



Reducing False Alarms in Seizure Prediction: A Specialized CNN Architecture and a Novel EEG Sampling Technique

Ameer Hayder Drbeeni¹, Haider Ismael Shahadi¹ and Arwa Hameed Mohammed Taqi¹

¹College of Engineering, University of Kerbala, Karbala, Iraq

Received 2 July 2024, Revised 16 January 2025, Accepted 29 January 2025

Abstract: Deep learning played a vital role in the seizure prediction challenge. However, most studies used generic architectures that fail to consider the distinct characteristics of multivariate time-series Electroencephalography signals. Additionally, many suggested methods depend on inadequate EEG segmentation techniques, resulting in unreliable results. This study presents an in-depth architectural design of a Convolutional Neural Network specifically tailored to extract features from wavelet-transformed EEG signals using Wavelet packet decomposition (WPD). In addition, the chosen testing strategy and data segmentation methodology ensures accurate, trustworthy, and reproducible performance results. This study introduces a data segmentation method to generate distinct intervals effectively capturing more temporal dynamics of the time-series data. The proposed model evaluation utilized 12 subjects' EEG data from the CHB-MIT dataset in a subject-specific manner, employing a Leave-One-Out cross-validation technique. The proposed architecture outperformed five reproduced state-of-the-art CNN models in the segment-based evaluation metrics. The proposed model achieved 78.00% accuracy, 65.17% sensitivity, and a high 90.83% specificity rate. Evaluation using the more straightforward KFold cross-validation technique demonstrated robust performance, achieving 96.68% accuracy, 97.41% sensitivity, and 95.95% specificity. The significant improvement in the model's specificity rates indicates a substantial reduction in false alarms, making the proposed model a reliable tool for seizure prediction.

Keywords: Epilepsy, Seizure Prediction, Electroencephalography, Convolutional Neural Network, Wavelet Packet Decomposition

1. INTRODUCTION

The brain is the central organ of the nervous system, orchestrating our thoughts, emotions, and actions. The phrase "I think, therefore I am" is the first and most popular principle in the philosophy of René Descartes. This quote underscores the very essence of human existence, emphasizing the importance of the brain in shaping our reality. One's mental health and physical well-being are deeply tied to the brain's health. Any deviation from the brain's natural functioning can heavily impact what life we can have.

Epilepsy is a brain illness characterized by unprovoked, recurrent, and spontaneous seizure events that disrupt normal brain functioning and affect patients' lives. The unpredictable nature of seizures can drastically degrade the quality of life and can be life-threatening directly or indirectly owing to drowning and other accidents [1]. The World Health Organization (WHO) reported that Epilepsy has affected more than 50 million individuals around the world. The burden of Epilepsy is compounded by the fact

that 30% of individuals exhibit resistance to anti-epileptic drugs [2]. This burden highlights the imperative need for innovative strategies to predict, detect, and prevent seizures [3].

Electroencephalography (EEG) is a commonly used tool for monitoring the brain's electric activity and studying Epilepsy [4]. The states of an epileptic brain can be one of four: inter-ictal (normal interval between seizure events), pre-ictal (interval immediately preceding a seizure event), ictal (period of a seizure), and post-ictal (interval immediately following a seizure). At its core, seizure prediction is a task of detecting the presence of the pre-ictal state. However, most studies depict seizure prediction as a binary classification task between the pre-ictal and inter-ictal states [5].

Seizure prediction utilizing the electric activity of an epileptic brain is a challenging research problem that has attracted more attention in recent years [6], [7], [8], [5]. The ability to issue warnings before seizures onset may lead to developing novel diagnostic methods and therapies [9],

[10]. Moreover, the focus on the algorithmic prediction of seizures might offer valuable insights into the mechanics of a seizure event. However, little to no attention has been paid to extracting meaningful and comprehensible characteristics about seizures and their dynamics [11].

EEG signals are complex, non-stationary, and require advanced signal processing and feature extraction to help provide meaningful interpretation [12]. Advances in deep learning have shown great promise in its capacity to learn good representations of EEG signals. Deep learning differs from traditional machine learning; it does not require extensive domain expertise or operator feedback on the feature extraction process. Nevertheless, deep learning models are prone to weak generalization and deliver less human comprehensible features [13].

Deep learning approaches offer a potential answer to developing a reliable seizure prediction system. However, the clinical adoption of deep learning models mandates a certain degree of model explainability. Although deep learning is considered a black box, it is possible to build human-comprehensible interpretations on the internals of deep learning models [14], [11]. Nevertheless, most studies utilized deep learning methods with generic architectural designs solely for accurate predictions. Such design choices greatly complicated the interpretation of model decisions. Another concern involving numerous studies is the potential for data bias in the data pre-processing and model evaluation techniques [15].

The following points highlight the key contributions of this study:

- A novel EEG time-series data framing and sampling technique to address the highly imbalanced class ratio of the two epileptic states. The proposed technique can also reduce the information loss due to the sampling process and maintain as much temporal span as possible. Furthermore, the sampling process is proven to be deterministic and avoids the uncertainty of other random sampling techniques.
 - A wavelet transform that can reveal the frequency (spectral) information of the EEG data without causing redundancy or increasing the dimensionality of the data. The model's compactness relies on exploiting the Wavelet Packet Decomposition (WPD) to enhance the data without increasing the dimensionality or redundancy.
 - A specialized Convolutional Neural Network (CNN) architecture designed to deal with the nature of multi-channel EEG time-series data. The design also accounts for compactness, simplicity, and easier understanding. The CNN architecture utilizes separable convolutions to reduce the number of learned parameters.
- Reproducing five state-of-the-art CNN-based models and their associated pre-processing pipelines from recent seizure prediction studies. The motive behind this objective is to contribute a head-to-head unbiased comparison using the same conditions. Furthermore, evaluating the proposed architecture performance employed multiple test sets generated by the Leave-One-Out Cross-Validation, also named Leave-One-Group-Out (LOGO).

The remainder of this manuscript is organized as follows: Section 2 gives a concise overview of recent state-of-the-art methods. Section 3 describes the data and materials used in this research, including the pre-processing techniques. Additionally, Section 3 lists and details the components of the proposed architecture, and presents the model training strategy and evaluation metrics. Section 4 delivers the performance results, while Section 5 thoroughly examines the comparisons of the reproduced methods. Section 6 encloses the manuscript with the study conclusions and discusses potential future works.

2. RELATED WORKS

The literature on seizure prediction has explored numerous manual feature extraction techniques [16], [17], [18]. However, the findings revealed no success in identifying the ideal set of features and the appropriate classifier to yield reliable prediction performance. As a concrete example, the winning study of the Melbourne University seizure prediction contest employed 11 classifiers and more than 3000 handcrafted features [19]. Thus, recent state-of-the-art methods utilized automated feature extraction, taking advantage of deep learning methods.

Tsiouris et al. [18] used the most common features, including the time domain, frequency domain, correlation, and graph theory features, to generate (643×1) feature vectors for each 5-second long EEG segment. The Long Short-Term Memory (LSTM) network was employed as the classifier for its inherent advantage in processing the sequential EEG time series. They also tested the LSTM network using raw EEG segments as the input in a feature extraction scenario. However, the feature extraction model was unable to achieve comparable results. The primary cause for this low performance observed while using raw EEG data was the model's low capacity and limited computing hardware capability, which prevented any attempt to increase the model size.

Khan et al. [20] designed a CNN architecture to extract features from the time-frequency domain of the EEG signals. The time-frequency domain was generated using the Continuous Wavelet Transform (CWT) of 1-second raw EEG segments. They argued that using wavelet-transformed EEG segments to include the frequency information achieves deeper feature extraction without additional model capacity. They used the Mexican-hat mother wavelet and nine dyadic scales to add the geometric frequency range

of 0.25 to 128 Hz in the frequency dimension. On the other hand, Truong et al. [21] operated the Short Time Fourier Transform (STFT) to transform the 30-second raw EEG segments into the time-frequency domain. The STFT generated segments that encompassed frequencies ranging from 1 to 128 Hz at the cost of losing information in the time dimension. The information loss occurs since STFT uses sliding windows of fixed lengths to generate the amplitude or power spectra.

