technologies

MDPI

*Article*

# Similarity-Driven Personalization and Optimization for Long-Horizon EEG Seizure Prediction

**Kiyan Afsari [1,*], Christian Ritz [2] and May El Barachi [3]**

[1] School of Engineering, University of Wollongong in Dubai, Dubai 20183, United Arab Emirates
[2] School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong 2522, Australia; critz@uow.edu.au
[3] School of Computer Science, University of Wollongong in Du bai, Dubai 20183, United Arab Emirates; maielbarachi@uowdubai.ac.ae
* Correspondence: kiyanafsari@uowdubai.ac.ae

## Abstract

Epileptic seizure prediction using an Electroencephalogram (EEG) can improve patient safety by enabling early intervention, yet most existing approaches focus on short prediction horizons with limited personalization or computational efficiency. This study presents a unified deep learning framework evaluated across ten pre-ictal prediction windows up to 300 min before seizure onset, using recordings from 161 patients and 1023 seizure events. At the 5 min horizon, the generalized model achieved 96.30% accuracy and 91.62% sensitivity. Two complementary personalization strategies are introduced: incremental transfer learning, which progressively fine-tunes the generalized model using patient-specific data, and Dynamic Time Warping (DTW)-based similarity personalization, which constructs a morphology-aware training cohort from a single reference seizure. Personalized models consistently outperform generalized baselines, particularly at longer horizons, with the DTW-based approach achieving 89.68% accuracy using only 70 similar patients. Reliable prediction is demonstrated up to 60 min prior to onset, while model optimization reduces computational complexity with minimal performance loss, supporting deployment in resource-constrained clinical environments.

**Keywords:** epilepsy; seizure prediction; biomedical signal processing

## 1. Introduction

Epilepsy affects approximately 50 million people globally, and the unpredictability of seizures poses severe risks to patient safety and quality of life [1]. Electroencephalogram (EEG) remains the gold standard for monitoring brain activity, and while deep learning has achieved high sensitivity in automated seizure detection [2–4], transitioning from offline clinical analysis to real-time ambulatory monitoring presents significant engineering challenges. Deploying seizure prediction algorithms on edge devices such as wearable headbands [5] or subcutaneous implants requires operating under strict constraints on battery life, memory, and processing power [6]. Standard deep learning models are computationally expensive, and the high spatial redundancy of multi-channel EEG recordings further compounds this problem, making channel selection and dimensionality reduction essential for lightweight edge-compatible deployment [7–9].

A further challenge is inter-patient variability: seizure onset patterns, background EEG activity, and temporal dynamics differ substantially across individuals, limiting the

Check for updates

Academic Editor: Valeri Mladenov

Received: 15 May 2026
Revised: 9 June 2026
Accepted: 11 June 2026
Published: 13 June 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

effectiveness of population-level models [10,11]. To address this, we adopt a similarity-driven personalization strategy in which patient similarity is quantified using EEG-derived features and Dynamic Time Warping (DTW)-based seizure pattern matching, enabling preferential integration of data from the most comparable patients during model adaptation. This staged approach balances population-scale robustness with patient-specific refinement, particularly at longer prediction horizons where variability is most pronounced.

Despite these advances, three critical gaps remain inadequately addressed in the existing literature. First, the overwhelming majority of seizure prediction studies focus on short prediction horizons of under 30 min, and systematic investigations of how model performance degrades as a function of prediction horizon, or where the temporal boundary of reliable prediction lies, remain scarce. Second, while personalization has been recognized as important for addressing inter-patient variability, existing approaches either require substantial patient-specific data, suffer from cold-start limitations, or treat personalization as a secondary post hoc step rather than a core design principle. Similarity-based personalization leveraging seizure morphology matching has not been systematically explored for EEG seizure prediction despite its demonstrated success in other biomedical AI applications. Third, model optimization for edge deployment is typically treated in isolation, decoupled from personalization and prediction horizon considerations, leaving a unified framework that jointly addresses all three dimensions absent from the literature. This study directly addresses all three of these gaps within a single cohesive framework.

Building on prior work [7,12], this study proposes a unified framework jointly addressing personalization, optimization, and long-horizon seizure prediction up to 300 min prior to onset. The main contributions are as follows:

1. An evaluation of EEG-based seizure prediction across horizons up to 300 min, exploring the temporal limits of reliability.
2. A similarity-driven personalization framework that enables data-efficient patient-specific adaptation using Dynamic Time Warping (DTW)-based seizure selection.
3. A structured optimization pipeline integrating temporal window selection, electrode reduction, and model compression for improved computational efficiency.
4. A unified framework that jointly addresses prediction accuracy, personalization, and deployability, which has not been systematically explored in prior work.

While CNN-LSTM hybrid architectures have been previously explored for seizure prediction [13,14], the proposed framework differs in several important ways. Existing CNN-LSTM studies predominantly focus on short prediction horizons, typically under 30 min, and treat personalization and optimization as secondary or absent considerations. Unlike existing studies that focus primarily on either personalization or model optimization in isolation, this work proposes a unified framework that jointly addresses personalization, computational efficiency, and long-horizon prediction. The novelty of this study lies not only in the integration of these components, but in the systematic investigation of seizure predictability across extended temporal horizons, enabling the identification of the boundary between clinically meaningful and unreliable prediction. Furthermore, the proposed similarity-driven personalization approach enables data-efficient adaptation by leveraging only the most relevant seizure patterns, while the multi-stage optimization pipeline significantly reduces model complexity without compromising performance. The novelty therefore lies not in the CNN-LSTM architecture itself, but in its systematic application within a unified framework that jointly addresses long-horizon prediction across ten temporal windows, data-efficient personalization through DTW-based patient selection, and computational optimization designed for edge deployment.

## 2. Related Work

The EEG-based seizure prediction has traditionally relied on handcrafted feature extraction combined with classical machine learning classifiers. Commonly used features include statistical moments, spectral band power, entropy measures, Hjorth parameters, and wavelet-based descriptors, which aim to capture changes in signal amplitude, frequency content, and non-linear dynamics preceding seizure onset [15,16]. These engineered representations are typically classified using algorithms such as support vector machines, random forests, and gradient boosting methods [17] and have demonstrated promising performance, particularly in patient-specific scenarios where seizure patterns are relatively stable [18].

With the advancement of deep learning, end-to-end approaches that operate directly on raw EEG signals have gained increasing attention. Convolutional neural networks (CNNs) have been widely adopted to learn spatial and temporal patterns from multichannel EEG [19,20], while recurrent architectures such as long short-term memory (LSTM) networks are used to model temporal dependencies inherent in seizure evolution [21]. Hybrid CNN–LSTM architectures further combine these strengths by extracting local representations with CNN layers and modeling longer temporal dependencies with recurrent units [13,14]. More recently, attention-based models and Transformers have been explored for EEG analysis, motivated by their ability to capture long-range dependencies through self-attention mechanisms [22,23].

Despite their success, feature-based and end-to-end approaches are often studied independently, with limited discussion on how input representation choices interact with personalization strategies, prediction horizon length, or computational efficiency. Moreover, while deep learning models reduce reliance on manual feature engineering, they often introduce substantial computational overhead, limiting their suitability for real-time or resource-constrained deployment. Table 1 highlights that most existing studies focus on short-term prediction using generalized models, motivating the need for personalized and long-horizon seizure prediction frameworks.

**Table 1.** Recent EEG-Based Seizure Prediction Studies.

| Study (Year) | Dataset | Model Type | Prediction Horizon | Key Limitation |
|---|---|---|---|---|
| **Liu et al. (2024) [14]** | CHB-MIT | Pseudo-3D CNN + BiConvLSTM3D | Short/unspecified | Focus on spatial features, limited long-horizon analysis |
| **Esmaeilpour et al. (2024) [21]** | CHB-MIT | CNN + Ensemble Classifier | Minutes (5 min implied) | No cross-subject adaptation reported |
| **Yuan et al. (2024) [24]** | CHB-MIT | Hybrid DenseNet-ViT with Attention | Minutes (preictal windows typical) | No personalization; global model |
| **Sadeghi Khansari et al. (2025) [25]** | CHB-MIT | DWT + Deep Learning (FNN) | Not explicitly long horizon | High performance but no personalization focus |
| **Upadhyay et al. (2025) [26]** | Public EEG | Explainable Hybrid DNN | Unspecified | Seizure vs. non-seizure classification not long horizon |
| **Li et al. (2025) [27]** | CHB-MIT & MSSM | Spatio-Temporal Attention Network | 15–45 min | Still relatively short horizons |

The prediction horizon, defined as the temporal distance between model prediction and seizure onset, is a critical yet underexplored dimension of seizure prediction research. The majority of existing studies focus on short-term horizons, typically ranging from a few minutes to 30 min prior to seizure onset, where pre-ictal EEG changes are more pronounced and easier to detect [28,29]. Short-horizon prediction is often framed as imminent seizure detection, aiming to trigger immediate interventions such as neurostimulation or

alerts. Several studies have explored moderately longer horizons, extending prediction windows to 60 or 90 min, often at the expense of reduced sensitivity or increased false alarm rates [30,31]. These results suggest that EEG signatures associated with seizure generation become increasingly subtle as the temporal distance from seizure onset increases. However, systematic investigations of long-horizon prediction beyond one or two hours remain scarce. Existing works rarely analyze how model performance degrades as a function of prediction horizon or whether certain architectures are more robust to long-range temporal uncertainty.

