Transformer Architectures, Foundation Models and TPU-Accelerated Intelligence for Advances in EEG-Based Seizure Prediction: A Survey

Kishori Sudhir Shekokar[1]Shashi Kant Gupta[2]
1 Postdoctoral Researcher, Lincoln University College, Malaysia;
2 Professor, Lincoln University College, Malaysia;
Email ID: shekokarkishori22@gmail.com , shashigupta@lincoln.edu.my

Abstract: Epilepsy is a chronic condition of the brain that manifests itself in the form of seizures and affects millions of people around the world. When it comes to predicting seizures correctly early on, it could make all the difference from a patient safety and clinical outcome point of view, using electroencephalography (EEG) signals. Generally, the traditional machine learning methods have shown poor capability because of the requirement of manually designing the features to extract them. One recent breakthrough in the deep learning field is the ability to learn complex representations of the spatiotemporal dynamics of EEG data automatically, especially with the use of transformer architectures. Moreover, the rise of the Foundation Models and computing using Tensor Processing Unit (TPU) has presented an opportunity to explore new directions for the development of scalable and efficient seizure prediction systems. This survey addresses the evolution of classical machine learning based models to Transformer inspired models, Foundation Models and intelligent systems based on TPUs for prediction of seizures. An existing comparative analysis to existing approaches is presented that shows strengths and weaknesses. Research gaps are discussed with regard to generalization, explainability, computational complexity and real-time deployment. Last but not least, future research directions related to self-supervised learning, Federated Learning, Explainable Artificial Intelligence (XAI) and TPU-accelerated Foundation Models are presented.

Keywords: Artificial Intelligence, Deep Learning, EEG, Epileptic Seizure Prediction, Foundation Models, Transformer, TPU

Introduction

Epilepsy is one of the most common neurological diseases with over 50 million people affected globally. The most common non-invasive method for measuring brain activity and identifying patterns of seizures is EEG. A key component of an early seizure prediction system is to provide the medical team with the timely interventions needed to prevent a medical emergency and to enhance the quality of life for patients. The traditional seizure prediction systems were based on manually engineered features from EEG and machine learning classifiers like Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF) and K-Nearest Neighbors (KNN). These approaches yielded moderate results, but were very domain specific and relied on manually-designed feature extraction. With the advent of deep-learning, automatic feature learning from raw EEG signals became possible. The Convolutional Neural Networks (CNNs) brought benefits to the spatial feature extraction and the Long Short-Term Memory (LSTM) networks brought benefits to learning the temporal dependency. But even such architectures still had difficulties in modeling long-range dependencies in EEG sequences.In recent years, Transformers have been shown to outperform other models in capturing the complexity of EEG relationships over longer temporal windows by using a self-attention mechanism [1]– [5]. At the same time, the potential of

Foundation Models and TPU-based computing in the field of healthcare intelligence has become a promising technology for scalable healthcare intelligence.

This survey will focus on a comprehensive review of Transformer-based EEG seizure prediction, Foundation Models and TPU enabled artificial intelligence systems.

II Related work

EEG-Based Seizure Prediction Field Has Evolved Over the Years-There are five major phases that could be used to categorize the development of seizure prediction methodologies:

1. Traditional Machine Learning
2. Deep Learning
3. Hybrid Deep Learning
4. Transformer-Based Learning
5. Foundation Models

Figure 1. The history of EEG-based seizure prediction

Traditional machine learning techniques used handcrafted features from frequency domain, time domain and wavelet analysis. Later, models based on deep learning automatically extracted features and resulted in a dramatic boost in prediction accuracy [6]– [8].

The newest developments are the transformer architectures that can model long-distance EEG dependency with the help of self-attention mechanisms [2, 3].

Transformer-Based EEG Analysis-Self-attention mechanisms in transformers have enabled them to capture the contextual relationships between words in the entire input sequence, revolutionizing AI.

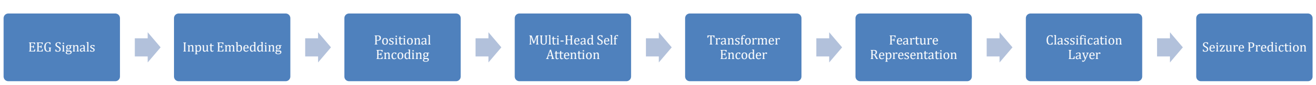

Figure 2. Transformer-Based EEG Seizure Prediction Framework

Transformer-based models offer several significant advantages that make them highly effective for a wide range of machine learning and natural language processing tasks. One of their key strengths is the ability to capture long-range dependencies, allowing the model to identify relationships between distant elements within a sequence without losing contextual information. Their parallel sequence processing capability enables simultaneous computation of all input tokens, resulting in faster training and improved computational efficiency compared to sequential architectures. Furthermore, transformers provide enhanced contextual understanding by employing self-attention mechanisms that dynamically assign importance to different parts of the input, leading to more accurate and meaningful representations. These models also demonstrate excellent scalability, as they can be efficiently trained on large datasets and adapted to increasingly complex tasks by expanding model size and computational resources. In addition, transformer architectures exhibit improved generalization across diverse datasets, enabling them to learn robust patterns that transfer effectively to unseen data, thereby enhancing prediction accuracy and overall model performance. In recent years, Vision Transformers and EEG Transformers have been shown to yield higher seizure prediction accuracy than CNN and LSTM based architectures [1]– [5].