Daoud and Bayoumi [22] conducted experiments using various deep learning architectures trained on raw EEG data without pre-processing except for the 5-second segmentation process. Their best-performing architecture with the least number of trainable parameters is a Convolutional Auto-encoder (CAE) for feature extraction and a Bidirectional LSTM (Bi-LSTM) for classification. The CAE architecture trains without supervision using data from several subjects to find lower-dimensional representations of the raw EEG data. The Bi-LSTM uses the learned features from the encoder component of the CAE to carry out the ultimate classification task.

Similarly, Xu et al. [23], and Jana and Mukherjee [24] introduced end-to-end deep learning models that directly learn from raw EEG data as input. Jana and Mukherjee [24] used 8-second long EEG segments and developed a CNN architecture with (3×3) convolution kernel size and a combination of (1×3) and (2×3) max-pooling kernel sizes. Different max-pooling kernel sizes are used because the input EEG segments have more temporal than spatial resolution. On the other hand, Xu et al. [23] used 1D convolution kernels having sizes of (1×20) and (1×10) for the first three blocks. Each of these blocks are followed by (1×10) and (1×5) 1D max-pooling operations. The extended kernel lengths can capture longer local temporal dependencies with an increased receptive field. The final two convolutional blocks have standard 2D convolution kernels with (3×3) size to capture spatial and temporal characteristics. Xu et al. [23] configured the EEG segmenting window size to be 20 seconds long to cover the extended temporal down-sampling.

An alternative method for expanding the receptive field of the convolution process involves using dilated kernels. Hussein et al. [25], Wang et al. [26], and Gao et al. [27] incorporated a dilation rate larger than one to enlarge the kernel length while avoiding increasing the number of weights within a kernel. Hussein et al. [25] divided the EEG data into segments, each having 30 seconds in length. Then, each segment is transformed into time-frequency representations using CWT with 100 scales. However, the wavelet mother and the distribution of the scales were not reported in this study. The illustrated time-frequency images indicated that the utilized scales produce a frequency range of 0.1 to 100 Hz. The suggested model consisted of 30 convolution layers arranged in parallel paths with varying dilation rates and a kernel size of (3×3) or (5×5) . While

using 30 convolution layers, the network is constructed to be only three convolution blocks deep. The suggested CNN architecture was created with numerous learnable parameters, requiring several hours of training for a single patient. Wang et al. [26] used STFT instead of CWT to transform the 30-second-long EEG segments. Additionally, they used 3D convolution kernels instead of 2D but did not report the number of kernels in each layer.

Lawhern et al. [28] constructed an EEG-specific architecture, called EEGNet, replicating the well-known Filter-Bank Common Spatial Pattern (FBCSP) feature extraction process. FBCSP performs separate spatial filtering on several frequency sub-bands to compute features with maximal variance to optimally discriminate between two classes [29]. Zhang et al. [30] utilized FBCSP to generate 2D images and reached cutting-edge performance using a simple CNN architecture. In contrast, EEGNet utilized raw EEG data to learn both temporal (frequency-related) and spatial features using a CNN directly instead of FBCSP. The EEGNet architecture incorporates a 1D temporal convolution layer with a kernel size of (1×64) , followed by spatial depthwise convolution layers with a kernel size of $(N \times 1)$ where N is the number of channels in the input EEG signal.

However, the EEGNet architecture designed by Lawhern et al. [28] was not specifically trained for a seizure prediction task. Jemal et al. [31] made minor adjustments to the EEGNet architecture and trained the model to predict seizures from 20-second-long raw EEG segments. The temporal kernel size is (1×128) to extract frequency features starting at 2 Hz from an EEG segment with a 256 Hz sampling rate. The architecture additionally included average-pooling layers having a 1D kernel length of (1×16) following both temporal convolution layers to enlarge the receptive field of the model. However, this study failed to report the number of kernels utilized in each stage of the proposed architecture.

Understanding is key when designing a model, even for a complex task. Otherwise, the internals of the deep learning model remain deeply vague, preventing the comprehensibility of the recognized patterns. Still, recent studies tend to develop complex architecture exclusively to gain more accuracy. Lu et al. [32] and Xu et al. [33] are example studies that yielded impressive results but raised the concern of interpretability and reproducibility of their models. Lu et al. [32] integrated the attention mechanism into a CNN architecture comprising six convolution layers. What elevated the complexity of the model is that they made their CNN architecture recursive by implementing the Bi-Directional LSTM (Bi-LSTM) as the final classifier.

The motive of this study was originally to explore whether simpler architectural designs can achieve similar or better performance compared to more complex ones. We believe that the seizure prediction field must strike a balance between accuracy and other practical considerations such

as efficiency and explanation. The EEGNet architecture, with its 1D convolution processes, inspired us to build a more robust architecture yet holds the same compactness and simplicity.

3. MATERIALS AND METHODS

A. Dataset

The proposed CNN architecture is trained using a subset of the CHB-MIT scalp EEG dataset gathered at the Children's Hospital Boston [34], [35]. The dataset comprises 664 recordings, grouped into 23 cases from 22 pediatric patients. Each case is denoted by the code chb_i , where i refers to the case ID. Each case has a summary *.txt* file that specifies the start and end timings, the EEG montage used for each recording, the number of seizure events, and their timing relative to the recording. Each case consists of recordings ranging from 9 to 42. Each recording has 10 minutes up to 4 hours of EEG data duration sampled at 256 Hz. This dataset utilized the international 10-20 EEG electrode placement system with bipolar montage. The proposed method uses only 18 channels common among subjects from the overall 23 EEG channels. The selected EEG channels and their corresponding electrode placements are depicted in Figure 1.

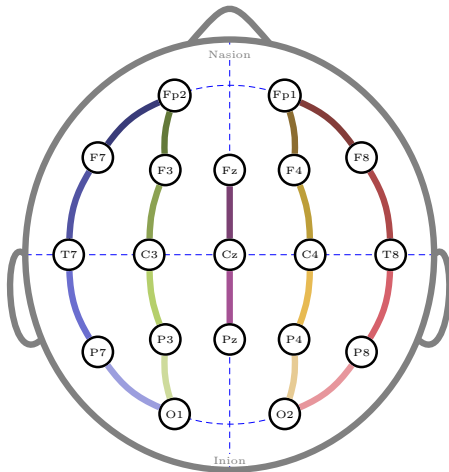


Figure 1. The montage of the 18 common EEG channels among the CHB-MIT subjects.

B. Data Segmentation

The proposed seizure prediction system employs the pre-ictal and inter-ictal states to perform a binary classification task. Performing binary classification on long continuous EEG data mandates establishing three parameters to extract the EEG samples (frames) and label them as one of the epileptic states:

- The pre-ictal period, which is immediately before a seizure. For this investigation, a 35-minute pre-ictal period length is extracted. Furthermore, seizure events with 30 minutes or less pre-ictal data are excluded to

ensure non-misleading validation results when testing on a limited amount of EEG samples.

- The inter-ictal distance refers to the duration between a seizure onset and the inter-ictal data. This period is excluded from the inter-ictal span to reduce the similarity between pre-ictal and inter-ictal states in the training phase. The proposed system uses two hours as the inter-ictal period distance, resulting in an hour and 25 minutes gap between the pre-ictal and inter-ictal periods.
- The post-ictal period is the interval immediately after a seizure that is not categorized as a pre-ictal or inter-ictal state. In this study, two hours of post-ictal period are excluded from the training set.

The inter-ictal distance ensures a sufficient temporal gap between the inter-ictal and pre-ictal states. Thus, the inter-class variance is uplifted, allowing for better temporal separation between these epileptic states. The motive of this procedure is due to the indefinite and subject-specific shift from the inter-ictal to the pre-ictal state. In conclusion, a clear temporal separation between the two states can enhance the model training [10]. Figure 2 shows a 150-minute EEG sample from subject *chb01* preceding the seizure of *chb01_15.edf* file.

The intervention time, also known as the Seizure Prediction Horizon (SPH) [10], is another essential parameter required to specify the timings of extracted periods. However, this parameter is only crucial in the event-based prediction (inference stage). For instance, a five-minute intervention time indicates that alarms induced in the five minutes immediately before a seizure are false alarms. Nevertheless, the intervention period is still considered pre-ictal; any alarm during this interval is a false warning since there should be enough time to intervene [10], [8].

A 35-minute pre-ictal duration to predict incoming seizures 30 minutes before the intervention interval yields a 30-minute Seizure Occurrence Period (SOP). A Seizure event must be preceded by at least a 2-hour (post-ictal) and 35 minutes (pre-ictal) of EEG data from the previous seizure termination. Otherwise, seizure events may overlap, and the pre-ictal period is short or non-existent. Only cases with three seizures or more that meet the aforementioned conditions are used to evaluate the proposed and replicated architectures.