Most seizure prediction systems are either population-level, benefiting from larger datasets but limited by inter-patient variability, or patient-specific, achieving higher accuracy when sufficient data are available but facing cold-start problems and poor scalability [32,33]. Hybrid approaches such as transfer learning and fine-tuning partially bridge this gap, yet personalization remains largely treated as a secondary step rather than a core design principle, and scalable frameworks that systematically balance population-level robustness with individual adaptation remain an open challenge. Similarity-based learning has been successfully applied in biomedical AI for patient clustering, metric learning, and personalized treatment recommendation, enabling models to leverage shared patterns across similar patients [34–36]. By identifying cohorts of similar patients, models can leverage shared patterns while avoiding the dilution effects associated with heterogeneous population-level training. In time-series analysis, distance metrics such as DTW and correlation-based measures enable selective integration of training data from comparable patients [37]. Despite this success, similarity-based personalization has not been systematically applied to EEG seizure prediction, leaving its potential for long-horizon forecasting largely unexplored.

For real-world deployment in continuous monitoring scenarios, EEG prediction models must meet strict constraints on power, memory, and latency. Prior work has explored channel selection, PCA, ICA, and lightweight architectures to reduce input dimensionality, alongside model-level optimizations such as pruning and parameter reduction [38–40]. However, optimization is typically treated as a post hoc step, decoupled from personalization and prediction horizon considerations, and a unified framework jointly addressing all three remains absent from the literature.

## 3. Material and Methods

### *3.1. Overview of the Framework*

This study proposes a unified framework for personalized, optimized, and long-horizon EEG-based seizure prediction that integrates three complementary components: similarity-driven personalization, multi-stage model optimization, and extended pre-ictal forecasting. Unlike conventional approaches that treat these dimensions independently, our framework addresses them jointly within a cohesive pipeline designed to balance predictive accuracy, computational efficiency, and clinical applicability.

The framework consists of five major stages, illustrated schematically in Figure 1. First, raw multi-channel EEG data from a large multi-patient cohort undergoes preprocessing and segmentation into pre-ictal, ictal, and inter-ictal windows corresponding to varying prediction horizons. Second, a population-level deep learning model based on a hybrid CNN-LSTM architecture is trained on aggregated data from all patients to learn generalized seizure-related patterns. Third, patient similarity is quantified using EEG-derived features and seizure characteristics, enabling the identification of clinically comparable patient subsets. Fourth, personalized models are derived through multi-step fine-tuning, where population-level models are progressively adapted using data from similar patients and finally refined with individual patient data. Fifth, a comprehensive optimization pipeline is

applied to reduce model complexity through channel selection, dimensionality reduction, and temporal window optimization.

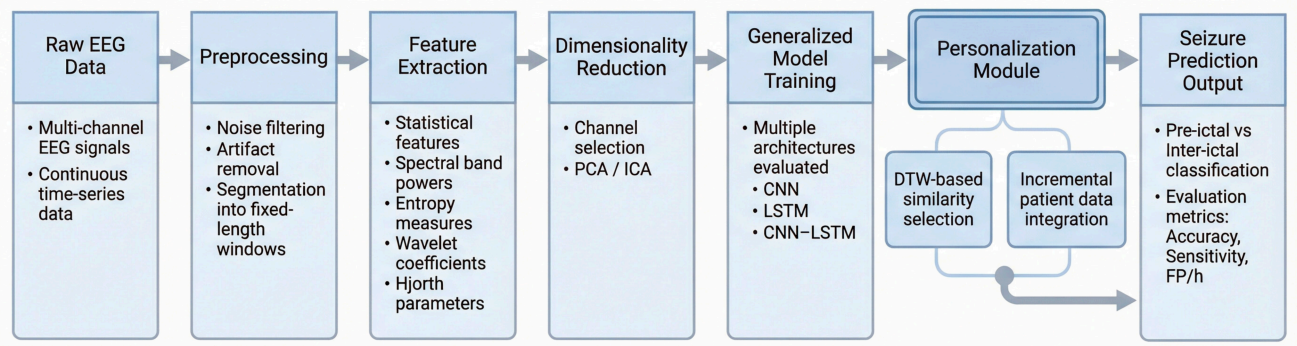



**Figure 1.** Overall system architecture of the proposed EEG-based seizure prediction framework, illustrating data acquisition, preprocessing, feature extraction, model training, personalization strategies, and evaluation.

### *3.2. Dataset and Preprocessing*

The dataset used in this study comprises long-term continuous EEG recordings from 161 patients diagnosed with epilepsy, obtained from the EPILEPSIAE database. A total of 1023 clinically annotated seizure events are included, with seizure frequency per patient ranging from 2 to 47 events. As illustrated in Figure 2, the distribution of seizure frequency is markedly skewed, with most patients experiencing between 4 and 10 seizures, yielding an average seizure count of approximately 7–8 events per patient. EEG recordings were acquired using the standard 10–20 electrode placement system with 19 to 23 channels depending on the recording protocol, and all signals were sampled at 256 Hz. Seizure duration statistics are summarized in Figure 2, which shows that seizure lengths span several orders of magnitude, with a median duration of approximately 60 s and the majority of events lasting between 30 and 120 s. Clinical annotations indicating seizure onset and offset times were provided by board-certified epileptologists based on visual inspection of the EEG recordings.

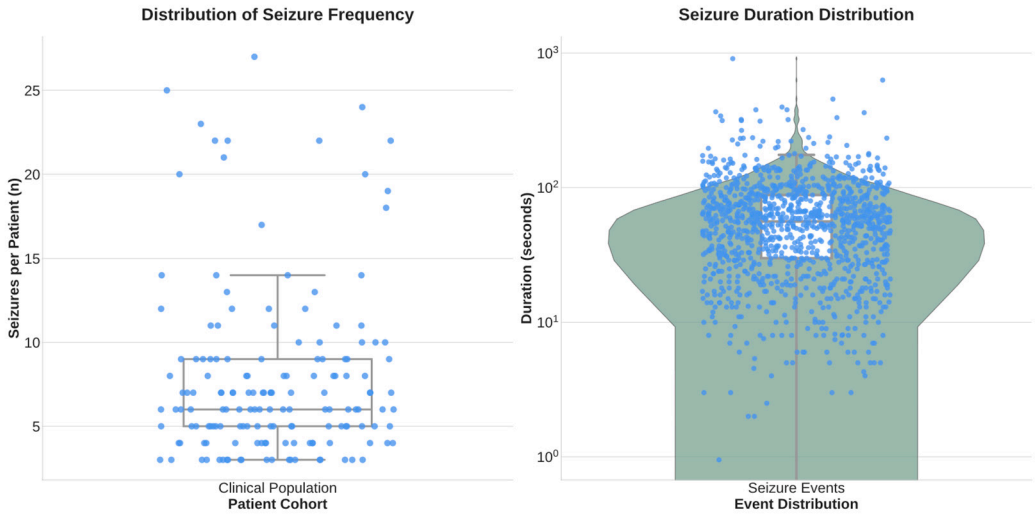



**Figure 2.** Distribution of seizure frequency (**left**) and seizure duration (**right**) across the 161-patient cohort. The **left** panel shows seizure counts per patient via boxplot and strip plot; the **right** panel presents event duration on a logarithmic scale using a combined violin and boxplot visualization.

