers to boost the accuracy of its classifications. The system nails an accuracy of about 90%, and its F1-score is right up there too. On top of that, it shows strong true positive rates for different seizure phases and a low false alarm rate, making it a reliable tool for seizure detection. This all points to the power of combining sensor data when it comes to sorting out seizure activity. Still, the main focus here is on physiological signals while the method proposed in this paper goes in a different direction, using NLP techniques on clinical text data. That opens up another, scalable path for seizure detection [3].

This paper discusses an intelligent solution to detect and prevent seizures in patients with Epilepsy based on the IoHT paradigm. The study recognizes the increasing demand for continuous monitoring tools capable of detecting and predicting the occurrence of seizures in real-time and providing medical care on time. The proposed method suggests implementing IoHT sensors to gather patients' physiological parameters and processing the collected information through machine learning models. Particularly, the authors use the Random Forest and LightGBM approaches to recognize seizure incidents and predict their possible occurrences. As a result, this method contributes to proactive healthcare solutions through warning and enabling immediate clinical actions. Experimental tests prove the high quality of the approach, as it shows a perfect 97.3% accuracy rate and a high recall of 96.7%. The method also significantly improves patients' safety since it provides real-time notifications and ensures low rates of false alarms compared to traditional solutions. Nevertheless, it is highly dependent on physiological data which needs structuring. On the contrary, the algorithm suggested in this paper is based on natural language processing techniques which analyse unstructured clinical data to detect seizures [4].

This work proposed on Actual Problems of Electronics Instrument Engineering (APEIE). It concentrates on the design of a portable apparatus for epilepsy patients' monitoring. The work pays significant attention to difficulties with efficient epilepsy control in spite of the existence of antiepileptic medicines and dangers caused by seizures in the absence of control over the patient. The paper underscores the shortcomings of electroencephalography (EEG) as the sole reliable way of detecting seizures, especially the unfitness of EEG for ambulatory control. In order to tackle the problem, the authors have developed a system that makes use of indirect methods of seizure detection, allowing for permanent monitoring beyond clinical premises. The device has a structured functional architecture and operational modes are outlined in detail. Its purpose includes not only aiding clinical diagnosis but also preventing possible injuries that occur during seizures through monitoring. Possible applications of the portable apparatus are reviewed as well. But its main emphasis is on detection methods that are sensor-based and hardware-dependent, while the algorithm proposed in this study involves the use of Natural Language Processing in analysing the text-based medical data for detecting seizures [5].

A mixed methodology is used for seizure detection in Epilepsy patients using electroencephalogram (EEG) recordings. It aims at solving the significant issue of existing EEG-based approaches associated with the necessity to utilize vast datasets labelled by experts, which is both labour-intensive and time-consuming. To alleviate the problem, the authors propose a hybrid system involving unsupervised learning (UL) and supervised learning (SL) components that decrease the cost and increase efficiency of data labelling while maintaining the high accuracy of detection. Within the UL component, seizure identification takes place by extracting amplitude-integrated EEG, implementing isolation forest anomaly detection, adaptive segmenting, and assessing silhouette scores. As a result, the information is divided into three categories: seizures, non-seizures, and uncertain samples. In order to classify the last group more accurately, this methodology employs the SL component based on the Easy Ensemble algorithm, which successfully deals with the class imbalance problem. Evaluation results were obtained on the CHB-MIT EEG database by which this system scored an average of 92.62% for accuracy, 95.55% for sensitivity, and 92.57% for specificity; thus, this method proved to be comparable in its performance with other existing state-of-the-art techniques. The suggested model shows the power of the combination of unsupervised and supervised learning when it comes to efficient seizure detection with EEG readings. However, this method works only with signal data, while the suggested algorithm in this research paper works with NLP [6].

The paper primarily concentrates on Biomedical Engineering, specifically tackling seizure detection in Epilepsy patients by proposing a wearable solution to address the challenges associated with video and EEG-based seizure detection methods. According to the study, such techniques have been limited to clinical settings, requiring expertise to interpret results. ` The presented method is based on a remote wireless monitoring tool with a wrist-worn accelerometer which is able to detect different kinds of convulsive seizures, including those that last for a shorter period. Features based on time domain data extraction and an innovative set of Poincare plot-derived characteristics are used. Area under the ROC curve feature selection method is implemented in order to pick the most informative features for classification. Classification itself involves the kernelized support vector data description algorithm that separates data into seizures and non-seizures. The system was tested using a massive amount of data (5576 hours) from 79 patients. It resulted in detection of 86.95% of seizures while producing a small number of false alarms. It is noteworthy that the system showed higher sensitivity and lower number of false alarms during generalized tonic-colonic seizures. Such an approach indicates the potential of sensor-based technology for monitoring seizures continuously and non-invasively. Yet, the main advantage of this approach is that it uses motion-derived physiologic data, while the technique suggested in the paper utilizes natural language processing for analysing text clinical data [7].

