Epilepsy & Behavior 155 (2024) 109732

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Research Paper

Attention-based deep convolutional neural network for classification of generalized and focal epileptic seizures

Taimur Shahzad Gill *, Syed Sajjad Haider Zaidi, Muhammad Ayaz Shirazi

Department of Electronics and Power Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan

ARTICLE INFO

Keywords:
Convolution Neural Network
Electroencephalogram
Epilepsy
Epileptic Seizures
Multi-Headed Attention Mechanism

ABSTRACT

Epilepsy affects over 50 million people globally. Electroencephalography is critical for epilepsy diagnosis, but manual seizure classification is time-consuming and requires extensive expertise. This paper presents an automated multi-class seizure classification model using EEG signals from the Temple University Hospital Seizure Corpus ver. 1.5.2. 11 features including time-based correlation, time-based eigenvalues, power spectral density, frequency-based correlation, frequency-based eigenvalues, sample entropy, spectral entropy, logarithmic sum, standard deviation, absolute mean, and ratio of Daubechies D4 wavelet transformed coefficients were extracted from 10-second sliding windows across channels. The model combines multi-head self-attention mechanism with a deep convolutional neural network (CNN) to classify seven subtypes of generalized and focal epileptic seizures. The model achieved 0.921 weighted accuracy and 0.902 weighted F1 score in classifying focal onset non-motor, generalized onset non-motor, simple partial, complex partial, absence, tonic, and tonic-clonic seizures. In comparison, a CNN model without multi-head attention achieved 0.767 weighted accuracy. Ablation studies were conducted to validate the importance of transformer encoders and attention. The promising classification results demonstrate the potential of deep learning for handling EEG complexity and improving epilepsy diagnosis. This seizure classification model could enable timely interventions when translated into clinical practice.

1. Introduction

The International League Against Epilepsy defines epilepsy as a temporary episode of signs and symptoms brought on by erratic neuronal activity and abnormal neural synchronization in the brain [1]. It is the most prevalent chronic brain neurological disorder, affecting over 50 million individuals globally due to the excessive electrical activity of the brain cells, and is characterized by epileptic seizures [2]. These epileptic seizures cause loss of consciousness, which may be implicated neurologically, physiologically, socially, and cognitively. There may be a mortality risk without competent monitoring and diagnosis [2].

The electroencephalogram (EEG) analysis is the gold-rated standard for diagnosing individuals with epilepsy [3]. However, it is laborious, time-consuming, and prone to error for a neurologist to visually observe and diagnose epileptic seizures in an EEG report [4]. To reduce the amount of long-term EEG data that needs to be evaluated by neurologists, it is crucial to develop an automated computer-aided system that allows patients and neurologists to identify and classify epileptic seizures [3,4]. Epileptic seizure classification from EEG signals requires a series of processes, including signal recording, data pre-processing, feature extraction, channel selection, classification, and performance analysis/decision-making. For classifiers to identify between various epileptic seizures correctly and accurately, appropriate and meaningful characteristics must be obtained due to the EEG signal's complex patterns and visual similarity between epileptic and normal signals.

Epilepsy classification is the most important clinical tool for diagnosing a person experiencing seizures [5]. Every clinical consultation is impacted by it, but its influence also extends to basic and clinical epilepsy research as well as the creation of novel treatments [6]. The classification process has numerous benefits, such as providing a framework for identifying the patient's seizure type, predicting the other types of seizures that are more likely, understanding the causes, and often determining their prognosis [7]. The classification also foretells the risk of comorbid conditions like learning difficulties, intellectual disabilities, mental traits like autistic spectrum disorder, and mortality risk like sudden unexpected death in epilepsy. Notably, classification often influences the choice of anti-epileptic treatments [8].

* Corresponding author.
E-mail addresses: taimur.beee19pnec@student.nust.edu.pk (T.S. Gill), sajjadzaidi@pnec.nust.edu.pk (S.S.H. Zaidi), ayaz.shirazi@pnec.nust.edu.pk (M.A. Shirazi).

https://doi.org/10.1016/j.yebeh.2024.109732
Received 9 June 2023; Received in revised form 3 February 2024; Accepted 27 February 2024
1525-5050/

The epilepsy type level comprises generalized epilepsy, focal epilepsy, combined generalized and focal epilepsy, and unknown epilepsy [9]. This paper focuses on two main types of epilepsy: generalized and focal. Generalized epilepsy originates in neural networks (NNs) that involve both hemispheres of the brain. It is characterized by bilateral symmetrical seizure activity [2]. Some examples of generalized seizure types are absence (petit mal), myoclonic, atonic, tonic, and tonic-clonic (grand mal) [9]. Focal epilepsy, previously called partial epilepsy, originates in just one hemisphere of the brain. It is characterized by lateralized or regional seizure activity [2]. Some examples of focal seizure types are simple partial, complex partial, and secondarily generalized [10].

This research is based on the classification of seven seizure types – focal onset non-motor seizures (FNSZ), generalized onset non-motor seizures (GNSZ), simple partial seizures (SPSZ), complex partial seizures (CPSZ), absence seizures (ABSZ), tonic seizures (TNSZ), and tonic-clonic seizures (TCSZ) [1]. Key differences between generalized and focal seizures are their characteristics and origin in the brain. Generalized seizures affect both hemispheres with bilaterally synchronous discharges on EEG. Focal seizures begin in a single area with unilateral or regional EEG abnormalities [9,10]. This paper provides a classification framework based on EEG characteristics for these seven epileptic seizure types across the generalized and focal epilepsies.

Accurately distinguishing between generalized and focal seizure types has important clinical implications that motivate the development of automated EEG classification systems. Identifying whether seizures originate diffusely or regionally guides prognosis, since generalized epilepsies typically have less favorable outcomes and higher treatment resistance [1,2]. Seizure type also informs optimal medication selection, as certain drugs work preferentially for generalized versus focal seizures [3]. Furthermore, recognizing focal onset seizures localized to one hemisphere or lobe is crucial for surgical evaluation, as resection may cure medication-refractory cases [4,5]. Automated seizure classification from EEG signals can assist clinicians in making more informed diagnostic and therapeutic decisions. This work develops deep learning (DL) models to accurately classify seven major generalized and focal seizure morphologies based on differences in their EEG patterns. Successfully classifying these seizure types demonstrates the utility of AI for improving clinical analysis of epilepsy.

2. Literature review

EEG-based automated seizure detection and classification has emerged as a promising technique to assist clinicians in diagnosis, monitoring and treatment of epilepsy [8,10]. This section first outlines the traditional machine learning methods for seizure detection, highlighting their limitations in multi-class classification, and then discusses recent advancements in multi-class seizure classification using DL techniques like convolutional neural networks (CNNs) and long short-term memory (LSTM) for automated multi-class seizure classification on the Temple University Hospital Seizure Corpus (TUSZ) [11–15]. It also analyzes state-of-the-art techniques, compares classification performance across focal, generalized and multiple seizure types, and discusses challenges and limitations in generalizing models to clinical practice.

2.1. Seizure detection and prediction

Seizure detection and prediction are critical for timely therapeutic interventions in epilepsy patients [11]. Early approaches relied on handcrafted features from wavelet and chaos analysis to design statistical classifiers. However, DL methods like CNNs and LSTMs have dominated recent research. Acharya *et al.* [12] proposed a 13-layer CNN architecture for automated seizure detection on EEG signals, achieving over 95% accuracy. Khan *et al.* [13] implemented an LSTM network combined with empirical mode decomposition, attaining 96.5% sensitivity for seizure detection. Although significant progress has been made, real-time seizure prediction remains challenging. Factors like patient-specific neural signatures and inter-subject variability impact consistency [14]. Domain adaptation techniques that account for distribution shifts have potential to enhance generalization capability. More recently, DL has opened new avenues for accurate seizure detection, laying the foundations for reliable prediction and responsive neuromodulation [15,16].

2.2. Seizure classification

Seizure classification using EEG signals has been an active area of research, with studies investigating various methodologies to accurately identify different seizure types. Researchers have focused on developing automated systems that can aid clinicians in diagnosis and treatment [14–20].