Preprocessing was performed in multiple stages to ensure signal quality and consistency across patients. First, all EEG recordings were resampled to a uniform sampling rate

of 256 Hz using polyphase filtering to minimize aliasing artifacts introduced during the resampling process. It is noted that anti-aliasing filtering is inherently applied during the original EEG acquisition prior to analog-to-digital conversion. Second, a bandpass finite impulse response (FIR) filter with cutoff frequencies of 0.5 Hz and 40 Hz was applied to remove low-frequency drift and high-frequency noise while preserving the clinically relevant EEG spectrum, which encompasses delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands. Third, electrode channels with excessive artifacts, signal dropouts, or impedance issues were identified through automated quality assessment based on signal variance and amplitude thresholds and subsequently excluded from further analysis. Missing or permanently noisy channels identified during this artifact rejection stage were excluded entirely from the analysis for the affected patient, with the remaining channels retained and used consistently across all prediction horizons for that patient. Additionally, segment-level artifact rejection was applied, discarding any 10 s window containing values beyond the defined amplitude thresholds. Following artifact rejection, all EEG signals were normalized using z-score normalization applied on a per-channel basis, where the mean and standard deviation used for normalization were estimated exclusively from the training partition and subsequently applied to the validation and test sets without refitting, ensuring that no information from held-out data influenced the preprocessing stage. This approach prevents indirect data leakage through normalization and ensures that the preprocessing pipeline is consistent with a realistic deployment scenario where test data statistics are unavailable during model development. Following preprocessing, the continuous EEG recordings were segmented into 50% overlapping epochs for classification. Three classes were defined: pre-ictal, ictal, and inter-ictal. Ictal segments corresponded to the time interval between clinically annotated seizure onset and offset. Pre-ictal segments were defined as the period immediately preceding seizure onset, with variable temporal extents corresponding to different prediction horizons as shown in Figure 3. Specifically, prediction windows were systematically investigated at intervals of 5, 10, 15, 30, 45, 60, 75, 90, 120, 180, 240, and 300 min prior to seizure onset. Inter-ictal segments were sampled from background activity at least four hours from any seizure event to avoid pre-ictal or post-ictal contamination.

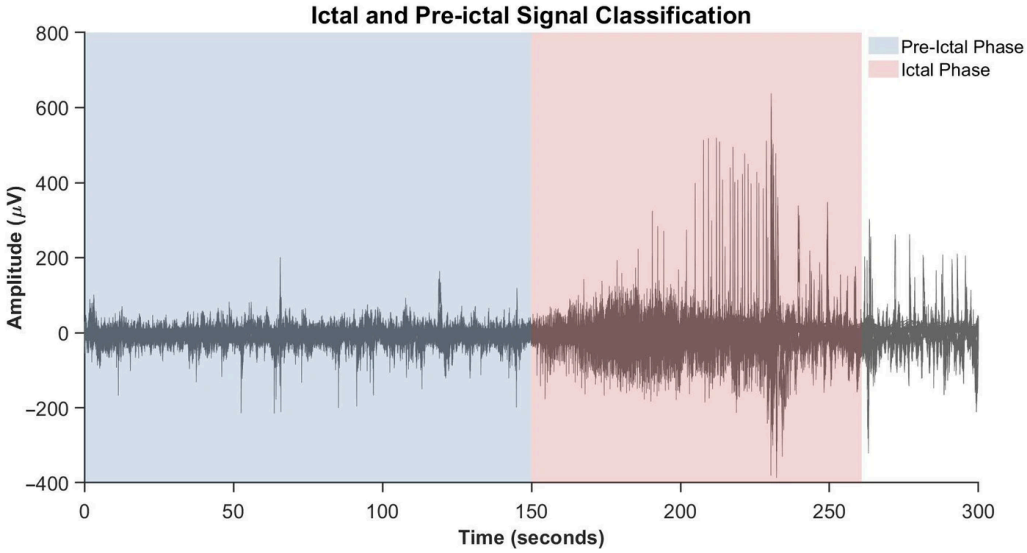



**Figure 3.** Definition of temporal segments for classification. The diagram illustrates the distinction between the ictal phase and the pre-ictal phase.

To address class imbalance, augmentation techniques including temporal jittering and Gaussian noise injection were applied to pre-ictal segments, while inter-ictal segments were under-sampled to achieve a balanced training distribution. Each EEG segment was divided into 10 s windows with 50% overlap, yielding tensors of shape (19, 2560) for a standard 19-channel, 256 Hz configuration and approximately 60 samples per seizure event, as illustrated in Figure 4. Data for each prediction horizon were extracted from the fixed pre-ictal interval immediately preceding that time point, ensuring no future information was accessible to the model.

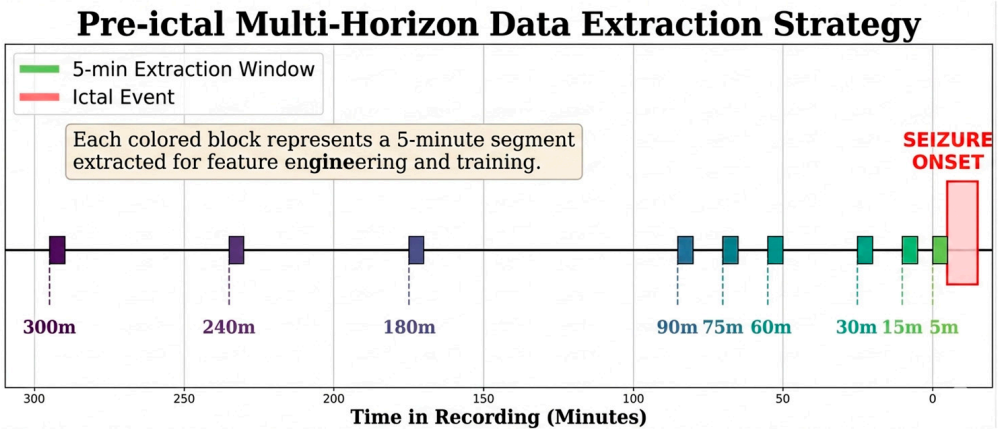



**Figure 4.** Pre-ictal Multi-Horizon Data Extraction Strategy chart showing 5 min segments extracted from a recording leading to a seizure onset.

Each prediction horizon is treated as an independent classification task, with separate models trained and evaluated using only the corresponding pre-ictal segments. This ensures that performance at each horizon reflects true predictive capability at that specific temporal distance, without contamination from shorter or longer pre-ictal intervals. In addition to signal segmentation and normalization, a comprehensive set of handcrafted features, as shown in Table 2, was extracted from each EEG segment to characterize the temporal, spectral, and time-frequency properties of preictal and interictal activity. Consistent with our prior work, these features were selected for their proven effectiveness, interpretability, and computational efficiency in seizure prediction tasks. Statistical descriptors were used to capture amplitude distribution and signal variability, while non-linear and complexity-based measures were included to reflect changes in EEG dynamics preceding seizure onset. Spectral features were computed to quantify frequency-specific energy distributions across clinically relevant EEG bands, and time-frequency features were extracted to capture transient, non-stationary patterns commonly observed in preictal EEG.

A seizure-wise partitioning strategy was employed to prevent data leakage, with each seizure event treated as an independent unit exclusively assigned to either training, validation, or test sets. The training set contained at least one seizure per patient to ensure broad inter-patient coverage, while validation and test sets consisted entirely of unseen seizure events. For each prediction horizon, only EEG data occurring strictly prior to the defined pre-ictal interval was used, ensuring no future information was introduced. For personalized modeling, additional patient-specific fine-tuning splits were defined as described in Section 3.4. It is important to note that these two strategies operate at different levels and are not mutually exclusive. Seizure-wise partitioning ensures that individual seizure events are not shared across training, validation, and test splits, preventing data leakage at the seizure level. Patient-independent evaluation, on the other hand, ensures that all seizure events belonging to a given test patient are entirely excluded from train-

ing, operating at the patient level. Together, these two strategies form a strict two-level partitioning protocol that prevents both seizure-level and patient-level data leakage.

**Table 2.** Handcrafted-features extracted from EEG.

| Feature Category | Feature | Description |
|---|---|---|
| **Statistical** | Mean | Average signal amplitude |
| | Standard Deviation | Signal variability |
| | Variance | Power dispersion |
| | Skewness | Signal asymmetry |
| | Kurtosis | Peakness and tail behavior |
| | Zero Crossing Rate | Temporal oscillation rate |
| **Hjorth Parameters** | Activity | Signal power |
| | Mobility | Mean frequency estimate |
| | Complexity | Signal shape variation |
| **Spectral** | Band Power ($\delta$, $\theta$, $\alpha$, $\beta$, $\gamma$) | Energy in standard EEG bands |
| | Relative Band Power | Normalized spectral contribution |
| | Spectral Entropy | Frequency distribution randomness |
| **Time–Frequency** | Wavelet Coefficients | Multi-resolution signal representation |
| | Wavelet Energy | Energy at different scales |
| **Spatial (when applicable)** | Common Spatial Pattern (CSP) | Discriminative spatial filtering |

To further guard against data leakage, several additional measures were implemented. First, all overlapping windows were generated independently within each seizure event, and no window was allowed to span across seizure boundaries or across train/test splits. Second, for each prediction horizon, only EEG data occurring strictly within the defined pre-ictal interval was used, ensuring no future information was accessible to the model. Third, for the personalization experiments, DTW similarity computations were performed exclusively on training data, with no information from the test or blind patient sets used during the similarity ranking or cohort selection process. Fourth, all preprocessing parameters, including normalization statistics and artifact thresholds, were estimated exclusively from the training partition and applied to validation and test sets without refitting, preventing any indirect leakage through preprocessing.