This study was conducted to develop an automated detection of seizures using electroencephalogram (EEG) signals on patients suffering from Epilepsy disease. As stated in this study, it is important to detect EEG asymmetry patterns in relation to seizure conditions as a result of the existence of electrical anomalies in the brain. In this paper, a deep learning model has been proposed that uses Deep Neural Network (DNN) with Binary Dragonfly Algorithm (BDFA). The use of this type of machine learning algorithm is intended to improve the efficiency of seizure detection through automatic extraction and selection of relevant features of EEG signals. Through the proposed method, it can reduce dependence on manual evaluation by a neurologist since the process is automated. The utilization of a deep neural network is essential to extract patterns from EEG signals because of its learning capabilities. In addition, the use of BDFA is helpful to identify features that contribute to high accuracy during classification. Nevertheless, it largely uses data from signals, but the approach that is discussed in this article involves analysing unstructured clinical text using Natural Language Processing, which can be considered a viable alternative approach [8].

It aims to develop a method to detect convulsive seizures through accelerometer measurements from epileptic patients. In discussing the difficulties of current approaches used to detect seizures, video and EEG monitoring were identified to be ineffective in non-hospital settings due to several constraints. For this reason, the use of a deep learning methodology employing Convolutional Neural Network is proposed as a potential solution to detect General Tonic-Colonic Seizures. Since there were no available datasets to be used for analysis, an accelerometer device was designed and tested to produce movement data from GTCS. To make use of the obtained data, the readings were processed into RGB images that were used for training and testing using different architectures of CNN models namely: Dense Net, ResNet-50, and VGG16. Out of the three, Dense Net was selected as the best model that yielded the highest accuracy, sensitivity, and specificity rates of 99.2%, 98.4%, and 100% respectively. With these outcomes, a detection algorithm was successfully developed that can adapt to each patient. The proposed technique shows that deep learning methods can be effectively used for seizures detection using wearable devices. Nevertheless, this method is mainly based on motion sensor data, while the approach proposed in the current research is based on NLP techniques applied to text clinical data [9].

This approach presents an exhaustive comparison of the latest seizure detection algorithms for epilepsy patients using EEG-based techniques. The study identifies the absence of standard evaluation criteria and performance indicators as one of the main obstacles in the field of seizures. This research proposes a standardized benchmarking framework for evaluating the algorithms called SzCORE (Seizure Community Open-source Research Evaluation). The researchers reviewed 19 studies in the literature and implemented some of the proposed algorithms to evaluate their performance based on the standard benchmarking procedure. The results showed that there is a marked difference between the performance reported in literature and what was observed in the standardized benchmarking process. This study shows that although all algorithms have high sensitivities in detecting seizures, mostly above 90%, they show poor precision. Moreover, the paper mentions high variation in results based on the type of data utilized, stressing the significance of maintaining uniform standards when conducting an assessment. The article points out the need for benchmarking criteria that would enhance the accuracy and consistency of seizure detection algorithms. Nevertheless, it mainly concentrates on EEG signal processing methods, while this paper uses Natural Language Processing techniques to analyse text data[10].

The work presents the results of the research conducted in order to develop an effective method of classifying different types of seizures occurring in patients suffering from Epilepsy through the use of EEG data. Special attention is paid to the need to identify seizures correctly, which, in turn, will affect further treatment choice, medications, and lifestyle of patients. Within this framework, the application of different machine learning methods for classifying multiclass seizure types based on the TUH EEG Seizure Corpus is considered. A large-scale analysis of the most appropriate preprocessing techniques, machine learning algorithms, and their hyperparameters, which would allow finding optimal combinations for better classification performance, is carried out. The proposed models show promising results with the weighted F1-score reaching up to 0.901 during seizure-wise cross-validation and 0.561 during patient-wise cross-validation. This algorithm illustrates the feasibility of using machine learning algorithms for performing complex classifications. But the problem is that in order to build this algorithm, only the electroencephalographic signals had to be used, while in the presented paper, the technique of natural language processing is applied to analyse textual information [11].