Table I displays the final subset of EEG data obtained from the CHB-MIT dataset. The table presents the number of seizure events, the overall length of the EEG data, the interictal and preictal data length, and the imbalance ratio between the two classes. The trainable seizures are the number of seizures used in the training process. As shown in Table I, the inter-ictal data are often long, and the pre-ictal data are limited by the pre-ictal length and number of seizure events. The number of inter-ictal EEG samples

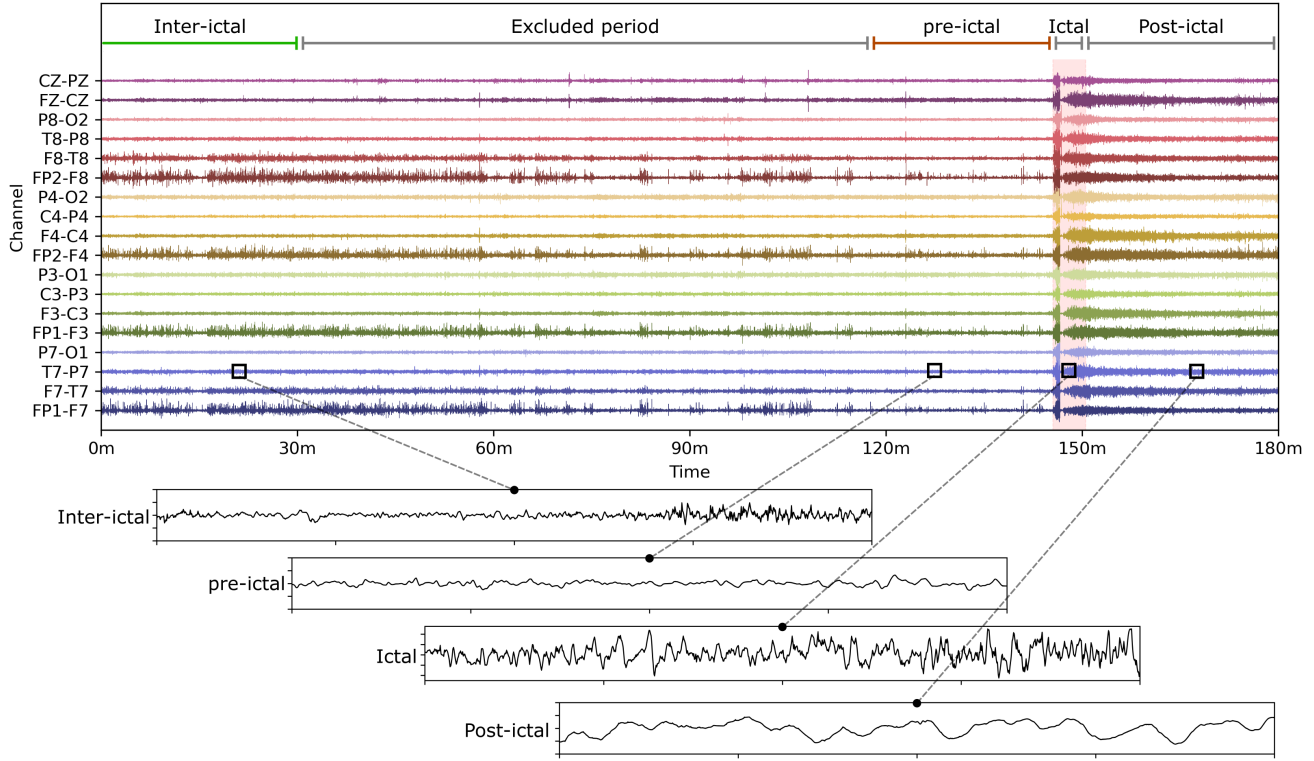


Figure 2. A 150 minutes of EEG data preceding a seizure event from subject *chb01*.

outnumbering the pre-ictal EEG samples causes data imbalance. To solve the imbalance issue, most studies employed random under-sampling techniques [20], [36], [22], [37], [38], [23], [24], [27]. Consequently, the certainty of these methods is prone to data bias due to randomly selecting a subset of the inter-ictal data. This study introduces a sampling technique that is formulated as below:

Let the subset of all extracted periods and their annotations from a single case be:

$$S = \{(x_{t,g}, y_{t,g}) | t = 0, 1, \dots, N_g - L, g = 0, 1, \dots, G\} \quad (1)$$

where $x_{t,g}$ is the EEG segment starting at time t and ending at time $t + L$ in group g , L is the selected window size of the EEG segment. $y_{t,g} \in \{0, 1\}$ is the binary class label of the corresponding EEG segment. N_g is the total number of EEG time points in group g . G is the total number of groups, where each group represents data associated with a seizure event. While seizure events that met the selection criteria have pre-ictal data of no less than 30 minutes, the seizure events is not necessarily preceded by inter-ictal data. Therefore, inter-ictal recordings are evenly divided among seizure events. In this study, the segmentation window size (L) is set to 4 seconds. The window size of 4 seconds resulted in a slightly better accuracy than that of 1, 2, and 8 seconds. Now let the majority class (inter-ictal class) and the minority class (pre-ictal class) be represented as follows:

$$S_{0,g} = \{(x_{t,g}, y_{t,g}) | y_{t,g} = 0\} \quad (2)$$

$$S_{1,g} = \{(x_{t,g}, y_{t,g}) | y_{t,g} = 1\} \quad (3)$$

where $S_{0,g}$ is the inter-ictal set and $S_{1,g}$ is the pre-ictal set. The complete set S represents all unique segments by hopping only one sample at a time. Truong et al. [21] employed a variable overlapping factor through a different hopping size for every subject. Using a large overlapping percentage increases the number of samples in the minority class. However, it does not add new information to the dataset and increases the training's computational cost. Thus, this study used a fixed overlap of 50%.

$$S'_{1,g} = \{(x_{t,g}, y_{t,g}) | y_{t,g} = 1, t = 0, n, 2n, \dots, N_g - L\} \quad (4)$$

where $n = 0.5 * L$ is an integer defining the hop size and equals half the segment size. In order to under-sample the majority class, we used a variable under-lapping factor. In contrast to random sampling, this strategy ensures that the chosen subset of the majority class contains as much temporal information as possible. The under-lapping hop size calculation is as follows:

$$k = \lfloor \frac{S_{0,g}}{S'_{1,g}} \rfloor \quad (5)$$

Where $\lfloor \cdot \rfloor$ is the floor function. The final under-sampled majority class is:

$$S'_{0,g} = \{(x_{t,g}, y_{t,g}) | y_{t,g} = 0, t = 0, k, 2k, \dots, N_g - L\} \quad (6)$$



TABLE I. Summary of the CHB-MIT dataset.

Subject	Num. of Seizures (Trainable)	Overall Length (minutes)	Interictal (minutes)	Preictal (minutes)	Imbalance (inter:pre)
chb01	7 (3)	2433	1296	104	12.5
chb02	3 (2)	2115	1564	69	22.7
chb03	7 (1)	2280	1643	33	49.8
chb04	4 (2)	9363	8509	70	121.6
chb05	5 (4)	2340	1233	139	8.9
chb06	10 (6)	4004	2363	209	11.3
chb07	3 (3)	4023	3437	105	32.7
chb08	5 (4)	1200	313	139	2.3
chb09	4 (3)	4072	3360	105	32.0
chb10	7 (5)	3001	1896	175	10.8
chb11	3 (1)	2087	1919	35	54.8
chb12	27 (3)	1241	73	100	0.7
chb13	12 (3)	1980	1202	104	11.6
chb14	8 (4)	1560	584	135	4.3
chb15	20 (8)	2400	428	279	1.5
chb16	10 (2)	1140	458	69	6.6
chb17	3 (2)	1260	907	70	13.0
chb18	6 (3)	2138	1738	104	16.7
chb19	3 (2)	1795	1620	70	23.1
chb20	8 (1)	1656	1202	34	35.4
chb21	4 (2)	1969	1465	68	21.5
chb22	3 (2)	1860	1258	70	18.0
chb23	7 (1)	1593	853	35	24.4

C. Wavelet Packet Decomposition

Wavelet-based and other frequency-time transformations can effectively analyze the non-stationary EEG signals [17], [39]. Both Wavelet Packet Decomposition (WPD) and Discrete Wavelet Transform (DWT) decompose the input signal into a pair of coefficients that reflect the low-frequency (approximation) and high-frequency (detail) components of the signal [40]. The distinction is that DWT uses only the approximation to generate the next-level approximation and detail decompositions. In contrast, WPD decomposes both previous level's approximation and detail components into four new components. Hence, DWT generates $(l + 1)$ coefficients whereas WPD produces (2^l) coefficients.