### *3.3. Baseline Model Architecture*

A hybrid CNN-LSTM architecture was selected as the baseline model, as it provides a favorable balance between accuracy, temporal modeling capability, and computational tractability across a wide range of prediction windows, as illustrated in Figure 5. The CNN component comprises three convolutional blocks (32, 64, and 128 filters) with batch normalization, ReLU activation, and max pooling, producing hierarchical temporal feature maps. These are passed to two stacked bidirectional LSTM layers (128 units each, dropout 0.5) that model long-range temporal dependencies in both directions. A two-layer fully connected classifier then maps the LSTM output to a sigmoid (binary) or softmax (multi-class) prediction. The model was trained using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 32. Training was conducted for a maximum of 30 epochs, with manual monitoring applied to stop training early in cases where signs of overfitting were observed, such as increasing validation loss alongside stable or improving training loss. The loss function used was binary cross-entropy for the two-class (pre-ictal vs. inter-ictal) setting and categorical cross-entropy for the three-class (pre-ictal, ictal, inter-

ictal) setting. Weight initialization followed the He normal initialization scheme, consistent with the ReLU activation functions used throughout the convolutional blocks. Dropout with a rate of 0.5 was applied to the bidirectional LSTM layers as a regularization measure. All experiments were conducted on an NVIDIA GPU with CUDA acceleration, implemented in MATLAB's Deep Learning Toolbox 2024b. Random seeds were fixed across all experiments to ensure reproducibility of results. Hyperparameter selection was performed through a structured grid search conducted on the validation set, prior to any evaluation on the test set. The following hyperparameters were systematically explored: number of convolutional filters (32, 64, 128), number of LSTM units (64, 128, 256), dropout rate (0.3, 0.5, 0.7), learning rate (0.0001, 0.001, 0.01), and batch size (16, 32, 64). The combination yielding the highest validation accuracy was selected as the final configuration and applied consistently across all prediction horizons and personalization experiments. This procedure was applied independently for the generalized model, and the selected hyperparameters were subsequently used as the initialization point for all personalization and optimization experiments, ensuring a consistent and reproducible experimental setup throughout the study.

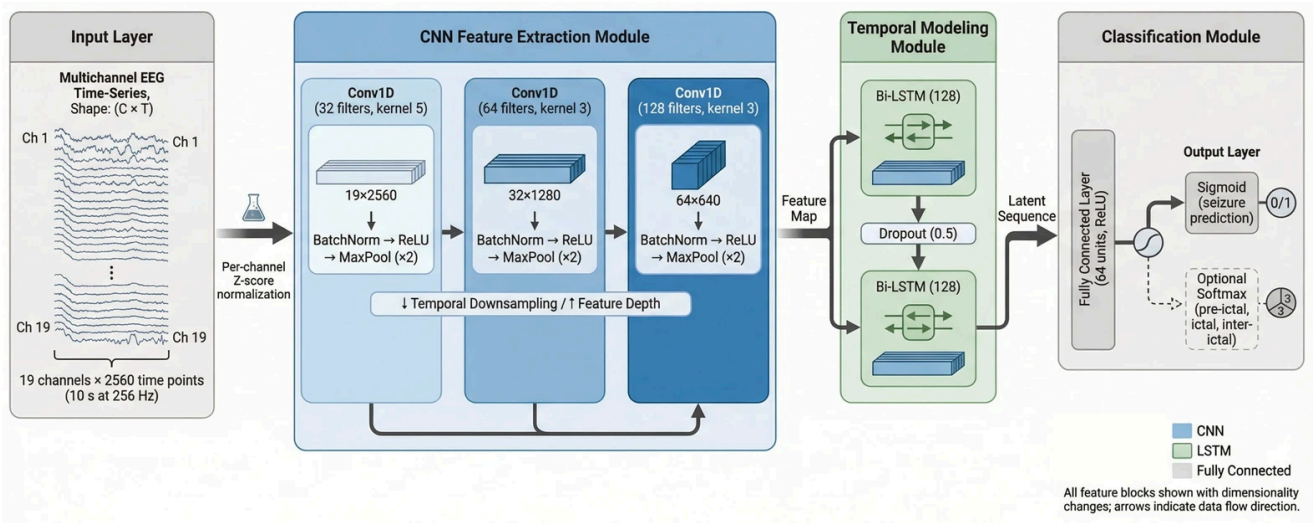



**Figure 5.** Schematic diagram of the baseline CNN-LSTM architecture showing the input layer, three convolutional blocks with batch normalization and max pooling, two bidirectional LSTM layers with dropout, and the fully connected classification module.

The baseline model is trained on the aggregated training set comprising data from all patients in the training partition, resulting in a population-level model that learns generalized seizure prediction patterns. This population-level model serves as the initialization point for subsequent personalized model adaptation, as detailed in Section 3.4. The baseline model contains approximately 63.4 million trainable parameters, concentrated primarily in the LSTM and fully connected layers, and serves as the performance benchmark against which all personalized and optimized variants are evaluated.

### *3.4. Similarity-Driven Personalization*

Given the substantial inter-patient variability in EEG characteristics and seizure dynamics, personalization is a critical component of the proposed seizure prediction framework. Rather than relying exclusively on population-level models or fully patient-specific training, we adopt two complementary personalization strategies that enable progressive adaptation while maintaining robustness to limited patient data. Both strategies are designed to integrate patient-specific information without introducing data leakage or compromising generalization.

3.4.1. Similarity-Based Personalization Using Dynamic Time Warping

DTW was selected as the similarity metric due to its ability to perform non-linear temporal matching through local stretching and compression, capturing variability in seizure duration, phase, and progression across patients. Unlike Euclidean or correlation-based measures, DTW handles temporal misalignment without requiring strict sequence alignment, and unlike embedding-based approaches, it operates directly on time-series data without additional training stages, making it lightweight and well-suited to the proposed deployment-oriented framework. DTW is a distance-based technique that measures similarity between temporal sequences by aligning them non-linearly in time as shown in Equations (1) and (2), making it well-suited for EEG signals that exhibit temporal distortions, variable seizure durations, and phase shifts across patients. To measure morphological similarity between seizures, DTW was employed as a temporal alignment metric. Given two time-series sequences $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_m]$, DTW computes a cost matrix based on pairwise distances:

$$D_{i,j} = (x_i - y_j)^2 \tag{1}$$

The DTW distance is defined as the minimum cumulative cost along an optimal warping path $\pi$ that aligns the two sequences while satisfying boundary and monotonicity constraints:

$$DTW(x, y) = min \sqrt{\sum_{(i,j)\in\pi} D_{i,j}} \tag{2}$$

DTW distances were computed between the target patient's pre-ictal and inter-ictal EEG segments and those of all training patients using MATLAB's built-in dtw function. For multichannel EEG signals, DTW distances were computed independently for each channel, and the resulting per-channel distances were aggregated by computing their mean across all retained channels to produce a single patient-to-patient similarity score. Prior to DTW computation, all EEG signals were z-score normalized on a per-channel basis using statistics estimated exclusively from the training partition, ensuring that amplitude differences across patients and recording sessions did not dominate the similarity measure. A Sakoe–Chiba band constraint with a warping window of 10% of the sequence length was applied to limit excessive temporal distortion and reduce computational cost. For each blind patient, DTW distances were computed between the selected reference seizure segment and all seizure segments available in the training set, with the final patient-level similarity score obtained by averaging distances across all segments and channels for each candidate training patient. Patients were then ranked in ascending order of this aggregated DTW distance, with lower distances indicating higher morphological similarity. The most similar patients were then selected to form a personalized training cohort, and the generalized model was fine-tuned on this similarity-constrained subset rather than the full population. All similarity computations were performed exclusively on training data to prevent information leakage.

It should be noted that in the current implementation, the reference seizure used for DTW-based similarity computation was randomly selected from the available seizures of the blind patient. To ensure that results were not biased by a fortunate or unrepresentative seizure selection, this random selection process was repeated five times and results were averaged across all five runs, providing a more robust and unbiased estimate of personalization performance. While this approach yields consistent and robust personalization performance, a formal sensitivity analysis investigating the impact of reference seizure selection on personalization outcomes was not conducted in this study and is acknowledged as a limitation.

3.4.2. Incremental Personalization via Progressive Patient Data Integration

The second strategy follows an incremental learning paradigm in which a generalized model is progressively fine-tuned using increasing proportions of the target patient's data. Four stages are defined in Table 3, ranging from minimal (10% patient data) to maximized (75%), with the remainder reserved for testing. This design emulates realistic clinical deployment, where patient recordings accumulate over time rather than being fully available at initialization, and enables systematic evaluation of how personalization depth influences predictive performance and robustness.