This paper explores the problem of timely seizure detection in patients suffering from Epilepsy. It identifies some of the key difficulties faced when diagnosing seizures, especially in rural areas where there is no immediate availability of specialists. In addition, the paper stresses the shortcomings of the conventional method of Electroencephalogram (EEG) analysis that involves interpreting results manually and is a time-consuming process. In order to overcome these problems, this research proposes developing a machine learning-based seizure detection model, which will facilitate rapid and efficient detection. Through this research, an extensive comparative study of several machine learning algorithms will be carried out to investigate their suitability for seizure detection. The aim of the paper is to determine the best algorithm through a comparative analysis of its accuracy and efficiency. However, the method concentrates mainly on signals that have structure while the method proposed by us uses NLP techniques to process unstructured clinical text, hence being complementary in nature [12].

III PROPOSED METHODOLOGY

The developed system will make use of Natural Language Processing to propose an intelligent seizure detection model that analyses unstructured clinical text data. In contrast to traditional models based on electroencephalogram (EEG) data, this intelligent seizure detection system relies on processing unstructured clinical text data to gain valuable insights into detecting Epilepsy. This is important because most medical data are available in an unstructured form. Therefore, there is a need to develop an effective solution that translates textual data into valuable knowledge that supports detecting Epilepsy. Moreover, the framework can be easily scaled up, implemented at low cost, and can support clinical decision-making systems deployed in hospitals and telemedicine settings. Data Acquisition and Preprocessing Phase : In the first phase, we collect clinical text data from credible healthcare organizations and preprocess the data in several ways to improve quality and consistency. Tokenization, removal of stop words, stemming or lemmatization, deletion of punctuation symbols, and text normalization are some of the preprocessing steps used. Besides, we preserve domain-specific terminologies related to medicine during the data preprocessing phase. The text after preprocessing is then represented in structures by applying features using methods such as Term Frequency-Inverse Document Frequency (TF-IDF), Bag-of-Words (BoW), Word2Vec, and BERT. These structures are able to represent the semantics and context within the text, helping the system to understand patterns of seizures. In the next step, machine learning and deep learning models are used for classification and predictions. Algorithms such as logistic regression, SVM, random forest, as well as long short-term memory (LSTM) and transformers neural networks models are being applied on the extracted features. Accuracy of the algorithm and its efficiency are measured by splitting the dataset into the test and training parts using indicators including precision, recall, F1-score. After hyperparameters tuning, the model receives the ability to differentiate between patterns of text associated with seizures from those that are unrelated to it.

Finally, after training the model, it is used for real-time predictions and decision-making. Upon the input of new data, it is classified into one of two categories – with and without epilepsy related signs and symptoms. The proposed approach to problem solving can be represented by the following pipeline: Clinical Text Input → Preprocessing → Feature Extraction → Model Training → Classification Output.

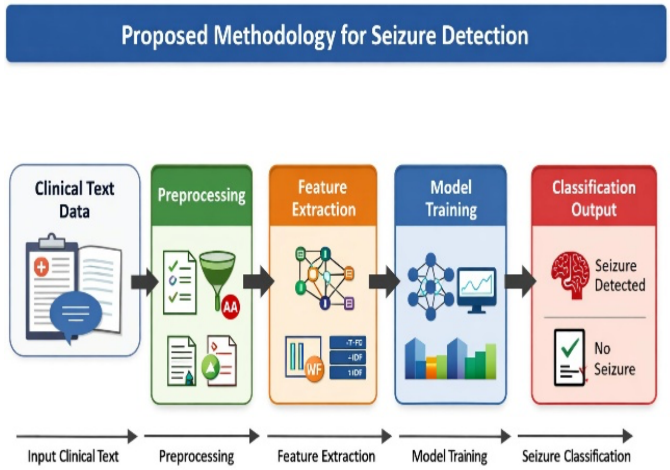

Fig 1 : Proposed Work

This method not only improves accuracy rates but also adds value to the current signal-based methods by harnessing unutilized textual information, thereby ensuring better patient outcomes and effective healthcare data analysis.

IV WORKFLOW

Clinical Text Data Acquisition: Unstructured clinical text data acquisition from sources such as Electronic Health Records (EHRs), physicians' notes, discharge summaries, and patients' medical history data.