2.3. Deep convolutional networks

Deep CNNs have emerged as a popular technique for seizure classification. Sriraam *et al.* [17] proposed a CNN framework to classify seven seizure types and non-seizure, achieving a classification accuracy of 82.14%. More complex CNN architectures like AlexNet, VGG16, VGG19 were explored by Raghu *et al.* [18], with the best accuracy of 88.30% using InceptionV3. Asif *et al.* [19] developed SeizureNet using an ensemble CNN architecture and obtained a weighted F1 score of 0.95 across multiple seizure types. Priyasad *et al.* [20] implemented a deep CNN with SincNet and Conv1D layers, extracting features directly from EEG without preprocessing and achieved an average weighted F1 score of 0.967.

Early breakthroughs using spiking neural networks by Ghosh-Dastidar and Adeli [21,22] highlighted supervised neural models for spike pattern detection. Recent surveys by Gu and Adeli [23] demonstrate deep convolutional networks substantially improve seizure detection versus earlier methods. However, challenges remain in model optimization, validation, and interpretability. Hybrid systems combining neural networks with wavelet feature extraction have potential to leverage the strengths of both techniques [24,25].

2.4. Wavelet-Based approaches

Wavelet transforms have been widely used for feature extraction in seizure classification. Pioneering work by Adeli *et al.* [26] first demonstrated wavelet methods for EEG decomposition and epilepsy diagnosis using discrete Daubechies and harmonic wavelets. Further studies like Sharma *et al.* [27] and Faust *et al.* [28] combined wavelet transforms with statistical features or neural network classifiers to detect seizures from EEG with over 90% accuracy. Saputro *et al.* [29] extracted features using mel-frequency cepstral coefficients and wavelet analysis, attaining a classification accuracy of 91.40%. Wijayanto *et al.* [30] employed empirical mode decomposition and wavelet features, achieving an accuracy of up to 93.33% with SVM classifier. Albaqami *et al.* [31] compared dual-tree complex wavelet transform, discrete wavelet transforms and wavelet packet decomposition, achieving highest weighted F1 score of 0.958 using statistical features.

2.5. Semi-Supervised and transfer learning

Semi-supervised learning using pretrained models has shown promise for seizure classification. Tang *et al.* [32] implemented a self-supervised pretraining strategy with diffusion convolutional RNNs, obtaining a weighted F1 score of 0.749. Transfer learning with CNNs like VGG16, GoogLeNet, ResNet was explored by Raghu *et al.* [14], achieving a maximum accuracy of 88.30%. Dang *et al.* [33] utilised transfer learning with multi-model classification probability fusion, attaining a weighted F1 score of 0.976.

2.6. Ensemble and hybrid models

Researchers have developed ensemble and hybrid models combining multiple techniques to improve classification performance. Liu *et al.* [34] proposed a hybrid bilinear DL model using CNN and RNN feature extractors, achieving a weighted F1 score of 0.974 for multi-class seizure classification. Asif *et al.* [15] implemented an ensemble architecture with multi-spectral feature embedding, obtaining a weighted F1 score of 0.95. Cao *et al.* [35] introduced a hybrid deep network with adversarial learning, attaining a classification accuracy of 94.70%.

Recent studies have developed deep neural networks like CNNs and transformers for automated seizure detection from EEG signals. Bhattacharya *et al.* [36] implemented a deep transformer model to predict seizures, achieving over 90% accuracy. Liu *et al.* [37] proposed a channel-perturbation CNN combined with LSTM for patient-independent seizure detection. Zhou *et al.* [38] introduced a self-organising fuzzy logic system for both patient-specific and cross-patient seizure recognition with strong performance. These works demonstrate capabilities of DL for robust EEG analysis and seizure recognition across diverse patient cohorts. However, standardized benchmarking on public datasets is needed to better compare strengths of different architectures and training mechanisms.

Another focus has been multi-modal neural networks that fuse different feature types for enhanced seizure detection. Yu *et al.* [39] developed a model combining CNN, LSTM, and transfer learning on multi-spectral EEG features, attaining high sensitivity and specificity. Wang *et al.* [40] introduced a dual-modal information bottleneck network integrating both EEG and ECG data, improving generalizability. Hybrid systems that leverage complementary modalities like neuroimaging show promise in mitigating challenges like variability across patients. But more research is required to determine optimal fusion strategies and sensor combinations tailored to seizure morphology.

Recent studies has also explored techniques to improve model optimization, evaluation, and clinical applicability. Nhu *et al.* [41] achieved an F1 score of 0.80 for interictal epileptiform discharge detection using CNN-transformer networks with multi-center EEG data and standardized protocols. Peh *et al.* [42] evaluated models across six centers to enhance generalizability. Zhao *et al.* [43] focused on developing interpretable attention-based networks that can explain seizure recognition. Such efforts are critical to translate advances into clinical impact. However, regulatory hurdles, clinician trust, and ethical considerations remain key barriers to real-world adoption.

2.7. Challenges and limitations

Despite significant progress, automated multi-class seizure classification remains challenging. Most studies have been performed on limited datasets, and model performance drops significantly under patient-specific cross-validation [44]. Class imbalance is another persistent issue, with researchers resorting to oversampling minority classes. There is a need for large multi-center datasets and standardized evaluation protocols to advance the field [14–18]. Generalizability to diverse patient populations is an open problem. Interpretability of complex DL models also needs to be enhanced for practical clinical adoption [45]. Therefore, seizure classification is a promising area but requires further research to address existing limitations.

3. Research methodology

This section details the methodology used in this work for multi-class

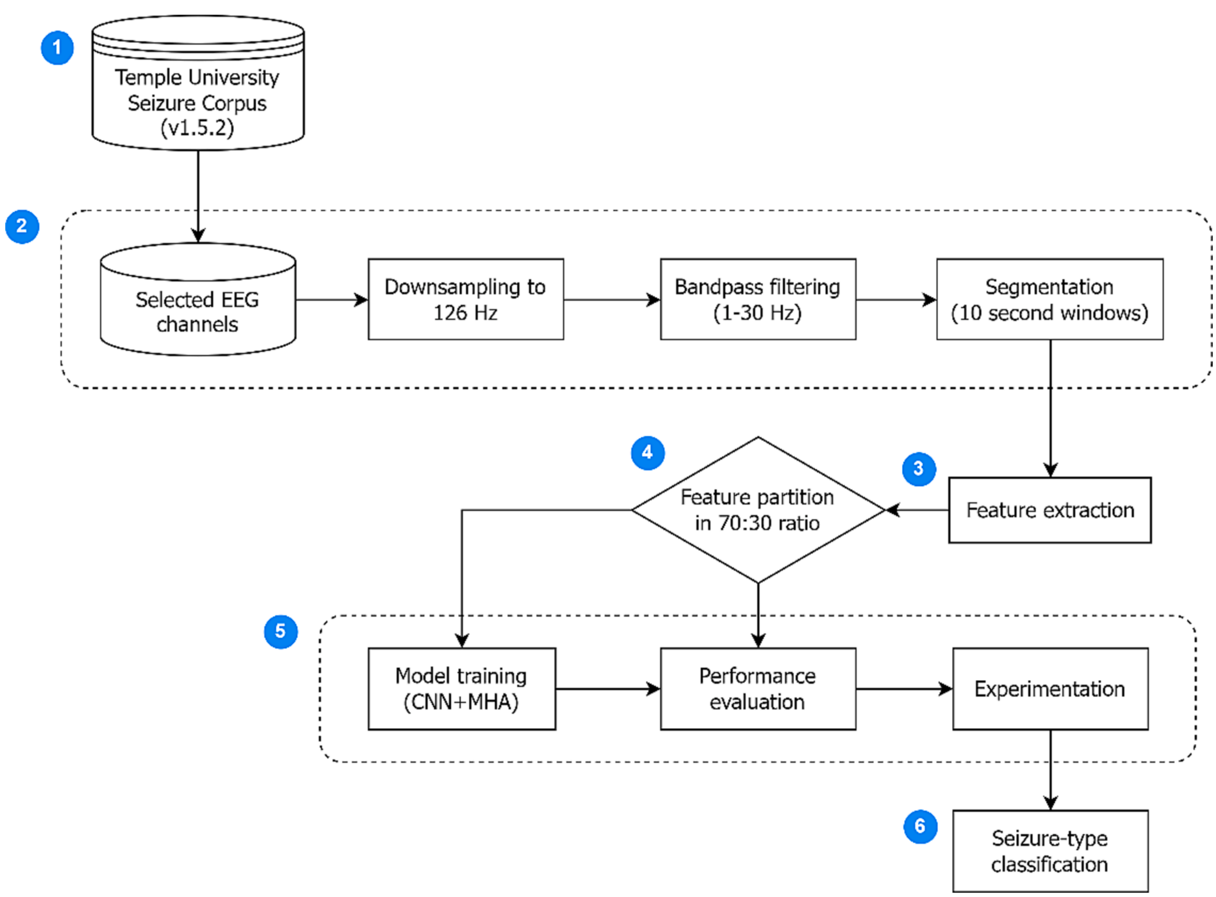

Fig. 1. The methodology flow diagram showing 6 main stages: (1) EEG data acquisition from Temple University Seizure Corpus (ver. 1.5.2), (2) preprocessing of the EEG signals, (3) feature extraction, (4) feature partition for training and testing, (5) training and evaluation of the proposed model, and (6) seizure-type classification.