**Table 3.** Incremental personalization stages for a single patient.

| Personalization Stage | Patient-Specific Input Data | Test Data |
|---|---|---|
| **Minimal** | 10% | 90% |
| **Low-resource** | 25% | 75% |
| **Moderate** | 50% | 50% |
| **Maximized** | 75% | 25% |

At each incremental stage, the model is retrained or fine-tuned using the augmented training set, and performance is evaluated on the held-out test data. This staged evaluation provides insight into the rate at which patient-specific information improves seizure prediction accuracy and whether diminishing returns are observed beyond a certain level of personalization. Furthermore, it allows direct comparison between similarity-based personalization and data-driven adaptation under varying data availability conditions.

*3.5. Model Optimization Pipeline*

A hierarchical optimization pipeline was adopted to reduce computational complexity while preserving predictive accuracy, as illustrated in Figure 6. Starting from a full-configuration baseline benchmark, the pipeline proceeds through four stages: architecture simplification (reducing CNN filters, LSTM units, and fully connected parameters); temporal window and electrode reduction using DTW, SHAP, and feature-importance ranking; dimensionality reduction via PCA and ICA; and deployment-oriented compression through pruning and quantization. Together, these stages provide a structured pathway from high-performance population-level models to lightweight, edge-deployable seizure prediction systems.

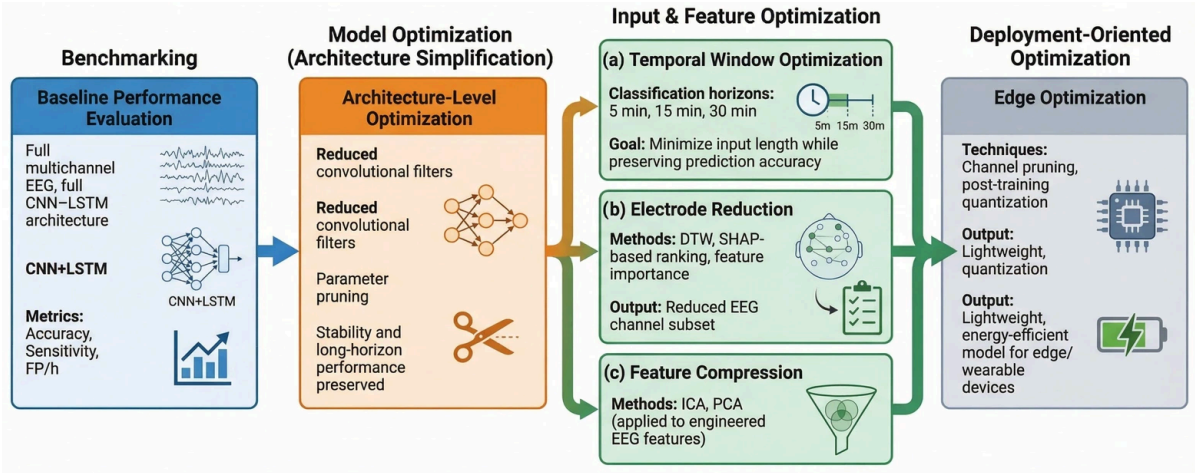



**Figure 6.** Multi-stage optimization pipeline for EEG seizure prediction, including baseline benchmarking, architecture simplification, input and feature reduction, and deployment-oriented optimization through pruning and quantization for edge devices.

### *3.6. Evaluation Metrics*

Model performance was evaluated using accuracy (Equation (3)), sensitivity (Equation (4)), specificity, and F1-score, with the number of undetected seizures reported as an additional clinical indicator. Accuracy reflects overall classification correctness, while sensitivity quantifies the proportion of seizures successfully predicted within the pre-ictal window, as missed seizures directly compromise patient safety. Specificity measures the model's ability to minimize false alarms, and F1-score provides a balanced assessment under class imbalance. A seizure was considered successfully predicted if at least one alarm was generated within the predefined pre-ictal window, consistent with standard practice in seizure prediction literature.

$$Accuracy = (TruePositive + TrueNegative)/(TotalSamples) \quad (3)$$

$$Sensitivity = TruePositive/(TruePositive + FalseNegative) \quad (4)$$

## 4. Results

### *4.1. Experimental Setup and Protocols*

All experiments were conducted under patient-independent evaluation, with EEG data partitioned into training, validation, and test sets as described in Section 3, and results averaged across patients to account for inter-patient variability. Five-fold Monte Carlo cross-validation was employed to improve robustness against random partitioning effects. Three evaluation settings are reported throughout: generalized (patient-independent train–test splits), blind (unseen patients without personalization), and personalized (generalized model adapted using patient-specific data). Performance was assessed primarily using accuracy and sensitivity, with undetected seizure counts reported to evaluate clinical reliability. It should be noted that while false positive per hour is reported for the generalized model at the 5 min horizon as a representative clinical usability indicator, a full systematic analysis of false alarm behavior across all prediction horizons and personalization settings is acknowledged as a limitation of the current study and is identified as a priority for future work.

### *4.2. Generalized Model Performance Across Architectures and Prediction Horizons*

This section presents the performance of all evaluated model architectures when trained and tested under the generalized, population-level paradigm. In this setting, models are trained on the aggregated multi-patient training partition and evaluated on held-out test patients, without any patient-specific adaptation. Two categories of input representation are examined: raw multichannel EEG and a comprehensive all-feature vector comprising statistical, spectral, and wavelet-based descriptors extracted from the EEG recordings. Performance is assessed across ten prediction horizons spanning 5 to 300 min prior to seizure onset, using accuracy and sensitivity as the primary evaluation metrics.

In addition to these primary metrics, the performance of the top-performing models is further evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC), along with complementary metrics including precision, F1-score, and specificity. The AUC analysis provides a threshold-independent assessment of classification performance, offering deeper insight into the models' ability to discriminate between pre-ictal and inter-ictal states under class imbalance conditions.

Figure 7 presents ROC curves and a detailed metric comparison for the top five models at the 5 min prediction horizon. The CNN-LSTM Hybrid achieves the highest AUC (0.973) along with 84.6% $\pm$ 0.81% sensitivity, 95.4% $\pm$ 1.17% specificity, and an F1-score of 92.3% $\pm$ 0.91%, reflecting superior separability and strong false alarm control.

CNN and ANN demonstrate more balanced profiles, with CNN achieving 90.2% ± 1.67% sensitivity and 90.7% ± 0.84% F1-score, and ANN achieving 88.1% ± 2.03% sensitivity and 89.6% ± 0.84% F1-score. DeepConvNet exhibits slightly higher sensitivity (89.9% ± 1.33%) at the cost of lower specificity (87.5% ± 1.51%, F1 88.4% ± 1.87%), while XGBoost achieves the highest specificity among feature-based models (93.6% ± 1.07%) but comparatively lower sensitivity (85.9% ± 1.22%, F1 89.1% ± 1.35%). Collectively, these results confirm the robustness of both feature-based and deep learning approaches at short horizons, while highlighting the CNN-LSTM Hybrid's overall superiority in discriminative capability.

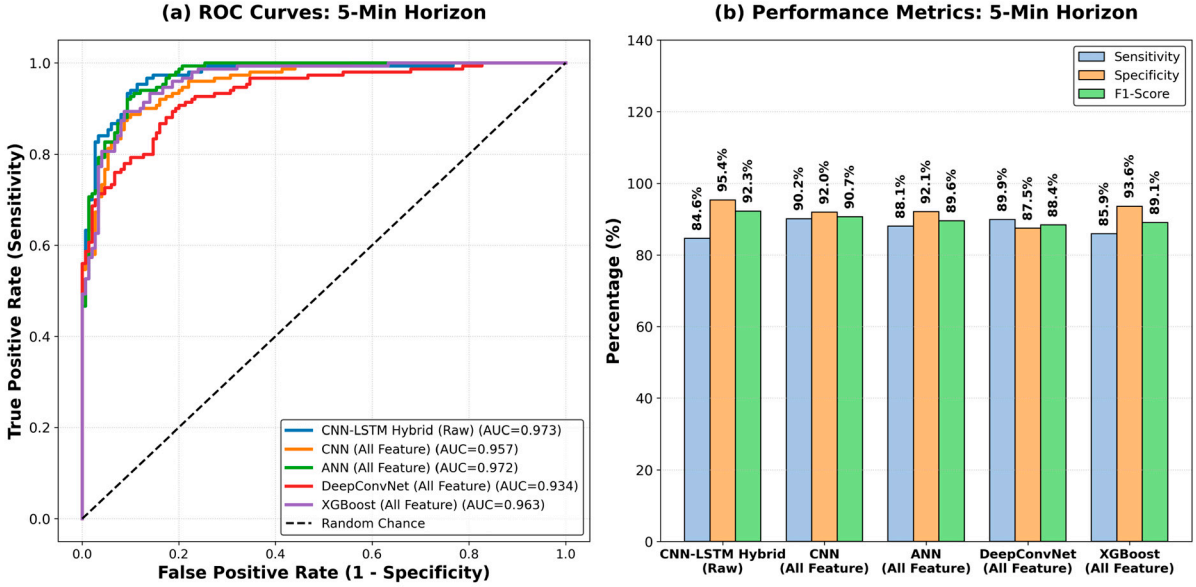



**Figure 7.** Dual-panel performance analysis for the top five models at the 5 min prediction horizon. (**a**) ROC curves showing discriminative capability, with the CNN-LSTM Hybrid achieving the highest AUC (0.973). (**b**) Sensitivity, specificity, and F1-score comparison highlighting architecture-level trade-offs between seizure detection and false alarm control.