DWT and WPD transformations do not add to the redundancy of the input signal due to the down-sampling process at each level. However, WPD coefficients have equal lengths due to undergoing the same number of down-sampling operations. Conversely, DWT has different time points between the approximation and detail coefficients because of the multi-level process. Hence, the equal-length coefficients generated by WPD are more appropriate for input to a CNN architecture. Figure 3 illustrates a 3-level WPD tree on a single channel of an EEG segment.

Let the chosen segment length be $L = 4$ seconds, where the number of time points in the segment would be $L \times f_s = 4 \times 256 = 1024$ time points. Applying a 4-level WPD using *Daubechies 4 (db4)* wavelet to the segment gives:

$$wpd : \mathbb{R}^{18 \times 1024} \rightarrow \mathbb{R}^{16 \times 18 \times 70}$$

where the number of generated coefficients for each channel is $2^4 = 16$. The down-sampling process at each level reduces the number of time points to 70. Note that $1024/2^4 =$

$64 \neq 70$ because the *db4* filter has a length of 8 points, causing the addition of 7 points to each level in the filtering process. Therefore, the down-sampling process of 4-level WPD yields the following sequence of coefficient lengths for each level:

$$1024 \rightarrow \frac{1024 + 7}{2} = 515 \rightarrow 261 \rightarrow 134 \rightarrow 70$$

The proposed seizure prediction model employs the WPD technique as the pre-processing pipeline. The WPD pipeline supplies the following feature extraction and classification processes with frequency information-rich EEG samples. WPD, with its frequency sub-bands separating capability, enhances the EEG data representation. Eventually, the proposed CNN architecture can extract deeper features with fewer parameters and less training time.

D. The Proposed CNN Architecture

Recently, one-dimensional (1D) CNN architectures have demonstrated exceptional performance in various signal processing applications, including the analysis of uni-variate signals such as ECG [41] or multivariate data such as EEG [28]. A significant advantage is the simplified interpretation of features extracted using 1D kernels on time-series data. For instance, employing a 2D kernel in an image classification task can be visually explained using basic geometry concepts such as edges and diagonal lines [42].

Similarly, 1D kernels can be interpreted in terms of temporal patterns or frequency information when utilized to extract features from a time series. Additionally, a 1D convolution along the channel axis of a multi-channel signal can indicate the channel that contributes the most to the extracted feature map. Hence, the developed CNN

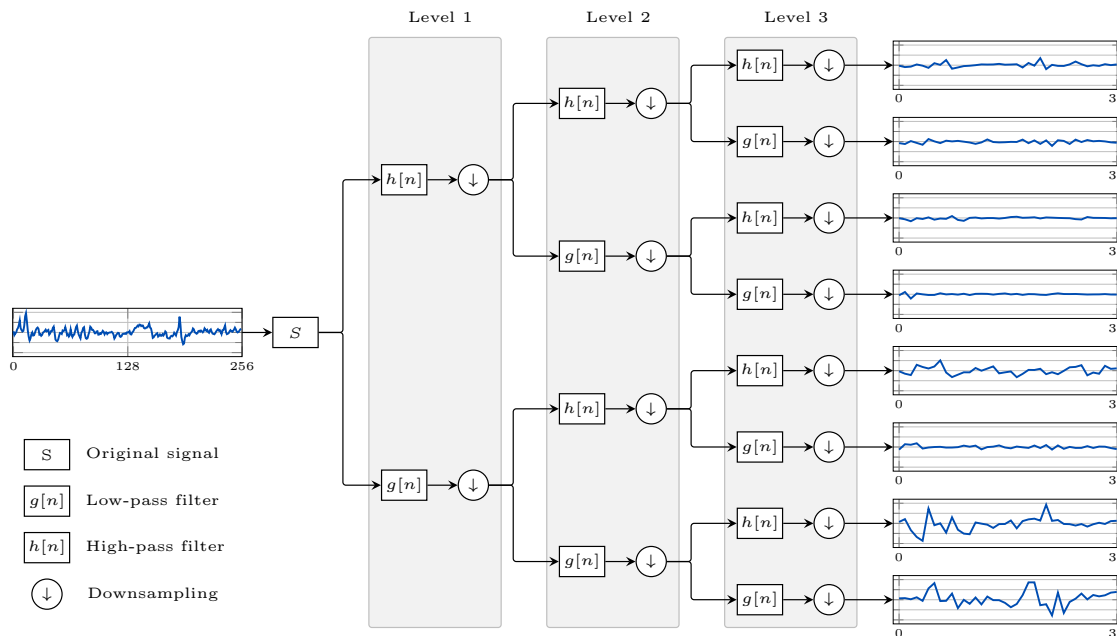


Figure 3. An example of 3-level WPD tree on a single channel of an EEG segment.

architecture comprises only 1D kernels using separable and depthwise convolutions. The proposed CNN architecture accepts as input a 3D matrix of the shape $(D \times C \times T)$, where the first dimension (D) represents the stacking of wavelet coefficients, (C) denotes the stacked EEG channels, and T denotes the number of points in the time domain. Figure 4a shows an example of a 6-channel 4-coefficient sample.

1) Spectral Convolutional Branch

The first convolution block, termed *Spectral Convolution Block*, applies a 1D kernel along the wavelet nodes of each channel separately, without interference between different channels. Each kernel generates different weighted combinations of the wavelet approximation and details coefficients. The spectral convolution process has a kernel size of $(D \times 1 \times 1)$ with no padding. Therefore, the output of the spectral convolution block has the shape of $(C \times T \times F)$, where F denotes the number of feature maps (number of filters). Figure 4b visually explain the process of the spectral convolution block.

The second layer, named *Temporal Convolution*, performs a separable convolution on the generated feature maps to extract deeper frequency-related information. The separable convolution comprises a 1D kernel that is one second long to extract frequency details at 1 Hz and above. A level 5 WPD generates an approximation coefficient with a bandwidth of 1 Hz up to $128/2^4 = 4$ Hz. However, the first convolution block aggregate wavelet details with different band-pass frequencies to include the frequency information from the overall 1-128 Hz bandwidth. In this sense, the temporal kernel receptive field spans a duration of 1 second and can extract frequency information of the range (1-128 Hz) while consisting of only (16) samples.

The second part of the temporal convolution performs a pointwise convolution to aggregate the extracted temporal features from different wavelet coefficient combinations and generate the final feature maps.

Until now, the network has not included any convolution across individual channels to extract the spatial information. The depthwise convolution block applies a 1D kernel along the channel dimension C left after the spectral convolution block. In the depthwise convolution, the network uses a 1D kernel of size $(C, 1)$ that only slides along the temporal dimension T . The depthwise layer learns a frequency-specific spatial filter for each feature map produced by the spectral convolution block. The depthwise convolution block has a parameter called *depth multiplier*. The depth multiplier controls how many channels the depthwise convolution produces. We set the depth multiplier to 4 to learn four spatial filters for each feature map separately. An average pooling layer of kernel size $(1 \times 1 \times 2)$ follows the depthwise convolution to reduce the temporal dimension further.