seizure classification from EEG signals. The overall workflow is shown in Fig. 1. First, EEG recordings are obtained from TUSZ ver. 1.5.2 containing annotations for seven seizure types (as shown in Table 2). Relevant EEG channels are selected and pre-processed through downsampling and filtering. The continuous signals are then split into segments to enable feature extraction on local windows (see Fig. 4). A CNN combined with a Multi-Head Attention (MHA) mechanism is proposed as the classification model (see Fig. 5). CNN automatically learns to extract spatial features, while the MHA learns to focus on relevant parts of the input sequence. After training on 70% of the data, the model performance is evaluated on the remaining 30% across various metrics. Multiple experiments are conducted by modifying model parameters and architectures. The optimized model is finally utilized for automated multi-class seizure classification on EEG test data.

The subsequent sub-sections will elaborate on the dataset, implementation specifics of the various data processing blocks, model architecture details, training methodology, performance evaluation metrics and results. The modular pipeline provides a structured framework to use DL for enhanced EEG-based diagnosis of epilepsy.

3.1. Dataset Description

This study uses TUSZ ver. 1.5.2, the largest publicly available dataset of multi-channel EEG recordings with annotated seizure events (see Table 1) [43,45]. The TUSZ contains over 30,000 EEG recordings collected since 2002 from 1,842 patients in various hospital settings, digitized at 250–1024 Hz with up to 128 channels in the conventional 10–20 system (see Fig. 2) [44]. It includes documentation like medical history and clinician reports to enable annotation of multiple seizure types. The training set has 4,599 EEG files with 2,377 marked seizures, predominantly focal non-specific (65%) and generalized non-specific (17%). The development and evaluation sets have 1,013 and 1,023 files respectively, with similar seizure type distributions [45].

As the only open-source EEG database with annotations for diverse seizure types, the TUSZ has been utilized in most state-of-the-art seizure detection and classification research. The multi-channel recordings contain real-world artifacts, and the varied collection environments and electrode configurations enable robust model development [41]. The annotated seizure events with detailed markings of onset, duration, and type are invaluable for benchmarking classification performance on focal, generalized, and mixed-type seizures. By using the TUSZ's clinical documentation, this study can correlate EEG signals with ground-truth seizure diagnoses to accurately detect and classify the seven key subtypes – FNSZ, GNSZ, SPSZ, CPSZ, ABSZ, TNSZ, and TCSZ (see Table 2).

Table 1
Seizure type descriptions for TUH EEG Seizure Corpus (TUSZ) ver. 1.5.2 [28].

Label	Seizure Type	Signs	Locality	Description
FNSZ	Focal Non-Specific Seizure	Electrographic	Hemispheric/ Focal	Focal seizures which cannot be specified with its type
GNSZ	Generalized Non-Specific Seizure	Electrographic	Generalized	Generalized seizures which cannot be further classified into one of the groups below
SPSZ	Simple Partial Seizure	Clinical & Electrographic	All	Partial seizures during consciousness; type specified by clinical signs only
CPSZ	Complex Partial Seizure	Clinical & Electrographic	All	Partial seizures during unconsciousness; Type specified by clinical signs only
ABSZ	Absence Seizure	Clinical & Electrographic	Generalized	Absence discharges observed on EEG; patient loses consciousness for few seconds (petit mal)
TNSZ	Tonic Seizure	Clinical & Electrographic	All	Stiffening of body during seizure (EEG effects disappear)
TCSZ	Tonic Clonic Seizure	Clinical & Electrographic	All	At first stiffening and then jerking of body (grand mal)

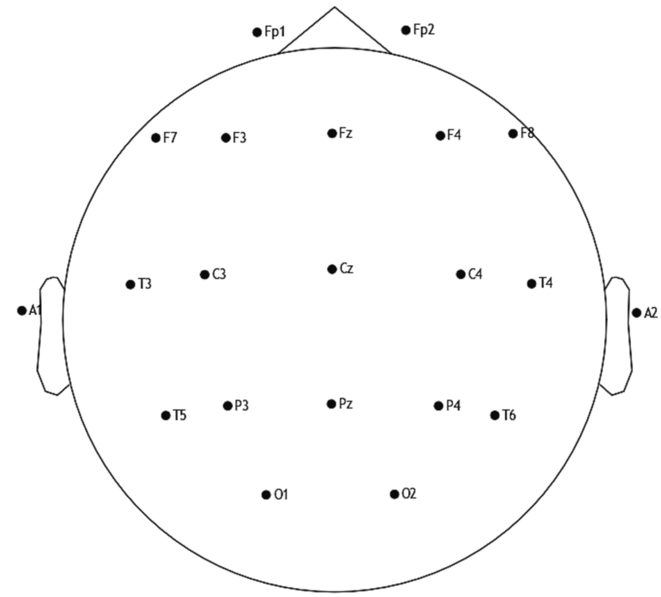

Fig. 2. The 10–20 system for EEG electrodes positioning used in the TUSZ ver. 1.5.2.

3.2. Data preparation and preprocessing

Before the feature extraction process, a series of preliminary steps were carried out to prepare the data for subsequent phases (see Fig. 1). The raw multi-channel EEG signals were loaded into the memory from the European Data Format (EDF) files and the EEG channels common to all the files were identified [46]. For each file, clinically-relevant EEG channels, consistent with those identified were loaded to reduce dimensionality and focus on channels known to provide useful signal content. The sampling frequency and channel names were also retrieved from the EDF file metadata. To generate the corresponding seizure event labels, the annotation files were loaded separately. These provide the start and end times of seizure events which were read and aligned time-wise with the raw EEG data (see Fig. 3).

The continuous raw EEG was then downsampled to 128 Hz from its original sampling rate if over 256 Hz to limit computational requirements while retaining sufficient signal quality. Bandpass filtering was applied to the data to extract frequencies between 1 Hz and 30 Hz, with known relevance for characterizing different seizure types [47].

Table 2
Distribution of the seven seizure types under consideration in TUSZ ver. 1.5.2 [32].

Seizure Type	Seizure Events	Duration (mins)	No. of Patients
FNSZ	1836	2019	150
GNSZ	583	995	81
SPSZ	52	36	3
CPSZ	367	605	41
ABSZ	99	14	12
TNSZ	62	20	3
TCSZ	48	92	12