Table I-Comparison of CNN, LSTM, and Transformer Models

Features	CNN	LSTM	Transformer
Spatial Learning	Strong	Limited	Strong
Temporal Learning	Moderate	Strong	Excellent
Long-Range Dependencies	Limited	Moderate	Excellent
Parallel Processing	High	Low	High
Scalability	Moderate	Moderate	Excellent
Generalization	Moderate	Good	Excellent

Foundation models are large-scale neural networks that are pretrained on extensive and diverse datasets before being adapted to specific downstream tasks through fine-tuning, enabling them to learn highly transferable and robust feature representations [10]. This paradigm has transformed artificial intelligence across multiple domains, with prominent examples including GPT for natural language processing and Vision Transformers (ViTs) for computer vision, while similar approaches are increasingly being adopted for electroencephalography (EEG) analysis [3,10]. Unlike conventional deep learning models that rely heavily on task-specific labeled datasets, EEG foundation models leverage self-supervised learning to extract meaningful representations from large collections of unlabeled EEG recordings, thereby significantly reducing the dependence on manual annotation [6,10]. These pretrained representations enhance transfer learning by allowing knowledge acquired during pretraining to be effectively adapted to downstream applications such as seizure prediction, seizure classification, sleep staging, and brain–computer interface systems [3,5–8]. Furthermore, EEG foundation models demonstrate improved cross-patient generalization by learning generalized neural patterns that remain effective across diverse patient populations and recording conditions, resulting in greater robustness to inter-subject variability, noise, and differences in acquisition protocols [3,5,6]. Recent studies have shown that integrating pretrained EEG representations with self-supervised learning frameworks and transformer-based architectures substantially improves seizure prediction and classification performance compared with conventional supervised approaches, highlighting their potential for developing accurate, scalable, and clinically reliable diagnostic systems [1–6,10]. Consequently, EEG foundation models represent a significant advancement in intelligent brain signal analysis by combining data-efficient learning, improved generalization, and robust feature extraction, thereby supporting the development of next-generation clinical decision support systems for neurological disorders such as epilepsy [5,6,9,10].

TPU-Accelerated Artificial Intelligence-The architecture of a transformer needs lots of matrix work and considerable computational power. Tensor Processing Units (TPUs) are specialized hardware accelerators that are specifically designed to handle large-scale deep learning tasks [9].

Table II-Hardware Platform Comparison

Parameters	CPU	GPU	TPU
Training Speed	Low	High	Very High
Parallel Processing	Limited	Good	Excellent
Transformer Optimization	Basic	Good	Excellent
Energy Efficiency	Moderate	Moderate	High

These systems using TPU can decrease training time considerably, enhance scalability and energy efficiency. As a result, TPUs are widely being used for healthcare-related applications that leverage the Transformer model.

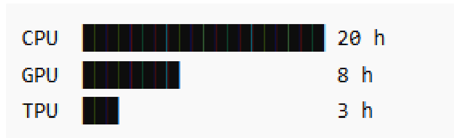

Figure 3. Illustrative Training Time Comparison

III Comparative analysis and research challenges.

Recent advances have been made in EEG-based seizure prediction with the use of advanced deep learning; there are still several challenges that hinder these methods from being effectively used in clinical practice. Another crucial problem is the poor generalization of many models, due to the differences in the EEG recording setup, patient characteristics and clinical environments, which might cause differences in model performance across datasets. One of the difficulties is the excessive number of experts being forced to label EEG data, which takes a lot of time, effort and domain knowledge to get right. Furthermore, more sophisticated architectures like Transformers have shown great potential in predicting information, though they tend to be resource-heavy, limiting their use in environments with limited resources. The opacity of models in their decision-making process also diminishes clinical adoption, as clinicians must be able to interpret the models' decision-making process to trust and adopt the automated predictions. In addition, due to the real-time nature of the application, computing power, memory requirement and energy consumption are still important challenges to be addressed in reliable prediction of seizures, especially in wearable and edge-based healthcare devices. Although existing research has sought to address some of these concerns in isolation, it is the lack of a single comprehensive solution that is needed to address all the major challenges at times. Specifically, there is limited research that looks at using Foundation Models, TPU-enabled computation, Explainable Artificial Intelligence (XAI), Federated Learning, and uncertainty-aware prediction simultaneously. Such a comprehensive approach has the potential to enhance the reliability of predictions, transparency of the model, privacy of the data and the efficiency of deployment, leading to more practical and clinically useful seizure prediction systems.