Thirteen model configurations are evaluated, spanning classical methods (Random Forest, SVM, Gradient Boosting), ANN, and deep learning architectures (CNN, DeepConvNet, ShallowConvNet, EEGNet, ResNet, CNN-LSTM Hybrid). Classical methods operate on engineered feature vectors, deep learning models are evaluated on both raw EEG and all-feature inputs, and the CNN-LSTM Hybrid is tested exclusively on raw EEG. As illustrated in Figure 8, all models converge to strong performance within the 60 min window, with leading architectures exceeding 90% accuracy and sensitivity, before declining sharply at longer horizons. At short-to-intermediate windows, deep learning architectures consistently outperform classical methods. The CNN-LSTM Hybrid achieved the highest accuracy at 5 min (96.3% ± 0.84%, sensitivity 91.62% ± 1.34%) and 15 min (94.0% ± 0.56%, sensitivity 90.1% ± 0.87%), while DeepConvNet with all-feature input led at 30 min (94.1% ± 1.64%, sensitivity 88.0% ± 1.81%) and 45 min (91.4% ± 1.28%, sensitivity 84.6% ± 2.64%). Notably, Gradient Boosting achieved the highest sensitivity at 15 and 30 min (91.96% ± 1.41% and 88.7% ± 1.94%), suggesting ensemble methods remain competitive where minimizing missed seizures is the priority. Simpler architectures including ShallowConvNet and ResNet lagged consistently across short-to-intermediate windows, likely due to training instability under the imbalanced multi-patient setting.

At 60 min, the CNN-LSTM Hybrid (Accuracy 91.3% ± 1.01%, sensitivity 83.0% ± 1.33%), DeepConvNet all-feature (Accuracy 90.4% ± 0.81%, sensitivity 88.8% ± 1.77%), and Gradient Boosting (Accuracy 90.5% ± 1.51%, sensitivity 81.9% ± 1.33%) all exceeded 90% accuracy, with AUC values of 0.969, 0.975, and 0.933 respectively (Figure 9), confirming

that robust pre-ictal EEG signatures remain detectable at this temporal distance. Beyond 75 min, a sharp decline is observed across all architectures, with only DeepConvNet all-feature maintaining reasonable performance at 78.1%. At 90 min and beyond, all models cluster between 50% and 64% regardless of architecture or input representation, defining a critical temporal boundary beyond which generalized prediction becomes unreliable and reinforcing the need for personalization strategies explored in subsequent sections.

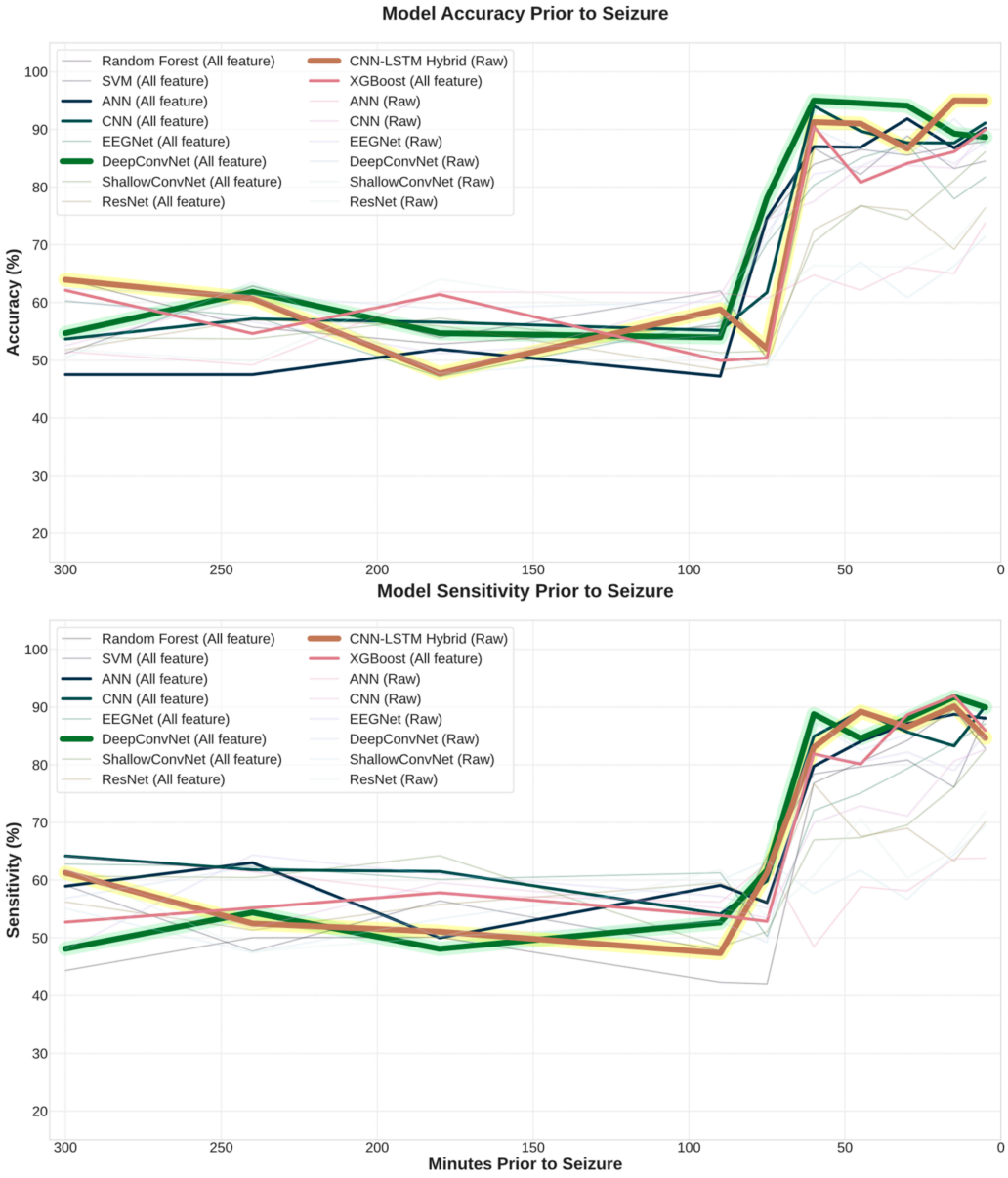



**Figure 8.** Accuracy (**top**) and sensitivity (**bottom**) of all evaluated architectures across ten prediction horizons from 300 to 5 min prior to seizure onset. CNN-LSTM Hybrid (raw) and DeepConvNet (all-feature) consistently achieve the highest performance across clinically relevant short-to-intermediate horizons, with all models converging toward chance level beyond 75 min.

Across all ten prediction windows, the CNN-LSTM Hybrid emerged as the best overall architecture, ranking first at the 5, 15, and 60 min horizons with a mean accuracy of 74.1%. Its strong short-horizon performance is particularly clinically significant, as 5 to 15 min prediction represents the most actionable window for real-time intervention. DeepConvNet with all-feature input achieved a higher mean accuracy of 75.7% due to stronger intermediate-horizon performance, but the CNN-LSTM Hybrid's superiority at clinically

critical short horizons makes it the preferred reference model for subsequent personalization and optimization experiments. It should be noted that the models evaluated in this section operate on heterogeneous input representations: classical methods operate exclusively on engineered feature vectors, while deep learning models are evaluated on both raw EEG and all-feature inputs, and the CNN-LSTM Hybrid is tested exclusively on raw EEG. Cross-modality comparisons are therefore intended to reflect input representation trade-offs and practical deployment considerations rather than strict architectural benchmarks, and results should be interpreted accordingly.