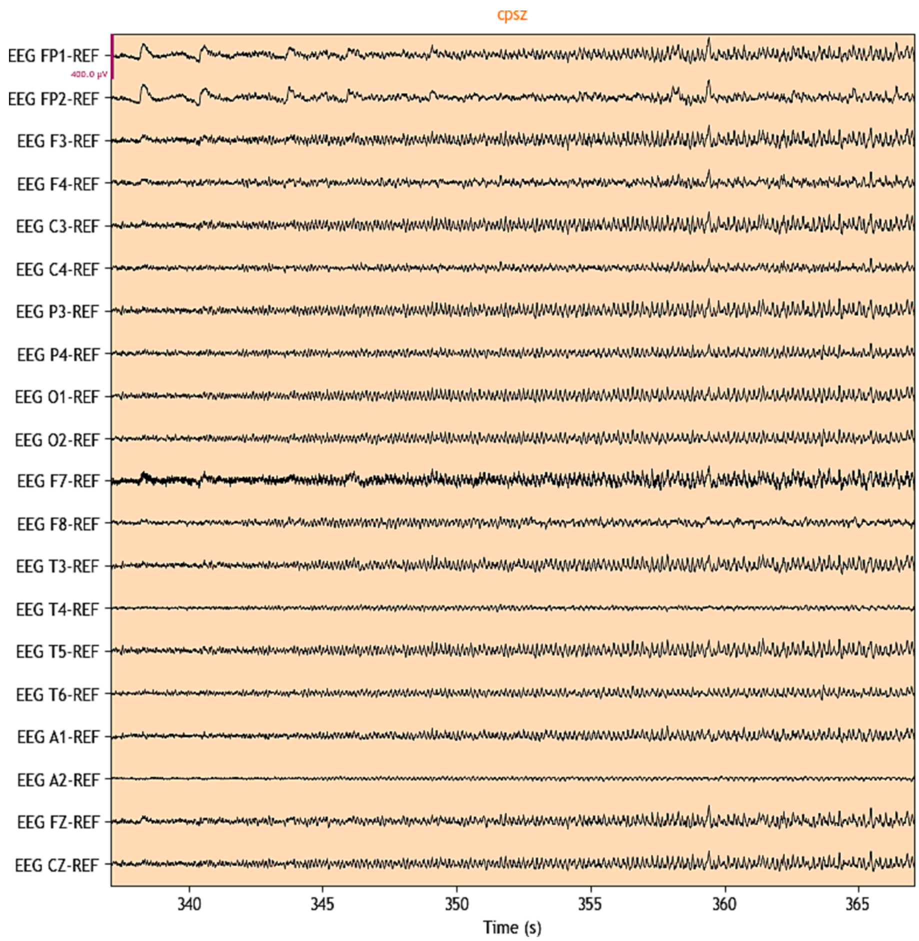

Fig. 3. An example of EEG recording for complex partial seizure (CPSZ) from TUSZ ver. 1.5.2.

Each channel was also z-normalized to have zero mean and unit variance to account for differences in units and variability across the different EEG channels. The filtered continuous data was then segmented into smaller, overlapping segments or 'windows' of 10 seconds, which helped transforming the continuous EEG recordings into a format that facilitates feature extraction on a per-window basis.

3.3. Feature extraction

The feature extraction phase was a critical step to convert the raw EEG data into a machine-readable format that captures the underlying physiological processes (see Fig. 4). In this phase, each windowed segment of EEG data was transformed into a multi-dimensional feature vector [48]. With the windowed EEG signals and label data, a set of features was extracted to enable learning informative representations for seizure classification. Simple statistical features like the mean, variance, skewness, and kurtosis of the amplitude were computed over sliding windows of 10 seconds to quantify changes in the amplitude distribution over time. These capture the presence of spikes, irregular activity, and similar statistical patterns that differentiate seizure types.

Spectral power within standard clinical frequency bands, 1–4 Hz, 4–8 Hz, 8–12 Hz, and 12–30 Hz is calculated to profile the band-specific energy (see Fig. 4). Relative power ratios between bands also highlight which frequencies exhibit the most pronounced activity. Dynamic bandpower ratios provided insight into the evolution of spectral content over time. Beyond basic statistical and spectral features, advanced methods were used to capture complex patterns within the EEG data. The Daubechies D4 transform was used to decompose the signal and the logarithmic sum of wavelet-transformed coefficients was calculated along with absolute mean, standard deviation and ratio of coefficients [49]. FFT-based correlation and eigenvalue features were calculated to capture the relationships between different frequency components. Similarly, time-domain correlation and eigenvalue features were used to understand the temporal relationships between different channels. These features offered a comprehensive view of the underlying brain activity, potentially increasing the discriminative power of the feature set for the classification task.

3.4. Proposed model architecture

The proposed neural network combines attention mechanism with the convolutional layers. The architecture consists of multiple transformer encoder blocks that process the input through layer normalization, MHA, and feed-forward (FF) neural networks (see Fig. 5). These blocks are followed by a Global Average Pooling (GAP) layer and a series of fully connected layers, ending in a softmax output layer for classification. The values of the key hyperparameters used in the model are shown in Table 3.

3.5. Input layer

The input layer of the model expects a multi-dimensional tensor shaped according to the dimensions of the training data, which is typically derived from EEG recordings. Specifically, the input shape is set according to the number of channels, time steps, and features in the EEG data [50]. This layer serves as the gateway for the neural network, enabling the model to take raw or preprocessed EEG signals and funnel them through subsequent layers for feature extraction and classification.

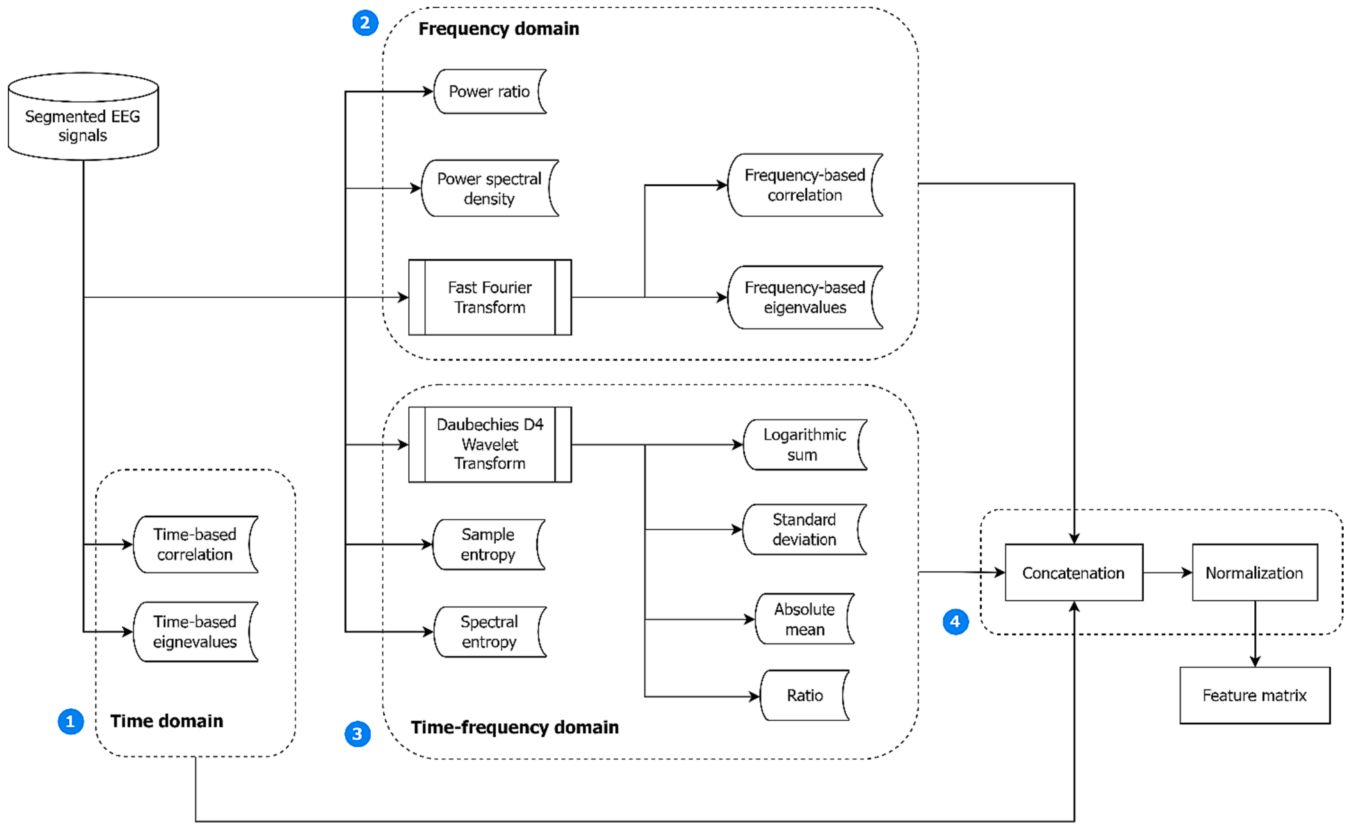

Fig. 4. Feature extraction pipeline showing the extracted features in time domain, frequency domain, and time–frequency domain.