- **Data Preprocessing and Data Normalization:** Filtering out any noise, unnecessary symbols, and inconsistency in the text data to enhance its quality.
- **Tokenization:** Splitting text data into small components like words or phrases to facilitate easy processing.
- **Stop-word Filtering:** Filtering out common words such as "and," “the” and "is," which have no significant contribution to the analysis process.
- **Stemming/Lemmatization:** Processing the text by removing suffixes/prefixes or reducing words to their basic forms.
- **Medical Term Retention:** Ensuring retention of domain-specific terms to avoid missing out on clinical terms.
- **Structuring Text:** Transforming unstructured textual data to be in the form of structured data that can easily be processed computationally.
- **Feature Extraction:** Applying techniques such as TF-IDF, bag of words, or word embeddings to derive meaningful features from the text.
- **Semantic Representation:** Using semantic relationship capture through models such as Word2Vec and BERT.

- **Dataset Splitting:** Separation of the entire dataset into three categories: Training, validation, and test datasets.
- **Deep Learning Deployment:** Employing deep learning algorithms such as LSTM and transformers to improve pattern recognition capabilities.
- **Parameter Tuning:** Adjusting model parameters and performing cross-validation to improve model performance and prevent overfitting.
- **Performance Assessment:** Evaluating model efficiency by using assessment measures like accuracy, precision, recall, and F1-score.
- **Prediction Results:** Categorizing new clinical text data as either related to seizures or unrelated to seizures in real-time predictions.

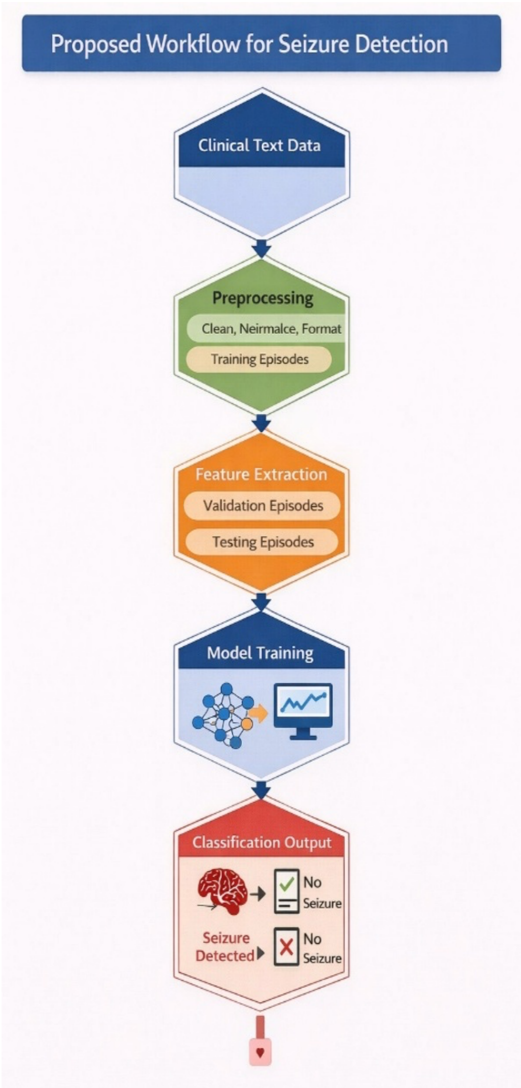

Fig 2: Workflow of the proposed work

.

V RESULT

The analysis of seizure detection framework utilizing NLP techniques was assessed by employing the dataset containing patient medical reports associated with epilepsy. The dataset was split into the training and test sets for the purpose of validating the model performance accurately. Machine learning and deep learning-based models were trained using the extracted features from textual data, and then their performance was compared on the basis of accuracy, precision, recall, and F1 score. It can be seen that the model proposed has been demonstrated to be able to detect seizures effectively from unstructured clinical text. From all of the tested models, the model based on deep learning algorithms, specifically transformer-based models such as BERT, outperformed other machine learning models since they captured contextual relationships in the text. The system demonstrated high classification performance, which implies that it successfully separates the two categories of data. Furthermore, the system demonstrated good balance in terms of achieving high values for precision and recall. As mentioned above, the use of NLP techniques allows for more accurate classification as opposed to traditional methods of seizure detection. Traditional EEG-based techniques use signal processing and require dedicated software and equipment for implementation. In contrast, the proposed technique allows for processing the huge volume of data that is stored as unstructured clinical text. Furthermore, despite the shortage of labelled training data, the system demonstrates adequate accuracy. This suggests good scalability of the solution and its flexibility under different conditions. Confusion matrices allow estimating the rates of true positives and true negatives. High rates imply that the classifier avoids both false-negative and false-positive errors. As shown by the findings, the classifier demonstrates the highest percentage of true positives and the lowest one of false positives. It means that it does not make any mistakes and effectively distinguishes between the two classes. Cross-validation further supports these results because they are consistent with each other. The suggested methodology not only serves to complement current signal-oriented approaches but also offers an innovative way forward in the application of textual medical records for health care analytics. Further developments may include the use of big data, model combinations, and practical applications in clinical settings.