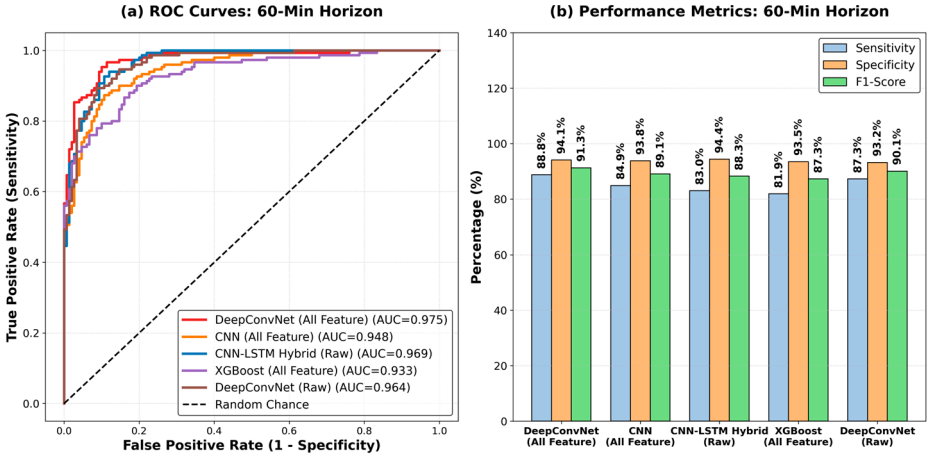



**Figure 9.** ROC curves and metric comparison for top models at the 60 min prediction horizon. DeepConvNet (all-feature) achieves the highest AUC (0.975), with sensitivity, specificity, and F1-score remaining well-balanced across architectures.

### *4.3. Baseline Performance and Blind Personalization Analysis*

To establish a reliable performance reference, the selected CNN–LSTM generalized model was evaluated under strictly patient-independent conditions. Particular emphasis was placed on the 5 min prediction horizon, as this interval provides clinically actionable warning while maintaining forecasting difficulty. At the 5 min horizon, the generalized model achieved an accuracy of 96.30% $\pm$ 0.41% and sensitivity of 91.62% $\pm$ 0.27%. These results demonstrate strong discriminative capability in identifying pre-ictal states shortly before seizure onset. Sensitivity above 90% indicates that the majority of seizures were successfully predicted within the defined pre-ictal window. Table 4 summarizes the performance metrics at the 5 min prediction horizon.

The model maintains high overall accuracy while preserving strong seizure coverage. To further illustrate inter-patient performance dispersion, Figure 10 presents patient-wise accuracy and sensitivity values for all 30 blind subjects.

**Table 4.** Generalized model performance at 5 min prediction horizon.

| Metric | Value |
|---|---|
| **Accuracy** | 96.30% $\pm$ 0.41% |
| **Sensitivity** | 91.62% $\pm$ 0.27% |
| **Specificity** | 97.47% $\pm$ 0.35% |
| **F1-Score** | 93.9% $\pm$ 0.38% |
| **False Positive/Hour** | 0.25 $\pm$ 0.04 |
| **Undetected Seizures** | 4 |

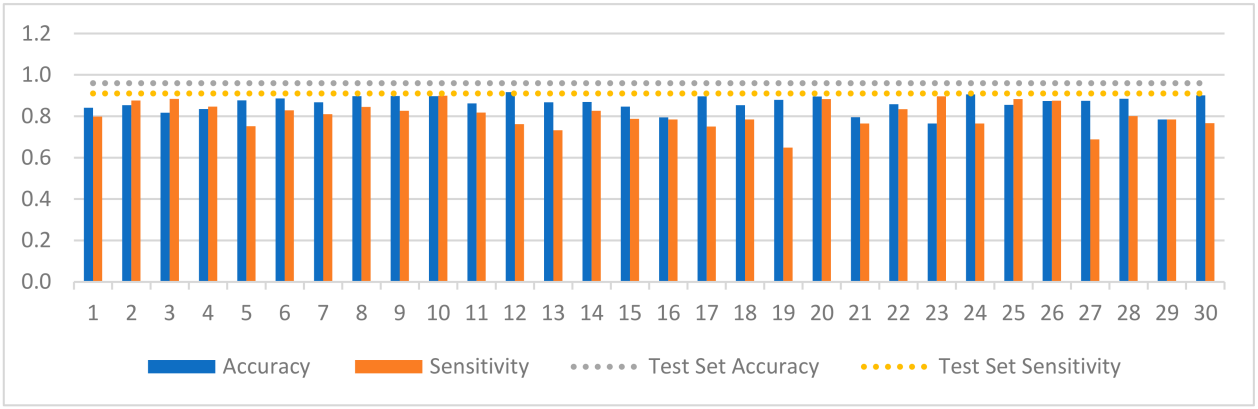



**Figure 10.** Patient-wise accuracy and sensitivity of the generalized CNN–LSTM model evaluated on the 30 blind patients at the 5 min prediction horizon.

To compare overall generalized performance against blind testing performance, Figure 11 presents aggregated metrics. Although the generalized model achieves high predictive performance at the 5 min horizon (Accuracy: 96.30% ± 0.41%, Sensitivity: 91.62% ± 0.27%), blind patient analysis reveals measurable inter-patient variability. This observation confirms that population-level learning alone cannot fully accommodate individual seizure dynamics.

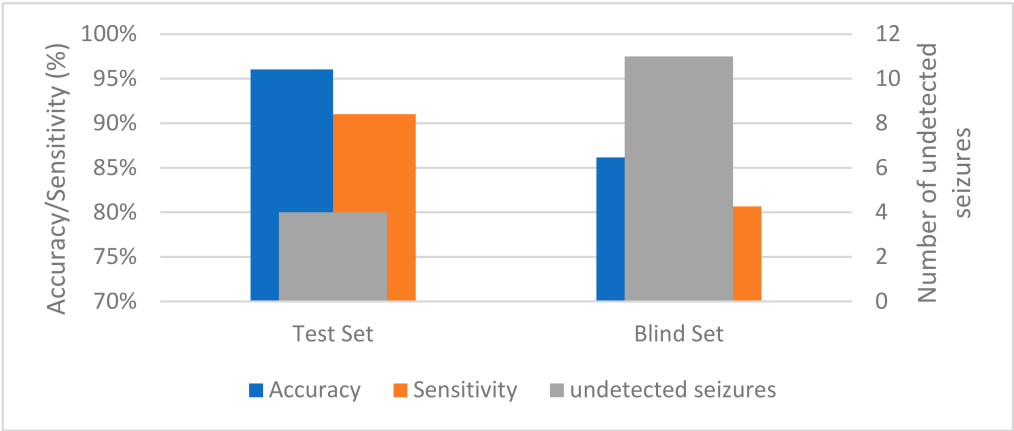



**Figure 11.** Comparison of generalized model performance on the test set and blind patient cohort, showing accuracy, sensitivity, and number of undetected seizures at the 5 min prediction horizon.

Therefore, subsequent sections investigate personalization strategies aimed at reducing missed seizures and stabilizing patient-level sensitivity without increasing false alarm rates.

### *4.4. Results of Incremental Personalization*

To evaluate the impact of progressive patient-specific adaptation, incremental personalization was implemented using a transfer learning framework. The previously trained generalized CNN–LSTM model served as the base model, and patient-specific data were incrementally introduced to fine-tune the network parameters. Rather than training a new model from scratch, the generalized model weights were retained and updated using the newly available patient data, enabling efficient adaptation while preserving population-level knowledge.

These stages simulate realistic clinical deployment, where patient EEG recordings accumulate progressively over time. At each stage, the generalized model was fine-tuned using the specified portion of patient-specific data, while the remaining data were reserved for testing. Performance was then evaluated and compared against the blind

benchmark, defined as the generalized model applied directly to the same patients without any personalization.

At the 5 min prediction horizon, the generalized model achieved an average accuracy of 86.16%, sensitivity of 80.67%, and 11 undetected seizures across the 30 blind patients. These values serve as the baseline reference for quantifying personalization gains. Figure 12 presents the performance across all 30 patients at each personalization stage.