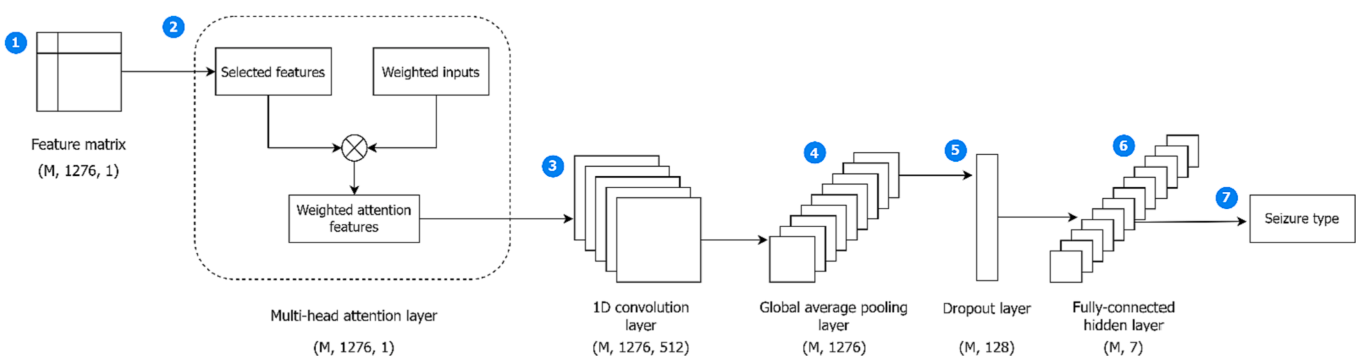

Fig. 5. Architecture of the proposed model showing the five key layers along with the input dimensions of each (M represents the batch size).

Table 3
Hyperparameters used in the proposed model architecture.

Hyperparameter	Value
Head size	256
Number of attention heads	8
Feed-forward dimension	512
Number of transformer blocks	2
MLP units	128
MLP dropout	0.5
Dropout	0.1
Learning rate	1e-4

3.6. *Transformer encoder blocks*

The heart of the proposed neural network model lies in its transformer encoder blocks, which are designed to process the EEG data in a way that captures both local and global contextual information [51]. These blocks are iterative units composed of layer normalization, MHA, and FF neural networks.

3.7. *Normalization layer*

Before the data undergoes attention mechanisms, it is first normalized using a layer normalization technique [52]. The normalization layer standardizes the features by computing the mean and variance for each feature across the input, followed by a scaling and shifting operation. The layer normalization is represented in (1):

$$X_{norm} = \gamma \cdot \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{1}$$

where, X is the input to the layer, μ is the mean, σ^2 is the variance, γ is the scale parameter, β is the shift parameter, and ϵ is a small constant for numerical stability, and is set to 1×10^{-6} in the implementation. Normalization is crucial because it helps in stabilizing the training and makes the model less sensitive to the initialization of parameters.

3.8. Multi-Head attention layer

The MHA layer is the basis of the transformer encoder and enables the model to focus on different parts of the input sequence simultaneously [53,54]. The *head size* specifies the dimensionality of the query, key, and value vectors, while the *number of heads* determines the number of such sets. The model comprises 8 attention heads, each of size 256 as shown in Table 3. Each head computes its own attention scores and produces its own context vectors, which are then concatenated and linearly transformed (see Fig. 5) [55]. Dropout is applied after the attention layer to prevent overfitting. This involves randomly setting some of the layer's outputs to zero during training, which aids in making the model more robust. A residual connection adds the original input back to the output of the attention layer. This helps in mitigating the vanishing gradient problem and allows for deeper models. The MHA mechanism is shown in (2) – (4):

$$\mathrm{MultiHead}(Q,K,V) = \mathrm{concat}(\mathrm{head}_1,\ \mathrm{head}_2,\ \cdots \mathrm{head}_h)^{WO} \tag{2}$$

where each head is shown as:

$$\mathrm{head}_i = \mathrm{Attention}\left(QW_i^Q,\ KW_i^K,\ VW_i^V\right) \tag{3}$$

The scaled dot-product attention is:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d^k}}\right)\cdot V \tag{4}$$

Here, Q, K, and V are the queries, keys, and values, W^O, W_i^Q, W_i^K, and W_i^V are parameter matrices, and d_k is the dimension of the key.

3.9. Feed-Forward network

The last component of the transformer encoder block is the FF network, which consists of 1D convolutional layers (see Fig. 5) [51]. The first convolutional layer uses ReLU (Rectified Linear Unit) as its activation function and has 512 filters with a kernel size of 1. This is followed by another convolutional layer that matches the output dimensions back to the input dimensions to maintain consistency and to facilitate the next residual connection. The FF network acts as a pointwise non-linearity applied to each position, essentially allowing the model to learn more complex representations of the data [56].

The FF network within the transformer block is described in (5):

$$\mathrm{FFN}(x) = \max(0, x\cdot W_1 + b_1)\cdot W_2 + b_2 \tag{5}$$

where W_1 and W_2 are the weights of the first and second convolutional layers, and b_1 and b_2 are their biases.

3.10. Global average pooling layer

Following the series of transformer encoder blocks, the architecture employs a GAP layer. This layer significantly reduces the dimensionality of the feature maps, averaging each feature across its temporal dimension [57]. The data format is set to *channels-first* to ensure that the pooling operation is conducted appropriately across the correct dimensions. The GAP layer serves a dual purpose: it not only simplifies the model by reducing the number of parameters that subsequent layers need to learn but also minimizes the risk of overfitting by providing a form of feature abstraction. The GAP layer is represented in (6):

$$\mathrm{GAP}(X) = \frac{1}{N}\cdot\sum_{i=1}^{N} X_i \tag{6}$$

where N is the number of features to pool over and X_i is the feature map of the i^{th} channel.

3.11. Fully connected layers

After the GAP, the data flows through a set of fully connected layers, often termed Multi-Layer Perceptrons. These layers are designed to learn non-linear combinations of the high-level features extracted by the preceding layers [58]. Each unit in these layers uses a ReLU activation function to introduce non-linearities into the model. Dropout of 0.1 is also applied between these layers to further mitigate the risk of overfitting. The fully connected layer is shown in (7):

$$Y = \max(0, XW + b) \tag{7}$$

where X is the input, W is the weight matrix, b is the bias, and Y is the output.

3.12. Output layer

The final layer in the architecture is the output layer, designed specifically for classification tasks. It employs a softmax activation function, which takes the raw output scores from the previous layer and converts them into probabilities [59]. The number of neurons in this layer corresponds to the number of classes the EEG signals can be classified into. The softmax function ensures that the sum of the probabilities across all classes is one, providing a clear and interpretable output from the model. The output from softmax output is shown in (8):

$$\mathrm{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{8}$$

where z is the input vector to the softmax function, and K is the number of classes.

3.13. Model evaluation

The dataset was split patient-wise in the 80:20 ratio, where 80% of the dataset was used for model training and 20% for evaluation. Since the classes are skewed, the proposed model was evaluated using weighted-average accuracy and F1 score to test the model's robustness and practicality.

3.14. Accuracy

Accuracy (ACC) measures the overall proportion of correct model predictions across all classes as shown in (9) [59]. The model's accuracy was computed on the held-out test set. Accuracy provides an intuitive measure of how often the model correctly classifies seizure types. However, accuracy can be misleading in contexts of significant class imbalance where always predicting the majority class yields high accuracy. Therefore, F1 score was used to ensure even rare classes were classified accurately and not skewing the results.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

3.15. F1 score

The F1 score calculates the harmonic mean of precision and recall, providing a balanced metric combining both properties as shown in (10) [60]. The F1 scores for each seizure type are reported individually, as well as the weighted-average across types to summarize overall performance in Table 4. The F1 score accounts well for class imbalance by considering both false positives and false negatives. High F1 demonstrates that the model achieves both strong sensitivity in detecting true seizure cases, and high positive predictive value in avoiding false positives. F1 scores close to 1 indicate excellent classification since the model is both detecting most real cases and rarely predicting incorrect labels. The F1 score was as the primary metric for model selection during

Table 4
Results obtained for multi-class classification of the seven seizure types.