Table III-Comparative Analysis of Seizure Prediction Approaches

Approach	Strength	Limitation
Traditional ML	Computationally Efficient	Handcrafted Features
CNN	Automatic Feature Learning	Limited Temporal Learning
LSTM	Temporal Modeling	High Complexity
Transformer	Long-Range Context Learning	Resource Intensive
Foundation Models	Transfer Learning	Large-Scale Pretraining
TPU-Accelerated Models	High Scalability	Hardware Dependency

IV Future Research Directions

EEG-based seizure prediction will likely benefit from the fusion of multiple complementary technologies and learning paradigms in the future. Self-supervised learning can be applied to take advantage of existing large-scale unlabeled EEG data, which otherwise would require expensive expert labels. The Foundation Transformer models can help learn generalized representations from different datasets, which can enhance the robustness and transferability of models among patient groups. For training and inference, the use of the TPU-accelerated frameworks can further optimize the training and inference efficiency and help build more advanced models within practical time limits. At the same time, Federated Learning offers a privacy-preserving solution for federated model training among different health institutions without the interposition of the data. Integrating the concepts of Explainable Artificial Intelligence (XAI) can enhance the transparency of AI-driven systems and facilitate clinical decision-making through the generation of meaningful insights into AI predictions. Additionally, Edge AI can be used to monitor and predict seizures in real-time via a wearable device without requiring extensive computational power. Last, uncertainty-aware prediction methods can be used to estimate the confidence level of the predictions of the model, which may result in more reliable and trusted clinical uses. These strategies can create a new era of intelligent, scalable and patient-centric seizure prediction systems.

V Conclusion

This survey covered the development of EEG-based seizure prediction from conventional machine learning approaches to Transformer-based architectures, Foundation Models and TPU-based intelligent systems. Transformer models have shown remarkable performance in modeling the complex dependencies within EEG signals, whereas Foundation Models have shown potential in transfer learning and generalized EEG intelligence. Further, TPU acceleration allows for the efficient training and deployment of models that are computationally intensive. There are, however, issues of generalisation, explanation and implementation in real-time which continue to be open research questions. To build powerful and clinically viable seizure prediction models, further research needs to be conducted in several key areas: Foundation Models, TPU acceleration, Federated Learning, and Explainable AI.

References

1. S. Yuan, K. Yan, S. Wang, J.-X. Liu and J. Wang, "EEG-based seizure prediction using hybrid DenseNet–ViT network with attention fusion," *Brain Sciences*, vol. 14, no. 8, Art. no. 839, 2024. https://doi.org/10.3390/brainsci14080839
2. E. Vafaei and M. Hosseini, "Transformers in EEG analysis: A review of architectures and applications in motor imagery, seizure, and emotion classification," *Sensors*, vol. 25, no. 5, Art. no. 1293, 2025.https://doi.org/10.3390/s25051293
3. Y. Huang, Y. Chen, S. Xu, D. Wu and X. Wu, "Self-supervised learning with adaptive frequency-time attention transformer for seizure prediction and classification," *Brain Sciences*, vol. 15, no. 4, Art. no. 382, 2025.https://doi.org/10.3390/brainsci15040382
4. D. P. Yedurkar, S. P. Metkar, T. Stephan, V. Mohan and S. Agarwal, "SPA-IoT with MCSV-CNN: A novel IoT-enabled method for robust pre-ictal seizure prediction," *BMC Medical Informatics and Decision Making*, vol. 25, Art. no. 354, 2025.https://doi.org/10.1186/s12911-025-03191-5
5. R. R. Borhade, S. S. Barekar, S. N. Ohatkar, P. K. Mathurkar, R. H. Borhade and P. M. Bangare, "ResNeXt-Lenet: A hybrid deep learning approach for epileptic seizure prediction," *Intelligent Decision Technologies*, vol. 18, no. 3, 2024.https://doi.org/10.3233/IDT-240923
6. N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, R. Boyle *et al.*, "Ten years of Tensor Processing Units," *Communications of the ACM*, vol. 67, no. 11, pp. 44–57, 2024.
7. J. Niu, A. Ravi, N. Florez Torres, D. Baum, T. Morshedzadeh, C. Aimone and M. Abou Jaoude, "Foundation models on wearable EEG using self-supervised learning," in *Proc. IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2025, pp. 1–5. https://doi.org/10.1109/EMBC58623.2025.11254670

8. Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, Art. no. 051001, 2020.https://doi.org/10.1088/1741-2552/ab260c
9. K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi and J. Qadir, "Explainable, trustworthy, and ethical machine learning for healthcare: A survey," *Computers in Biology and Medicine*, vol. 139, Art. no. 104855, 2021.https://doi.org/10.1016/j.compbiomed.2021.104855
10. World Health Organization, "Epilepsy fact sheet," World Health Organization, 2025. https://www.who.int/news-room/fact-sheets/detail/epilepsy