2) Spatial Convolution Branch

The proposed architecture has a second branch that starts with a channel-wise convolution. The channel-wise convolution block, named *Spatial Convolution Block*, performs a 1D convolution operation across EEG channels, ensuring no interference between frequency bands. The interference between wavelet coefficients is prevented by applying a 1D kernel with a $(1, C, 1)$. The spatial convolution block learns a spatial filter across different frequency bands to generate feature maps of shape $(D \times T \times F)$. The spatial convolution branch consists of layers similar to the spectral convolution branch layers, except that the depthwise convolution has a kernel size of $(D, 1)$ in the spatial branch. The second

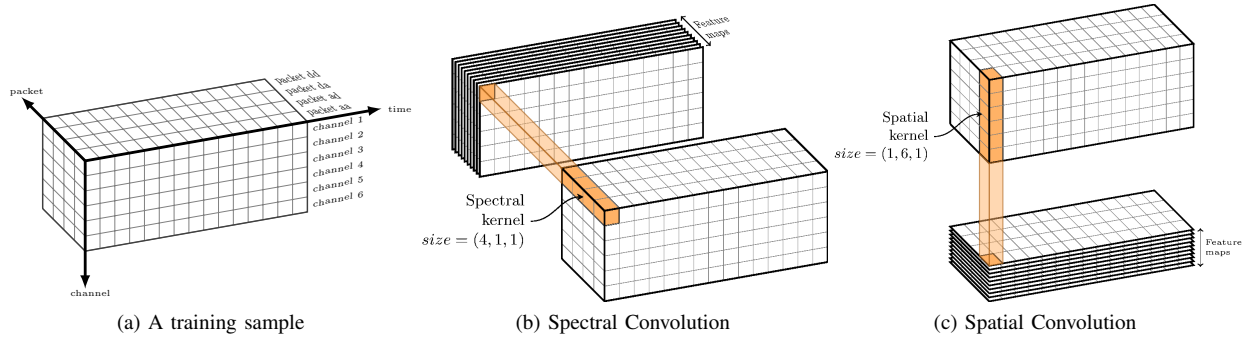


Figure 4. An example of a 6-channel 4-coefficient sample to illustrate the process of (a) The structure of a single training sample, (b) The spectral convolution block, where the convolving kernel slides along the channel and temporal dimension to account for the same wavelet coefficients across individual channels. (c) The spatial convolution block, where the kernel slides over individual channels. The packet (aa), (ad), (da), and (dd) represent the equally divided low-to-high frequency sub-bands of 4-level WPD, respectively.

branch is concatenated with the first branch to form the final feature map. Figure 4c explains the process of the spatial convolution block.

Ultimately, the network applies a 1D temporal convolution to allow the accumulation of concepts learned across the temporal dimension with a kernel half the size of the previous temporal kernel. The final feature map then passes through two dense layers with units equal to 32 and 1, respectively. The network’s output is derived using a *sigmoid* activation function, resulting in a single scalar value representing the predicted class probability. Figure 5 presents complete configuration details of utilized layers of the proposed CNN architecture developed in this study.

E. Training Strategy

The proposed CNN architecture is trained in a subject-specific scheme. The EEG recordings are z-score normalized before the segmentation process. The EEG data of a subject is divided into groups, as discussed in section 3-B, where each group involves the data related to a seizure event. The designed CNN architecture was trained and tested using the Leave-One-Group-Out (LOGO) cross-validation. In this testing strategy, the EEG data of one group is used as a test set, leaving the rest of the groups as a training set. If the subject has N seizure events, N models are trained using $(N - 1)$ seizure events data and tested on the remaining seizure event data. Ultimately, the LOGO cross-validation for the test data split provides robust evaluations of the model generalization to unseen data. Figure 6 illustrates the LOGO test data split method. On the contrary, the K-Fold cross-validation divides the whole set into K groups randomly. The randomization in the split process causes the model to be evaluated on a mixture of EEG samples from different seizure events.

The training set is additionally split into training and validation sets. While most studies employed a randomly chosen holdout validation set, we used an ordered selection of the validation set. The validation set uses every fifth segment within a class from the training set to construct

the 20% validation set. Compared to the random selection of validation samples, the proposed split method ensures that the training and validation sets hold samples with similar temporal characteristics. Additionally, the ordered selection delivers deterministic results, and the training process becomes repeatable.

The holdout validation set prevents the model from overfitting using an early stopping condition. The parameter *validation patience* is configured to 25, which halts the training if the validation loss shows no improvement after 25 epochs. The model training utilizes an Adam optimizer with a 0.001 learning rate and a 512 batch size. The model is trained for 500 epochs or until violating the early stopping criteria. Additionally, the model with the best parameters before halting the training process was restored to provide the evaluations on the test set. Given that data labels are binary, the training phase employs the binary cross-entropy loss term. This study discusses only the segment-based results, as the event-based results can be easily derived by applying any post-processing method. The accuracy, sensitivity, and specificity metrics were used to assess the model’s performance compared to the replicated models.

4. RESULTS

The designed CNN architecture and the developed EEG segmentation and sampling method were tested using EEG data of 67 seizure events from 12 patients subset of the CHB-MIT dataset. We compared the designed CNN architecture against five models replicated from the studies of Khan et al. [20], Truong et al. [21], Lawhern et al. [28], Xu et al. [23], and Jemal et al. [31]. We closely replicated the same architectures, their pre-processing pipelines, and the related hyperparameters as the original studies. Nevertheless, there are still some differences that were not taken into account. For instance, Khan et al. [20] incorporated the ictal class to train the model in a multi-class classification task.

Our replication process is conducted to deliver a fair comparison using the same data segmentation and cross-

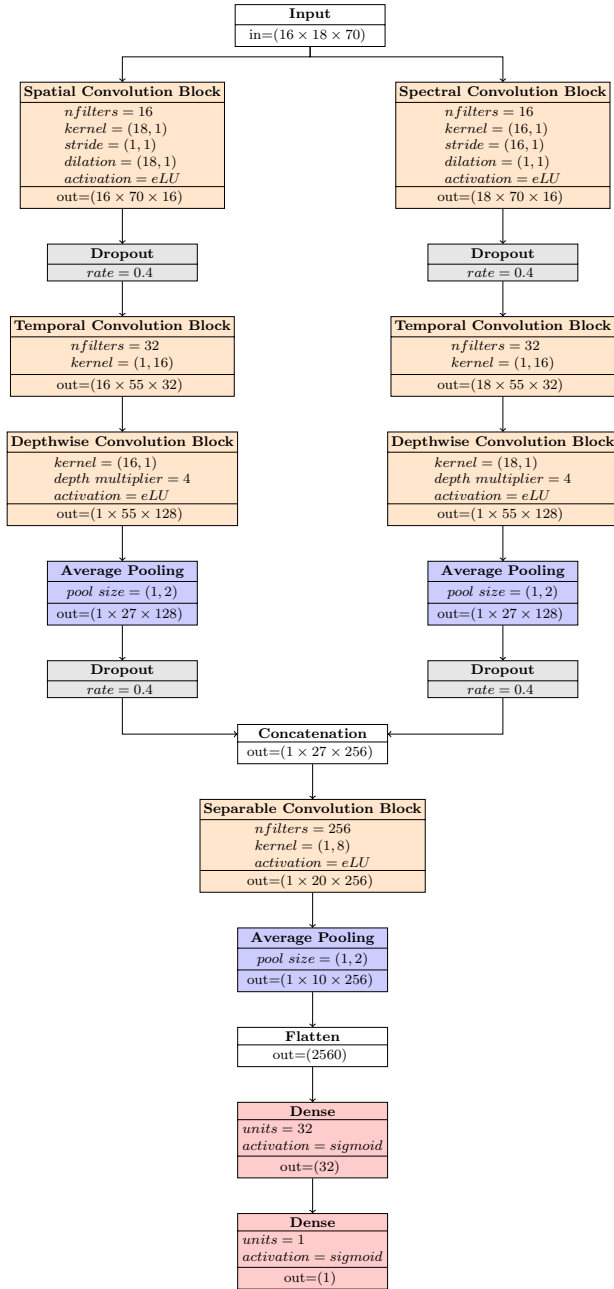


Figure 5. The proposed CNN architecture.

validation techniques presented in this study. Using the same cross-validation method ensures that the results are comparable and that the data selection does not bias the model performance. Table II presents the comparison results of the proposed architecture and the replicated models. The table presents the number of epochs, the elapsed time, and the models' performance metrics using the KFold and LOGO cross-validation techniques. We set $K = 5$ to estimate the results of each model across five folds of K-Fold.

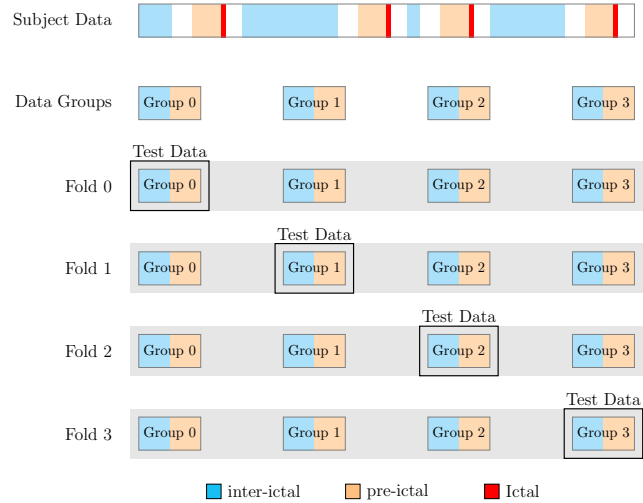


Figure 6. Grouping and test data split using the Leave-One-Group-Out cross-validation method.