Seizure Type	Accuracy	F1 score
FNSZ	0.920	0.897
GNSZ	0.949	0.920
SPSZ	0.929	0.911
CPSZ	0.900	0.902
ABSZ	0.893	0.876
TNSZ	0.875	0.898
TCSZ	0.898	0.889
Weighted-average	**0.921**	**0.902**

hyperparameter tuning given its robustness to skew.

$$\mathrm{F1} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})} \tag{10}$$

4. Results

4.1. Experimental settings

In this study, all experiments were conducted using a high-performance Dell Precision 5550 workstation with an 11th generation Intel Core i7 processor, 32 GB RAM, and 1 TB SSD. The models were trained on 80% of the total dataset for all experiments (unless otherwise stated) and evaluated on the unseen test set which was 20% of the dataset.

4.2. Performance evaluation

The proposed model achieved an overall weighted accuracy of 92.1% across all seizure types on the held-out test set. The class-averaged F1 score was 90.2%, providing a balanced measure of performance (see Table 3). Both metrics indicate promising generalization, with the majority of seizures correctly classified by seizure type [61]. The confusion matrix on the test set further illustrates the breakdown of predictions across classes [59]. Most seizures were correctly identified, with some confusion between particular types as expected from their electrographic similarities.

The model performance was assessed for each individual seizure class to determine if certain types were classified more accurately (see Table 3). FNSZ were detected with 92% accuracy and F1 score of 89.7%, indicating the model reliably identified this most frequent class [62]. Accuracy was also strong at 94.9% for GNSZ [63]. However, the rarer seizure types proved more challenging. The model achieved the lowest scores of 87.5% accuracy and 87.6% F1 for TCSZ. This per-class analysis indicates where model improvements should be targeted. However, the strong individual class performance demonstrates the model's classification capabilities across diverse seizure morphologies. Further analysis of incorrectly classified cases revealed patterns causing errors. For example, FNSZ with muscle artifact were sometimes mislabeled as TCSZ. Using amplifier noise and muscle contamination as model features could prevent these false positives [61].

4.3. Ablation study

Ablation experiments were performed to validate the core mechanisms in the proposed model and select appropriate hyperparameters that balance accuracy, efficiency, and generalization [64]. The goal was to empirically derive insights into the model's behavior that could guide improvements.

4.4. Model capacity

The impact of model capacity was evaluated by reducing the number of trainable parameters. Decreasing the transformer head size from 256 to 64 dimensions decreased accuracy from 92.1% to 82.4% while improving training time (see Table 5). Using fewer attention heads also reduced accuracy. The smaller model still performed reasonably but the higher capacity architecture better captured subtle seizure patterns. This demonstrates a tradeoff between representational power and computational efficiency that should be tuned based on requirements [65].

4.5. Attention mechanism

Attention mechanism ablation assessed the importance of MHA for seizure feature learning [66]. The transformer encoders were replaced with a basic CNN, removing the attention components (see Table 5). This significantly reduced accuracy from 92.1% to 76.7%, validating attention's value. By enabling focus on salient input relationships, attention provides more discriminative representations than convolutional feature extraction alone. The results empirically highlight attention as a core model mechanism.

4.6. Regularization

The model's regularization and generalization capabilities were evaluated by modifying key hyperparameters (see Table 5). Increasing the dropout rate from 0.1 to 0.3 introduced more randomness but had minimal impact on accuracy, only decreasing it within 3–4%. This demonstrates the model can regularize well even with high dropout [67].

5. Discussion

The proposed attention-based deep CNN achieved strong performance in automated seizure classification, with 92.1% weighted accuracy and 90.2% F1 score across diverse EEG recordings in the TUSZ ver. 1.5.2 [33]. The attention-based CNN demonstrated effective learning of discriminative patterns to distinguish FNSZ, GNSZ, CPSZ, and ABSZ [62,63]. Ablation studies provided insights into model optimization, highlighting the importance of transformer encoders and MHA for maximizing classification accuracy.

5.1. Research contribution

This research focuses specifically on classification of 7 major subtypes of clinical relevance across the focal and generalized categories using their differences in EEG patterns. The proposed deep CNN architecture with multi-head self-attention is designed to capture both spatial and temporal relationships in EEG signals. It builds on recent advances in transformer networks for enhanced feature learning [74,76]. The model is rigorously evaluated on the extensive TUSZ ver. 1.5.2 using proper cross-validation protocols to assess real-world efficacy [33]. This work moves beyond seizure detection to address the specialized and clinically valuable task of automated multi-class classification. The optimized neural network demonstrates the capabilities of AI for characterizing epileptic seizures and assisting diagnostic decision-making.

5.2. Practical implications and limitations

The promising results validate the potential of DL for automated seizure classification from EEG. The model performance aligned with recent literature demonstrating accuracies exceeding 90% on the TUSZ ver. 1.5.2 using advanced neural networks (see Table 6) [68,69]. The confusion matrix revealed clinically expected overlaps between particular seizure morphologies reflecting their electrographic similarities [59]. However, rare seizure types like TNSZ proved most challenging, indicating opportunities to tailor model improvements towards better feature learning for minority classes [61]. Despite the challenges, the findings reinforce deep CNN with attention mechanisms as state-of-the-art approaches for handling the complexity of EEG data for seizure

Table 5
Results of the ablation study conducted by varying head size, number of heads, and dropout[1].

Model	Head Size	Number of Heads	Trainable Parameters	Dropout	Time/epoch (min)	Weighted-Average Accuracy
CNN with MHA	256	8	195,470	0.3	34.2	0.891
	256	6	188,302	0.3	25.6	0.863
	256	4	181,134	0.3	17.8	0.872
	256	2	173,966	0.3	9.4	0.873
	256	8	195,470	0.2	34.1	0.876
	256	8	195,470	0.1	34.9	0.921
	128	8	181,134	0.3	22.1	0.882
	64	8	173,966	0.3	15.5	0.824
CNN without MHA	N/A	N/A	3,580	0	0.50	0.756
	N/A	N/A	3,580	0.1	0.50	0.767
	N/A	N/A	3,580	0.2	0.40	0.753
	N/A	N/A	3,580	0.3	0.40	0.745

[1] The number of epochs and batch size were held constant at 10 and 4, respectively.

Table 6
Comparison of existing literature on seizure classification with present study.

Author(s)	Methods	Number of Seizure Types	Accuracy (%)
Dwi Saputro et al. [21]	SVM	4	91.40
Sriraam et al. [17]	CNN (AlexNet)	7	84.06
Song et al. [14]	PaHMM	7	81.00
Raghu et al. [18]	CNN (Inceptionv3)	7	88.30
Li et al. [29]	CE-stSENet	7	92.00
Shankar et al. [33]	CNN	3	89.91
Present Study	MHA-CNN	7	92.10

classification. This could pave the way for real-time monitoring systems that alert clinicians whenever a patient's seizure type changes, enabling rapid personalized interventions.

A key strength of the proposed model is its utilization of MHA, which enables seizure-specific feature learning from EEG inputs by focusing on relevant spatial and temporal relationships [17,68]. The transformer architecture extracts higher-level representations compared to standard CNNs, boosting classification performance [69]. Techniques like layer normalization and residual connections aid training convergence and stability [70]. The model generalizes well due to regularization through dropout and early stopping [71]. It achieves balances between accuracy and efficiency that can be tailored based on computational constraints [72]. The modular architecture also allows flexibility, where components can be modified or expanded as classification needs evolve.

Despite promising results, several limitations present opportunities for future work. The model could be improved by using raw EEG instead of hand-engineered features, enabling end-to-end learning [73]. Class imbalance remains a concern, which oversampling, or loss weighting could address [74]. Though the TUSZ is extensive, testing on more heterogeneous EEG data would further validate model robustness [75]. Attention visualization techniques could provide interpretability regarding how the model focuses on salient seizure morphology [76]. Architectural enhancements like convolutional feature extractors or recurrent layers may boost performance [31]. Moreover, deploying the model in clinical environments and longitudinal trials would be essential for assessing real-world efficacy in assisting diagnostics and closed-loop interventions.

6. Conclusion and future work

This work developed an attention-based deep CNN for automated multi-class seizure classification from EEG signals. The model was trained and tested on the extensive TUSZ, using expert annotations for four key generalized seizure types. The transformer architecture with normalization and MHA achieved over 90% accuracy in classifying complex seizure morphologies. Ablation studies provided insights to guide architecture optimization and training strategies.