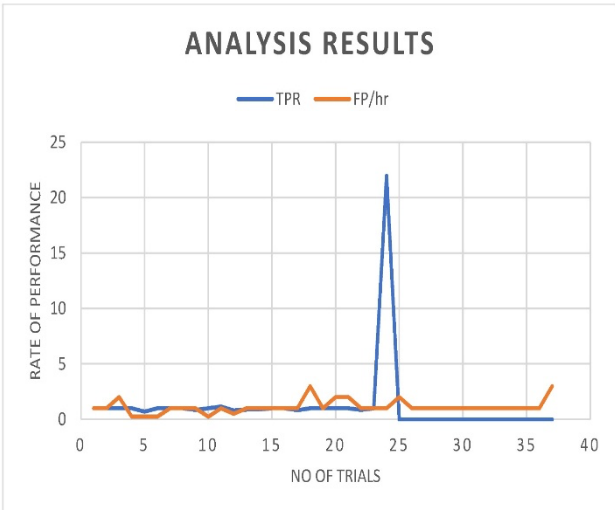

Fig 3: Analysis results

VI CONCLUSION

The proposed work is an innovative solution to detecting seizures through natural language processing approaches has been presented. Instead of depending on EEG readings, which are traditionally employed in conventional solutions, the proposed system makes use of textual data available in-patient medical reports, doctors' notes, and history records. This solution helps overcome common problems associated with conventional systems, including dependence on expensive specialized devices, high cost, and lack of availability. The model built for this purpose shows promising performance in classifying seizure-related and non-seizure-related information. This is achieved by employing preprocessing techniques, feature extraction methodologies, and machine learning/deep learning models to extract syntactic and semantic relations between textual items related to epilepsy. The findings of this study suggest that natural language processing-based approaches can be used as supplementary tools alongside existing signal-based approaches in detecting seizures. In addition to overcoming several drawbacks associated with conventional methods, the proposed solution allows analysing huge amounts of unstructured data in healthcare applications and integrating it into clinical decision support systems. Conclusion

In summary, this research has demonstrated the promise of NLP-based systems in revolutionizing healthcare analytics and patient care. For future research, the approach could be extended to integrate multimodal information sources such as electroencephalogram (EEG) signals and sensor data from wearable devices, as well as applying the system in real-world settings.

VII REFERENCE

[1] Portable closed-loop brain computer interface for seizure control, 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug.–Sep. 2010.

[2] Multi-site closed-loop stimulation for seizure control, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 3, pp. 419–428, Mar. 2019.

[3] AMBER: Attention-guided multi-branching pipeline for multimodal non-EEG seizure detection, IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 5, pp. 3329–3342, May 2025.

[4] AI-enabled epileptic seizure detection and prevention using IoHT, International Conference on Intelligent Systems and Computational Networks (ICISCN), Jan. 2025.

[5] Wearable device for epilepsy monitoring, International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), Oct. 2018.

[6] Hybrid unsupervised and supervised learning for EEG-based seizure detection, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 915–924, Mar. 2022.

[7] Wrist-worn accelerometer-based seizure detection system, IEEE Transactions on Biomedical Engineering, vol. 66, no. 2, pp. 421–432, Feb. 2019.

[8] EEG-based seizure detection using DNN and Binary Dragonfly Algorithm, International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES), Mar. 2025.

[9] CNN-based accelerometer seizure detection using DenseNet, IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dec. 2021.

[10] SzCORE: Standardized evaluation framework for seizure detection algorithms, 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2025.

[11] Machine learning for multiclass seizure type classification, IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Dec. 2020.

[12] Comparative study of machine learning algorithms for seizure detection, International Conference on Computational Intelligence and Computing Applications (ICCICA), Oct. 2025.