The fastest converging method was the CNN model of Truong et al. [21], where the proposed architecture is three convolution layers deep. Our proposed CNN architecture has the 2nd fastest convergence training time. The table also presents the performance metrics in two cross-validation methods: Random data splitting using K-Fold cross-validation and the robust LOGO cross-validation methods. In the K-Fold cross-validation, the proposed architecture achieved an accuracy, sensitivity, and specificity of 96.68%, 97.41% and 95.95%, respectively. However, the randomness of the data splitting may introduce bias in the model performance and does not provide deterministic results. The replicated methods' accuracy dropped down to around 60% in the LOGO cross-validation. The highest performance drop was observed in the sensitivity metric, where the seizure events are from intervals with distinct temporal characteristics. The designed CNN architecture offered significant improvement over other models in the LOGO cross-validation method. Our specialized architecture achieved an accuracy of 78.00%, sensitivity of 65.17%, and specificity of 90.83%. Khan et al [20] CNN model achieved the 2nd highest sensitivity of 47.97%.

5. DISCUSSION

Sensitivity quantifies the model's potential to detect a pre-ictal segment accurately, whereas specificity measures the model's accuracy in classifying inter-ictal segments. Therefore, it is crucial to enhance sensitivity to improve the ability to predict a seizure occurrence. However, enhancing the specificity is essential to avoid inaccurate pre-ictal classifications, which might lead to false alarms. Our model successfully predicted all seizure occurrences in LOGO cross-validation in subjects *chb01*, *chb06*, *chb08*, *chb12*, and *chb14*, with a sensitivity of $94\% \pm 1.5\%$. Nevertheless, the gain in model performance is driven by the fact that our model is reaching a specificity of $90.83\% \pm 1.6\%$ besides yielding comparable or higher sensitivity rates.



TABLE II. The results of the proposed method compared to the replicated architectures.

Author	Avg. training time (seconds)	KFold CV			LOGO CV		
		Acc	Sen	Spec	Acc	Sen	Spec
Khan et al. [20]	243	80.02	82.00	78.05	60.80	47.97	73.62
Truong et al. [21]	229	92.91	93.84	91.98	59.98	39.71	80.25
Lawhern et al. [28]	487	86.94	85.93	87.95	65.24	47.85	82.62
Xu et al. [23]	470	93.06	96.49	89.64	59.62	37.97	81.28
Jemal et al. [31]	332	90.38	90.71	90.04	61.64	41.95	81.34
<i>This study</i>	233	96.68	97.41	95.95	78.00	65.17	90.83

Except for subject *chb14*, which has the lowest specificity of $73.89 \pm 5.6\%$, the proposed architecture consistently reached a specificity of $94.13\% \pm 0.9\%$. The only downside of this architecture requiring further investigation is subject *chb18*, which has a 0% sensitivity. Figure 7 showcases the subject-specific performance metrics obtained using the LOGO cross-validation technique to examine the performance of the proposed architecture.

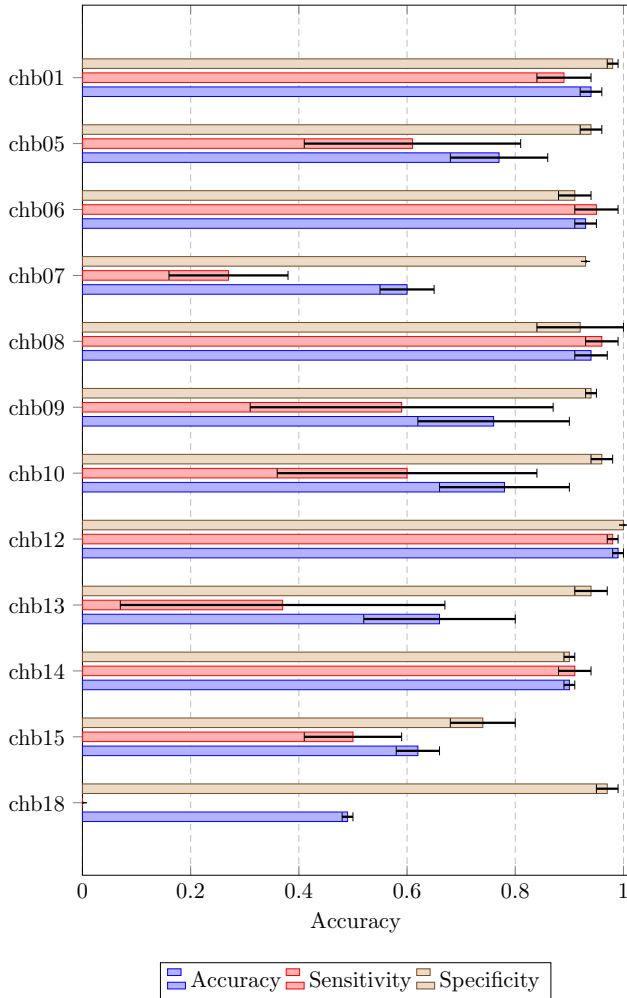


Figure 7. The subject-specific accuracy, sensitivity and specificity metrics of the proposed architecture using LOGO cross-validation. The error bars denote the metrics' standard error.

A. Systematic data sampling

The proposed sampling approach contributes to the reliability and consistency of the results, ensuring reproducible and deterministic evaluations. Otherwise, the model performance may be biased due to inaccurate data selection or grouping. The proposed approach removed any randomness in extracting the EEG segments from the continuous EEG recordings. Moreover, the sampling method offers a practical solution to the data imbalance issue. The practicality and effectiveness are attributed to the systematic nature of over-sampling the majority class and under-sampling the minority class compared to other techniques. Figure 8 presents the results of the proposed approach compared to two sampling methods: Random sampling and Synthetic Minority Oversampling Technique (SMOTE).

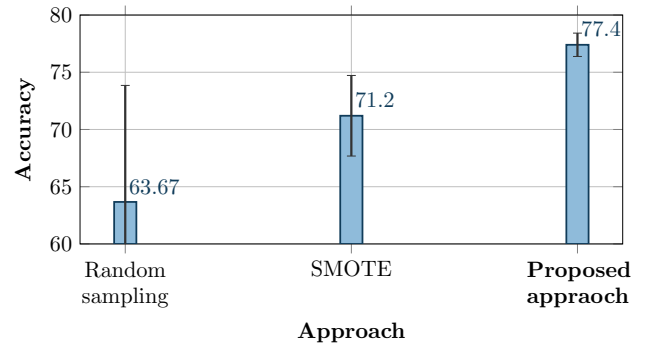


Figure 8. The classification performance comparison between different sampling methods.

B. architectural design choices

The replicated methods were selected based on two factors:

- The availability of the architecture design details.
- The reproducibility of the pre-processing technique.

The reproduced models employed various pre-processing techniques to compare the performance of WPD with different transformation methods. For instance, Khan et al. [20] utilized CWT to extract the frequency-related features, whereas Truong et al. [21] used STFT. The CNN architectures of these studies employed the EEG channels as a feature map channel dimension. Hence, the first layer's

kernel of these models must have as many channels as the EEG input segment, resulting in a wide spatial receptive field. The configured kernel sizes are either (3×3) or (5×5) , resulting in a low receptive field in the frequency and temporal dimensions. Truong et al. [21] used an FFT window hop size of around 133 samples to down-sample $(30 \times 256 = 7680)$ time points in the original 30-second EEG segment to only 59 time points in the temporal dimension. The down-sampling of STFT leads to an expansion of the temporal receptive field of the model. However, this trade-off causes a substantial loss of temporal information to accommodate the frequency dimension.