Automated seizure classification can significantly improve diagnostic workflows and improve treatment for epilepsy patients [77]. DL promises clinically useful performance as showcased by these results [68]. By quickly and accurately discriminating seizure types, such systems could enable timely interventions and reduced monitoring burdens [75]. Reliable classification also facilitates personalized medication selection and seizure control [78]. This work demonstrates transformers and attention as promising techniques for handling EEG complexity. Further optimization could make these models suitable for deployment in ICUs and wearables for real-time monitoring [79]. Therefore, advanced neural networks present new possibilities for enhancing epilepsy management and quality of life for patients through data-driven analytics.

While this work focused on deep CNNs, the field of automated seizure classification is rapidly evolving. Promising new techniques like neural dynamic classification [80], dynamic ensemble learning [81], and finite element machines [82] have emerged for pattern recognition and time-series modeling. Using self-supervised learning [83] and other innovative algorithms could enhance performance and generalization. Competitive benchmarking of these methods in standardized frameworks would help evaluate their relative strengths and weaknesses on seizure morphologies. Moreover, evaluating synergistic fusion of complementary techniques may also unlock new capabilities exceeding individual models.

Future epilepsy research should continuously evaluate state-of-the-art machine learning paradigms beyond DL. Advances in model interpretability, uncertainty quantification, and trustworthiness will also be critical to translate these technologies from the lab to clinical practice [7,54]. However, multidisciplinary collaboration and infrastructure for robust real-world validation are imperative to instill confidence for clinical adoption [38–43]. This work aimed to demonstrate the potential of optimized NNs for multi-class seizure classification. But continued exploration of innovative techniques through collaborative efforts will be key to convert automated EEG analytics into real-world impact for patients.

CRediT authorship contribution statement

Taimur Shahzad Gill: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Syed Sajjad Haider Zaidi:** Project administration, Resources, Supervision. **Muhammad Ayaz Shirazi:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Shellhaas RA. Seizure classification, etiology, and management. Handb Clin Neurol 2019;162:347–61.
[2] World Health Organization, "Epilepsy," World Health Organization, Feb. 09, 2022. https://www.who.int/news-room/fact-sheets/detail/epilepsy.
[3] Tiwari S, Sharma V, Mujawar M, Mishra YK, Kaushik A, Ghosal A. Biosensors for epilepsy management: state-of-art and future aspects. Sensors 2019;19(7):1525.
[4] Kaur T, Diwakar A, Mirpuri P, Tripathi M, Chandra PS, Gandhi TK. Artificial intelligence in epilepsy. Neurol India 2021;69(3):560.
[5] Tatum WO, Rubboli G, Kaplan PW, Mirsatari SM, Radhakrishnan K, Gloss D, et al. Clinical utility of EEG in diagnosing and monitoring epilepsy in adults. Clin Neurophysiol 2018;129(5):1056–82.
[6] Wilson SJ, Baxendale S, Barr W, Hamed S, Langfitt J, Samson S, et al. Indications and expectations for neuropsychological assessment in routine epilepsy care: Report of the ILAE Neuropsychology Task Force, Diagnostic Methods Commission, 2013–2017. Epilepsia 2015;56(5):674–81.
[7] Engel J. Seizures and epilepsy, Vol. 83. USA: Oxford University Press; 2013.
[8] Johnston CA, Crawford PM. Anti-epileptic drugs and hormonal treatments. Curr Treat Options Neurol 2014;16:1–9.
[9] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017;58(4):512–21.
[10] Hirsch E, French J, Scheffer IE, Bogacz A, Alsaadi T, Sperling MR, et al. ILAE definition of the idiopathic generalized epilepsy syndromes: position statement by the ILAE task force on nosology and definitions. Epilepsia 2022;63(6):1475–99.
[11] Adeli H, Ghosh-Dastidar S, Dadmehr N. A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. IEEE Trans Biomed Eng 2007;54(2):205–11.
[12] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med 2018;100:270–8.
[13] Khan H, Marcuse L, Fields M, Swann K, Yener B. Focal onset seizure prediction using convolutional networks. IEEE Trans Biomed Eng 2017;65(9):2109–18.
[14] Song X, Aguilar L, Herb A, Yoon SC. In: March. Dynamic modeling and classification of epileptic EEG data. IEEE; 2019. p. 49–52.
[15] Acharya UR, Hagiwara Y, Adeli H. Automated seizure prediction. Epilepsy Behav 2018;88:251–61.
[16] Nogay HS, Adeli H. Detection of epileptic seizure using pretrained deep convolutional neural network and transfer learning. Eur Neurol 2021;83(6):602–14.
[17] Sriraam N, Temel Y, Rao SV, Kubben PL. In: July. A convolutional neural network based framework for classification of seizure types. IEEE; 2019. p. 2547–50.
[18] Raghu S, Sriraam N, Temel Y, Rao SV, Kubben PL. EEG based multi-class seizure type classification using convolutional neural network and transfer learning. Neural Netw 2020;124:202–12.
[19] Asif, U., Roy, S., Tang, J. and Harrer, S., 2020. SeizureNet: Multi-spectral deep feature learning for seizure type classification. In Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology: Third International Workshop, MLCN 2020, and Second International Workshop, RNO-AI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 3 (pp. 77-87). Springer International Publishing.
[20] Priyasad D, Fernando T, Denman S, Sridharan S, Fookes C. Interpretable seizure classification using unprocessed EEG with multi-channel attentive feature fusion. IEEE Sens J 2021;21(17):19186–97.
[21] Ghosh-Dastidar S, Adeli H. Improved spiking neural networks for EEG classification and epilepsy and seizure detection. Integr Comput-Aided Eng 2007;14(3):187–212.
[22] Ghosh-Dastidar S, Adeli H. A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. Neural Netw 2009;22(10):1419–31.
[23] Gu B, Adeli H. Toward automated prediction of sudden unexpected death in epilepsy. Rev Neurosci 2022;27;33(8):877–887. https://doi.org/10.1515/revneuro-2022-0024.
[24] Adeli H, Ghosh-Dastidar S. Automated EEG-based diagnosis of neurological disorders: Inventing the future of neurology. CRC Press; 2010.
[25] Shanir PM, Khan KA, Khan YU, Farooq O, Adeli H. Automatic seizure detection based on morphological features using one-dimensional local binary pattern on long-term EEG. Clin EEG Neurosci 2018;49(5):351–62.
[26] Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci Methods 2003;123(1):69–87.
[27] Sharma P, Khan YU, Farooq O, Tripathi M, Adeli H. A wavelet-statistical features approach for nonconvulsive seizure detection. Clin EEG Neurosci 2014;45(4):274–84.
[28] Faust O, Acharya UR, Adeli H, Adeli A. Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure 2015;26:56–64.
[29] Dwi Saputro, I.R., Maryati, N.D., Solihati, S.R., Wijayanto, I., Hadiyoso, S. and Patmasari, R., 2019, May. Seizure type classification on EEG signal using support vector machine. In Journal of Physics: Conference Series (Vol. 1201, No. 1, p. 012065). IOP Publishing.
[30] Wijayanto I, Hartanto R, Nugroho HA, Winduratna B. In: August. Seizure type detection in epileptic EEG signal using empirical mode decomposition and support vector machine. IEEE; 2019. p. 314–9.
[31] Albaqami H, Hassan G, Datta A. Comparison of WPD, DWT and DTCWT for Multi-Class Seizure Type Classification. In: In 2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE; 2021. p. 1–7.
[32] Tang L, Zhao M, Wu X. Accurate classification of epilepsy seizure types using wavelet packet decomposition and local detrended fluctuation analysis. Electron Lett 2020;56(17):861–3.
[33] Dang N, Shao K, Chen L, Yang M. February. Multi-model decision-making seizure types classification based on transfer learning. In: In 2022 International Symposium on Control Engineering and Robotics (ISCER). IEEE; 2022. p. 192–201.
[34] Liu T, Truong ND, Nikpour A, Zhou L, Kavehei O. Epileptic seizure classification with symmetric and hybrid bilinear models. IEEE J Biomed Health Inform 2020;24(10):2844–51.
[35] Cao X, Yao B, Chen B, Sun W, Tan G. Automatic seizure classification based on domain-invariant deep representation of EEG. Front Neurosci 2021;15:760987.
[36] Bhattacharya A, Baweja T, Karri SPK. Epileptic seizure prediction using deep transformer model. Int J Neural Syst 2022;32(02):2150058.
[37] Liu G, Tian L, Zhou W. Patient-independent seizure detection based on channel-perturbation convolutional neural network and bidirectional long short-term memory. Int J Neural Syst 2022;32(06):2150051.
[38] Zhou J, Liu L, Leng Y, Yang Y, Gao B, Jiang Z, et al. Both cross-patient and patient-specific seizure detection based on self-organizing fuzzy logic. Int J Neural Syst 2022;32(06):2250017.
[39] Yu Z, Albera L, Le Bouquin Jeannes R, Kachenoura A, Karfoul A, Yang C, et al. Epileptic seizure prediction using deep neural networks via transfer learning and multi-feature fusion. Int J Neural Syst 2022;32(07):2250032.
[40] Wang J, Ge X, Shi Y, Sun M, Gong Q, Wang H, et al. Dual-modal information bottleneck network for seizure detection. Int J Neural Syst 2023;33(01):2250061.
[41] Nhu D, Janmohamed M, Shakhatreh L, Gonen O, Perucca P, Gilligan A, et al. Automated Interictal Epileptiform Discharge Detection from Scalp EEG Using Scalable Time-series Classification Approaches. Int J Neural Syst 2023;33(01):2350001.
[42] Peh WY, Thangavel P, Yao Y, Thomas J, Tan YL, Dauwels J. Six-center assessment of CNN-Transformer with belief matching loss for patient-independent seizure detection in EEG. Int J Neural Syst 2023;33(03):2350012.
[43] Zhao Y, He J, Zhu F, Xiao T, Zhang Y, Wang Z, et al. Hybrid Attention Network for Epileptic EEG Classification. Int J Neural Syst 2023;33(06):2350031.
[44] Shah V, Von Weltin E, Lopez S, McHugh JR, Veloso L, Golmohammadi M, et al. The Temple University Hospital Seizure Detection Corpus. Front Neuroinf 2018;12:83.
[45] Li Y, Liu Y, Cui WG, Guo YZ, Huang H, Hu ZY. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. IEEE Trans Neural Syst Rehabil Eng 2020;28(4):782–94.
[46] Harati A, Lopez S, Obeid I, Picone J, Jacobson MP, Tobochnik S. The TUH EEG CORPUS: A big data resource for automated EEG interpretation. In: In *2014 IEEE signal processing in medicine and biology symposium (SPMB)*. IEEE; 2014, December.. p. 1–5.
[47] Golmohammadi, M., Harati Nejad Torbati, A.H., Lopez de Diego, S., Obeid, I. and Picone, J., 2019. Automatic analysis of EEGs using big data and hybrid deep learning architectures. *Frontiers in human neuroscience*, *13*, p.76.
[48] Obeid I, Picone J. The temple university hospital EEG data corpus. Front Neurosci 2016;10:196.
[49] Shankar A, Dandapat S, Barma S. Seizure type classification using EEG based on Gramian angular field transformation and deep learning. In: In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021, November.. p. 3340–3.
[50] Vidyaratne LS, Iftekharuddin KM. Real-time epileptic seizure detection using EEG. IEEE Trans Neural Syst Rehabil Eng 2017;25(11):2146–56.
[51] Samiee K, Kovacs P, Gabbouj M. Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform. IEEE Trans Biomed Eng 2014;62(2):541–52.
[52] Simons FJ, Loris I, Nolet G, Daubechies IC, Voronin S, Judd JS, et al. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. Geophys J Int 2011;187(2):969–88.
[53] Wei X, Zhou L, Chen Z, Zhang L, Zhou Y. Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. BMC Med Inf Decis Making 2018;18(5):71–80.
[54] Gong L, Li M, Zhang T, Chen W. EEG emotion recognition using attention-based convolutional transformer neural network. Biomed Signal Process Control 2023;84:104835.
[55] Li W, Tian Y, Hou B, Dong J, Shao S, Song A. A Bi-Stream hybrid model with MLPBlocks and self-attention mechanism for EEG-based emotion recognition. Biomed Signal Process Control 2023;86:105223.
[56] Liu, L., Liu, J. and Han, J., 2021. Multi-head or single-head? an empirical comparison for transformer training. *arXiv preprint arXiv:2106.09650*.
[57] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing 2021;452:48–62.
[58] Chaudhari S, Mithal V, Polatkan G, Ramanath R. An attentive survey of attention models. ACM Trans Intell Syst Technol (TIST) 2021;12(5):1–32.
[59] Ganea, O., Bécigneul, G. and Hofmann, T., 2018. Hyperbolic neural networks. *Advances in neural information processing systems*, *31*.

[60] Deng Z, Li C, Song R, Liu X, Qian R, Chen X. EEG-based seizure prediction via hybrid vision transformer and data uncertainty learning. Eng Appl Artif Intel 2023;123:106401.
[61] Gao Z, Xie J, Wang Q, Li P. Global second-order pooling convolutional networks. In: In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*; 2019. p. 3024–33.
[62] Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Machine Learn Technol 2020;2(1):37–63.
[63] Chinchor N. In: The statistical significance of the MUC-4 results. Association for Computational Linguistics; 1992. p. 30–50.
[64] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45(4):427–37.
[65] Jia G, Lam HK, Althoefer K. Variable weight algorithm for convolutional neural networks and its applications to classification of seizure phases and types. Pattern Recogn 2022;121:108226.
[66] Hussain SF, Qaisar SM. Epileptic seizure classification using level-crossing EEG sampling and ensemble of sub-problems classifier. Expert Syst Appl 2022;191:116356.
[67] Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2023;13(2):e1484.
[68] Bonissone PP, Subbu R, Eklund N, Kiehl TR. Evolutionary algorithms+ domain knowledge= real-world evolutionary computation. IEEE Trans Evol Comput 2006;10(3):256–80.
[69] Gu P, Wu T, Zou M, Pan Y, Guo J, Xiahou J, et al. Multi-Head Self-Attention Model for Classification of Temporal Lobe Epilepsy Subtypes. Front Physiol 2020;11:604764.
[70] Zheng Q, Yang M, Yang J, Zhang Q, Zhang X. Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process. IEEE Access 2018;6:15844–69.
[71] Liang, H., Zhang, S., Sun, J., He, X., Huang, W., Zhuang, K. and Li, Z., 2019. Darts+: Improved differentiable architecture search with early stopping. *arXiv preprint arXiv:1909.06035*.
[72] Craley J, Jouny C, Johnson E, Hsu D, Ahmed R, Venkataraman A. Automated seizure activity tracking and onset zone localization from scalp EEG using deep neural networks. PLoS One 2022;17(2):e0264537.
[73] Golmohammadi M, Ziyabari S, Shah V, Obeid I, Picone J. In: December. Deep architectures for spatio-temporal modeling: Automated seizure detection in scalp EEGs. IEEE; 2018. p. 745–50.
[74] Saqib M, Zhu Y, Wang M, Beaulieu-Jones B. In: July. Regularization of deep neural networks for EEG seizure detection to mitigate overfitting. IEEE; 2020. p. 664–73.
[75] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.
[76] Selvaraj A, Pj S, Thomas J, Rajamanickam Y, Menon RN, Agastinose Ronickom JF. In: Multi-Class Seizure Type Classification Using Features Extracted from the EEG. IOS Press; 2023. p. 68–71.
[77] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. J Neural Eng 2018;15(3):031005.
[78] Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 2018;106:249–59.
[79] Cui J, Lan Z, Sourina O, Müller-Wittig W. EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network. IEEE Transactions on Neural Networks and Learning Systems. 2022.
[80] Rafiei MH, Adeli H. A new neural dynamic classification algorithm. IEEE Trans Neural Networks Learn Syst 2017;28(12):3074–83.
[81] Alam KMR, Siddique N, Adeli H. A dynamic ensemble learning algorithm for neural networks. Neural Comput & Applic 2020;32:8675–90.
[82] Pereira DR, Piteri MA, Souza AN, Papa JP, Adeli H. FEMa: A finite element machine for fast learning. Neural Comput & Applic 2020;32:6393–404.
[83] Rafiei MH, Gauthier LV, Adeli H, Takabi D. Self-supervised learning for electroencephalography. IEEE Transactions on Neural Networks and Learning Systems. 2022.