Xu et al. [23] utilized unprocessed EEG segments directly as the input for their model. The model architecture utilized lengthy 1D max-pooling kernels in the last convolution layer, with a cumulative temporal divisor of 1000. Using large max-pooling kernel sizes, specifically (1×10) , expanded the temporal receptive field. However, this came at the expense of sacrificing the temporal resolution. The model underwent training using EEG segments of 20 seconds in length. Therefore, the last convolution layer achieved a temporal receptive field of 12 seconds, using a kernel size of (3×3) . However, the spatial receptive field was not adequately developed since there was no pooling in the channel dimension, except in the last two convolution layers. Jemal et al. [31] utilized a max-pooling kernel size of (1×16) after the leading two convolution processes. The EEGNet architecture utilized a depthwise convolution kernel of (18×1) kernel size to extract the spatial features. Hence, the spatial receptive field equals the number of channels in the EEG segment.

Our model employs the decompositions of the WPD to further extract the frequency-related information. Furthermore, WPD down-samples the time domain of the EEG segment without losing any temporal information. This concept is embedded in the fact that WPD is lossless and can fully reconstruct the original signal from the coefficients. Hence, WPD enhances the representation of the EEG segments, isolates the frequency sub-bands in the segments, and reduces the temporal dimension. To address the limitations of replicated CNN designs, the proposed CNN architecture has 1D kernels with large receptive fields along the temporal, frequency sub-bands (spectral), and channel dimensions.

The proposed architecture consists of two branches that carry out distinct aggregation of features. The first branch extracts temporal features aggregated from enhanced coefficients across the channels. In contrast, the second branch extracts temporal features aggregated from enhanced channels utilizing the coefficients. The third layer of the architecture uses a depthwise kernel to further extract the spatial features in the first branch and coefficient features from the second branch. Thus, this design allows the model to have a receptive field equal to the number of coefficients and channels. Moreover, the down-sampling process of

WPD enables the model to have an extended temporal receptive field using small kernel sizes without reducing the temporal resolution of the input segment. Table III presents the architecture and its employed input dimensionality of the replicated model and our proposed model. The table also presents the number of parameters in the CNN model, the pre-processing method, the segmentation window size, and the input shape of the EEG segment.

C. Performance comparison

Our proposed model achieved slightly higher or comparable performance for several subjects. Nevertheless, our model achieved significantly higher AUC values for subjects *chb05*, *chb06*, *chb07*, *chb09*, *chb14*, and *chb15*. Additionally, we can observe that for subject *chb01*, *chb13*, *chb14*, only the original EEGNet could deliver comparable results to our model. The only limitation of our proposed architecture is the inability to classify any pre-ictal segments of subject *chb18*. This limitation could be due to the large gap between the recordings (around 50 hours of missing EEG data from the 90 hours of the subject session). Figure 9 presents the AUC metric values of the proposed architecture and the replicated models using the LOGO cross-validation technique.

The seizure prediction task is highly challenging due to the high similarity between the pre- and inter-ictal segments. Additionally, it has been observed that the longer the temporal distance between the training and testing data result in a decrease in the model's performance. Therefore, the accuracy drop in the specificity is lower than that of the sensitivity. Using an under-lapping factor to under-sample the longer inter-ictal data segments improved the model's performance by including more temporal characteristics. However, this holds true only for the inter-ictal class where there are much longer periods available in the subject's data. Therefore, to increase the seizure prediction performance, the model should be trained on longer periods of pre-ictal data.

6. CONCLUSION AND FUTURE WORK

This study presented a novel wavelet-based 1D CNN architecture to build a segment-based seizure prediction model. Concerning the nature of EEG signals, the proposed model employed Wavelet Packet Decomposition to enhance the representation of the EEG signals. Additionally, WPD can extract frequency-related details without increasing the complexity or redundancy of the EEG data. Moreover, the WPD pre-processing technique can reduce the temporal dimension in a lossless manner. Therefore, it is possible to enlarge the receptive field of the model of small-sized kernels without sacrificing the resolution when using traditional down-sampling techniques such as pooling or kernel dilation. The proposed model utilized separable and depthwise convolution layers to further enhance and extract deeper features from the wavelet packets.

We replicated five state-of-the-art CNN-based models from recent studies to establish a head-to-head comparison.



TABLE III. The results of the proposed method compared to the replicated architectures.

Author	Num. of Params.	Pre-process	Segment	Input shape
Khan et al. [20]	188K	CWT	1s	$10 \times 128 \times 18$
Truong et al. [21]	195K	STFT	30s	$115 \times 59 \times 18$
Lawhern et al. [28]	2.1K	Raw	1s	18×256
Xu et al. [23]	942K	Raw	20s	18×5120
Jemal et al. [31]	2.8K	Raw	5s	18×1280
<i>This study</i>	160K	Raw	4s	288×70

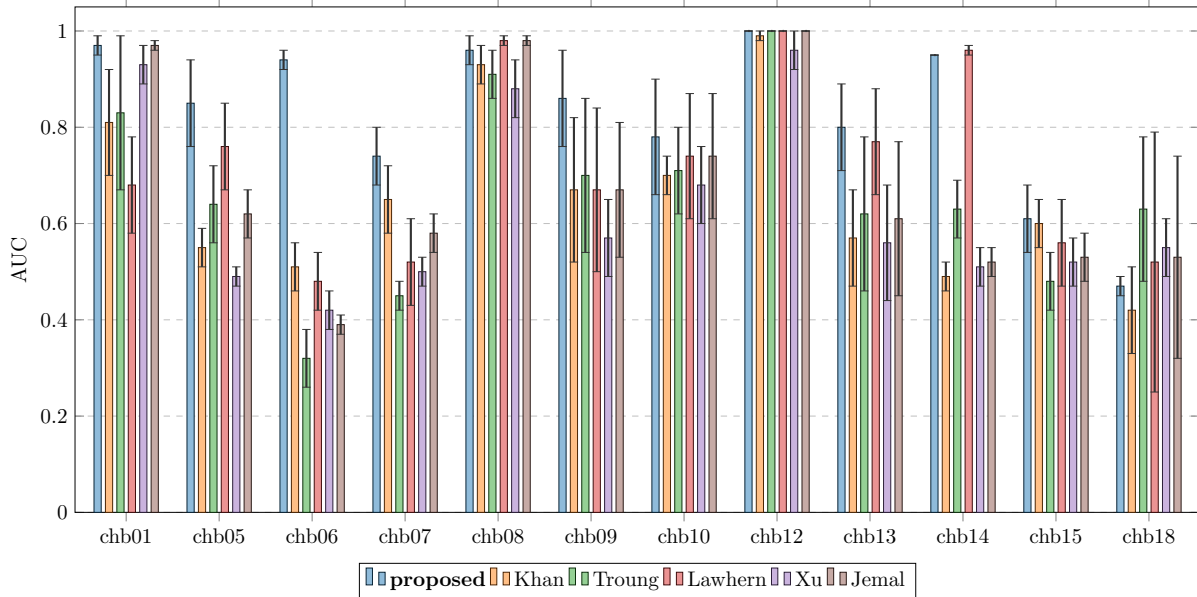


Figure 9. The subject-specific AUC values of the proposed architecture and the replicated methods using LOGO cross-validation. The error bars represent the standard error of the AUC values.

Analysis of these replicated models' architectural limitations motivated the development of the proposed model. The proposed design outperformed the replicated models, significantly improving the seizure prediction ability. We trained and evaluated these models using two cross-validation methods: Random K-fold and Leave-One-Group-Out. While K-Fold delivers nondeterministic results, the LOGO cross-validation technique showcases a practical testing situation and assesses the model's generalization ability. The CHB-MIT dataset is utilized to benchmark the proposed model and the replicated methods. The proposed model outperformed the replicated models, proving the robustness and reliability of the proposed architecture.

The EEG data typically contains long intervals of inter-ictal data, while seizure events with valid pre-ictal periods are rare, leading to a high data imbalance. Most studies utilized random under-sampling techniques to reduce the quantity of inter-ictal samples. This study introduced an under-lapping factor to under-sample the inter-ictal data samples while maximizing the incorporation of temporal characteristics. The suggested under-sampling method improved the certainty of the model's performance and increased the specificity of the model, resulting in infrequent false alarms.

However, this under-sampling technique only applies to the inter-ictal class, where the data is abundant. To further boost the model's sensitivity and overall performance, the model's training requires more pre-ictal samples.

The configured event-related parameters constrain the model training. These fixed parameters include the pre-ictal duration, inter-ictal distance, and post-ictal length. Incorrect parameter settings may lead to training on incorrectly labeled data. Therefore, a more accurate definition of the epileptic states involves utilizing a continuous variable instead of discrete labels. In this sense, the seizure prediction problem can be considered as a regression task instead of a classification task. An example of a continuous variable can be the temporal distance between the processed EEG sample and the impending seizure event. This viewpoint may deliver more insights into the difference between the two epileptic states and how a seizure develops.

A medical application of seizure prediction models requires a high level of explainability and interpretability. While the proposed architecture has an explainable design, further investigation using explainable AI techniques is required to shed light on the model's internals and interpret

its decision-making process. Further design enhancements require integrating the attention mechanism to amplify the most informative parts of the wavelet-transformed EEG signals. The attention module can also deliver a degree of explainability to the model's predictions.

REFERENCES

- [1] O. Devinsky, A. Vezzani, T. J. O'Brien, N. Jette, I. E. Scheffer, M. De Curtis, and P. Perucca, "Epilepsy," *Nature Reviews Disease Primers*, vol. 4, no. 1, p. 18024, May 2018.
- [2] World Health Organization, *Epilepsy: A Public Health Imperative*. Geneva: World Health Organization, 2019.
- [3] W. C. Stacey and B. Litt, "Technology Insight: Neuroengineering and epilepsy—designing devices for seizure control," *Nature Clinical Practice Neurology*, vol. 4, no. 4, pp. 190–201, Apr. 2008.
- [4] W. Tatum, G. Rubboli, P. Kaplan, S. Mirsatari, K. Radhakrishnan, D. Gloss, L. Caboclo, F. Drislane, M. Koutroumanidis, D. Schomer, D. Kasteleijn-Nolst Trenite, M. Cook, and S. Beniczky, "Clinical utility of EEG in diagnosing and monitoring epilepsy in adults," *Clinical Neurophysiology*, vol. 129, no. 5, pp. 1056–1082, May 2018.
- [5] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O'Brien, and A. Razi, "Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 139–155, 2021.
- [6] S. Ramgopal, S. Thome-Souza, M. Jackson, N. E. Kadish, I. Sánchez Fernández, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy & Behavior*, vol. 37, pp. 291–307, Aug. 2014.
- [7] K. Gadhomi, J.-M. Lina, F. Mormann, and J. Gotman, "Seizure prediction for therapeutic devices: A review," *Journal of Neuroscience Methods*, vol. 260, pp. 270–282, Feb. 2016.
- [8] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction — ready for a new era," *Nature Reviews Neurology*, vol. 14, no. 10, pp. 618–630, Oct. 2018.
- [9] L. D. Iasemidis, "Seizure Prediction and its Applications," *Neurosurgery Clinics of North America*, vol. 22, no. 4, pp. 489–506, Oct. 2011.
- [10] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, Feb. 2007.
- [11] Mauro, F. Pinto, A. Leal, F. Lopes, J. Pais, A. Dourado, F. Sales, P. Martins, and C. A. Teixeira, "On the clinical acceptance of black-box systems for EEG seizure prediction," *Epilepsia Open*, vol. 7, no. 2, pp. 247–259, Jun. 2022.
- [12] S. Sanei and J. Chambers, *EEG Signal Processing*, 1st ed. Wiley, Sep. 2007.
- [13] Z. Ren, X. Han, and B. Wang, "The performance evaluation of the state-of-the-art EEG-based seizure prediction models," *Frontiers in Neurology*, vol. 13, p. 1016224, Nov. 2022.
- [14] A. Rai, "Explainable AI: From black box to glass box," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 137–141, Jan. 2020.
- [15] S. Shafieezadeh, G. M. Duma, G. Mento, A. Danieli, L. Antoniazzi, F. Del Popolo Cristaldi, P. Bonanni, and A. Testolin, "Methodological Issues in Evaluating Machine Learning Models for EEG Seizure Prediction: Good Cross-Validation Accuracy Does Not Guarantee Generalization to New Patients," *Applied Sciences*, vol. 13, no. 7, p. 4262, Mar. 2023.
- [16] S. M. Usman, M. Usman, and S. Fong, "Epileptic Seizures Prediction Using Machine Learning Methods," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–10, 2017.
- [17] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packet decomposition for automated epileptic seizure detection and prediction," *Biomedical Signal Processing and Control*, vol. 39, pp. 94–102, Jan. 2018.
- [18] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, vol. 99, pp. 24–37, Aug. 2018.
- [19] L. Kuhlmann, P. Karoly, D. R. Freestone, B. H. Brinkmann, A. Temko, A. Barachant, F. Li, G. Titericz, B. W. Lang, D. Lavery, K. Roman, D. Broadhead, S. Dobson, G. Jones, Q. Tang, I. Ivanenko, O. Panichev, T. Proix, M. Náhlik, D. B. Grunberg, C. Reuben, G. Worrell, B. Litt, D. T. J. Liley, D. B. Grayden, and M. J. Cook, "Epilepsyecosystem.org: Crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG," *Brain*, Aug. 2018.
- [20] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal Onset Seizure Prediction Using Convolutional Networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 2109–2118, Sep. 2018.
- [21] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104–111, 2018.
- [22] H. Daoud and M. A. Bayoumi, "Efficient Epileptic Seizure Prediction Based on Deep Learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, Oct. 2019.
- [23] Y. Xu, J. Yang, S. Zhao, H. Wu, and M. Sawan, "An End-to-End Deep Learning Approach for Epileptic Seizure Prediction," in *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, Aug. 2020, pp. 266–270.
- [24] R. Jana and I. Mukherjee, "Deep learning based efficient epileptic seizure prediction with EEG channel optimization," *Biomedical Signal Processing and Control*, vol. 68, p. 102767, Jul. 2021.
- [25] R. Hussein, S. Lee, R. Ward, and M. J. McKeown, "Semi-dilated convolutional neural networks for epileptic seizure prediction," *Neural Networks*, vol. 139, pp. 212–222, Jul. 2021.
- [26] Z. Wang, J. Yang, and M. Sawan, "A Novel Multi-scale Dilated 3D CNN for Epileptic Seizure Prediction," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. Washington DC, DC, USA: IEEE, Jun. 2021, pp. 1–4.



- [27] Y. Gao, X. Chen, A. Liu, D. Liang, L. Wu, R. Qian, H. Xie, and Y. Zhang, "Pediatric Seizure Prediction in Scalp EEG Using a Multi-Scale Neural Network With Dilated Convolutions," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–9, 2022.
- [28] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, Oct. 2018.
- [29] Kai Keng Ang, Zhang Yang Chin, Haihong Zhang, and Cuntai Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. Hong Kong, China: IEEE, Jun. 2008, pp. 2390–2397.
- [30] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, Feb. 2020.
- [31] I. Jemal, N. Mezghani, L. Abou-Abbas, and A. Mitiche, "An Interpretable Deep Learning Classifier for Epileptic Seizure Prediction Using EEG Data," *IEEE Access*, vol. 10, pp. 60 141–60 150, 2022.
- [32] X. Lu, A. Wen, L. Sun, H. Wang, Y. Guo, and Y. Ren, "An Epileptic Seizure Prediction Method Based on CBAM-3D CNN-LSTM Model," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 417–423, 2023.
- [33] X. Xu, Y. Zhang, R. Zhang, and T. Xu, "Patient-specific method for predicting epileptic seizures based on DRSN-GRU," *Biomedical Signal Processing and Control*, vol. 81, p. 104449, Mar. 2023.
- [34] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
- [35] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, Jun. 2000.
- [36] M. Shahbazi and H. Aghajan, "A Generalizable Model for Seizure Prediction Based on Deep Learning Using Cnn-Lstm Architecture," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Anaheim, CA, USA: IEEE, Nov. 2018, pp. 469–473.
- [37] A. R. Ozcan and S. Erturk, "Seizure Prediction in Scalp EEG Using 3D Convolutional Neural Networks With an Image-Based Approach," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 11, pp. 2284–2293, Nov. 2019.
- [38] X. Wei, L. Zhou, Z. Zhang, Z. Chen, and Y. Zhou, "Early prediction of epileptic seizures using a long-term recurrent convolutional network," *Journal of Neuroscience Methods*, vol. 327, p. 108395, Nov. 2019.
- [39] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomedical Signal Processing and Control*, vol. 31, pp. 398–406, Jan. 2017.
- [40] Y. Meyer and D. H. Salinger, *Wavelets and Operators*, ser. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge university press, 1992, no. 37.
- [41] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, Apr. 2021.
- [42] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.