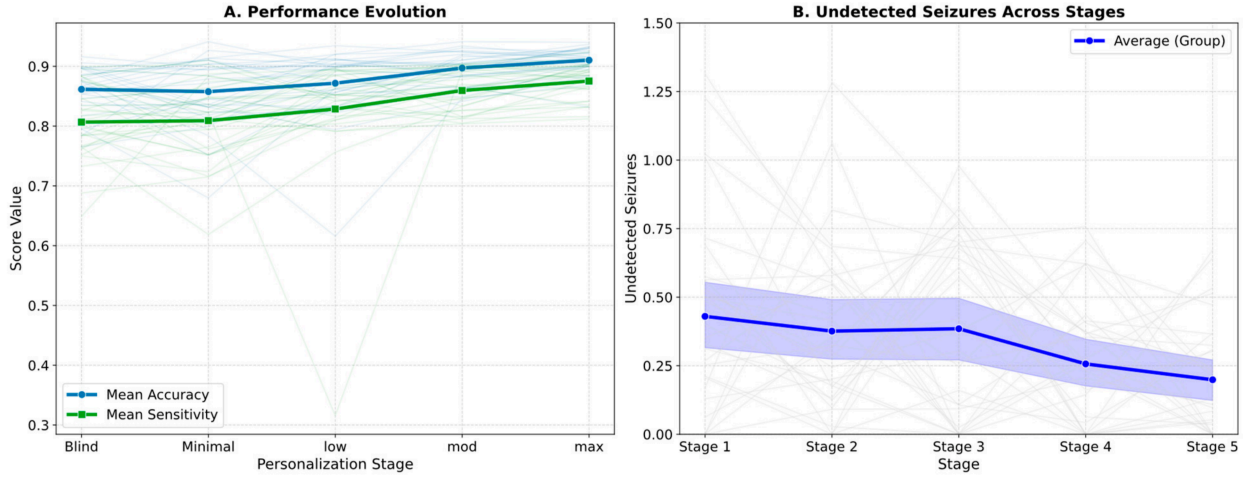



**Figure 12.** Patient-wise evolution of accuracy, sensitivity, and average undetected seizures across the five stages of personalization for all 30 patients. The results demonstrate progressive performance stabilization, improved patient-specific adaptation, and enhanced seizure coverage through successive personalization stages.

The results demonstrate that incremental personalization through transfer learning systematically enhances seizure prediction performance. While minimal personalization (10%) does not consistently outperform the blind model, moderate and maximized stages produce clear gains in accuracy, sensitivity and reduction in undetected seizures. These findings suggest that the generalized model effectively captures population-level seizure dynamics and that fine-tuning with patient-specific data allows the model to adapt to individual electrophysiological signatures without discarding previously learned representations.

### *4.5. Results of Similarity-Based Personalization*

To further enhance patient-specific performance with minimal labeled data from a blind subject, we implemented a similarity-driven personalization strategy based on DTW. Unlike incremental retraining that progressively incorporates patient data, this method restructures the training dataset itself by selecting patients that are morphologically similar to the blind patient before model training. The objective is to construct a compact yet highly representative dataset tailored to the new patient.

The proposed approach consists of two sequential stages. First, one seizure is randomly selected from the blind patient and used as a query template. This seizure is compared against all 130 patients available in the global training set. DTW distance is computed for each pairwise comparison, allowing temporal alignment despite variations in seizure duration or phase shifts. All patients are then ranked in ascending order of DTW distance (lower distance indicates higher similarity). In the second stage, only the top-N most similar patients are selected to construct a new personalized training dataset. Two configurations were evaluated: Top 30 (~20% of the dataset) and Top 70 (~50% of the dataset) similar patients, as shown in Figure 13. For each blind patient, a new model is trained exclusively on the selected subset and then evaluated on that patient's data. Two similarity pool sizes were selected based on elbow analysis of clustering results derived from the DTW

similarity matrix. The elbow point suggested two natural clusters, leading to the definition of narrow and moderate similarity pools corresponding to the Top-30 and Top-70 most similar seizure events.

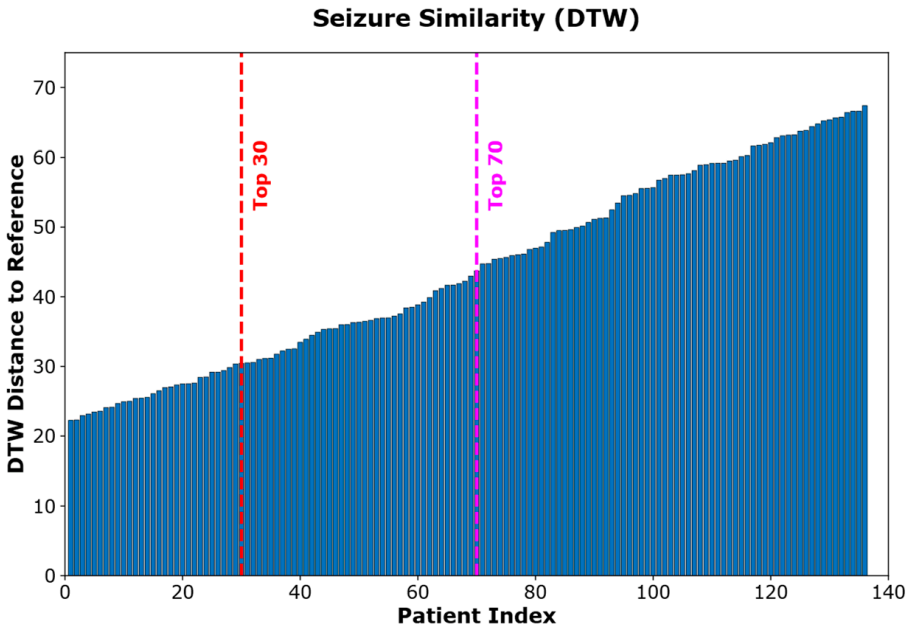



**Figure 13.** Distribution of Seizure Morphology Similarity. Dynamic Time Warping (DTW) distances were calculated for 130 patients against a gold-standard reference. Vertical markers at $n = 30$ and $n = 70$.

When training on the Top 30 most similar patients, the method achieved an average accuracy of 87.38% and average sensitivity of 83.80%. Performance was generally stable across patients, with most cases exceeding 80% sensitivity. However, a small number of outliers slightly reduced the overall average, suggesting that using only 30 samples may occasionally limit robustness.

Expanding the similarity pool to the Top 70 most similar patients improved performance across all metrics. The average accuracy increased to 89.68%, sensitivity improved to 86.70% and number of undetected seizures reduced to 4 seizures only. Compared to the Top 30 configuration, this corresponds to approximate improvements of +2.3% in accuracy and +2.9% in sensitivity, as shown in Figure 14. The broader similarity pool appears to provide a better balance between personalization and variability, enhancing model stability without diluting patient-specific characteristics.

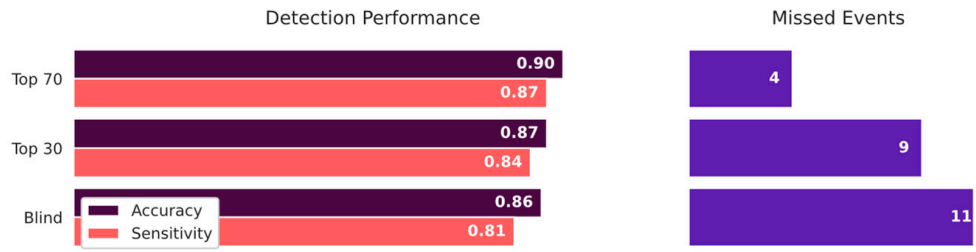



**Figure 14.** Performance comparison between similarity-based personalization strategies using Top-30 and Top-70 patient subsets, illustrating the impact of similarity pool size on accuracy and sensitivity.

The Top-70 configuration offers the most consistent and robust performance, demonstrating that even a single reference seizure is sufficient to construct a high-performing personalized model.

### *4.6. Optimization*

#### 4.6.1. Baseline Performance (Benchmark)

The reference CNN–LSTM detection architecture, depicted in Figure 5, incorporates the complete set of 19 EEG channels alongside comprehensive feature representations derived from 128-sample windows, yielding 132 total inputs. Under this full configuration, the model achieves 96% ± 0.83% accuracy, 91% ± 0.94% sensitivity, and 0 undetected seizures, with 63.4 million trainable parameters and a preprocessing time of 0.002 s. This strong detection performance, however, comes at substantial computational cost, limiting applicability in resource-constrained and real-time scenarios. Consequently, the subsequent enhancement phases focus on reducing input complexity and architectural demands while sustaining reliable detection capability.

Regarding computational complexity of the DTW-based similarity procedure, the pairwise DTW computation between a target patient's reference seizure and all 130 training patients scales as $O(N \times L^2)$, where N is the number of training patients and L is the sequence length. In practice, with the Sakoe–Chiba band constraint applied, this reduces significantly to $O(N \times L \times W)$ where W is the warping window size. For the current dataset configuration, the total DTW similarity computation for a single target patient required approximately 3.20 s on a single CPU, making it feasible for offline personalization prior to deployment. The incremental transfer learning personalization procedure adds minimal overhead relative to full model retraining, as only the final layers are fine-tuned using the patient-specific data subset.

#### 4.6.2. Input Reduction

Pearson's correlation coefficient as shown in Equation (5) was used to quantify redundancy between EEG channels, where elevated absolute correlations indicate shared information content. For an N-electrode recording, a comprehensive inter-channel correlation matrix, shown in Equation (6), was constructed, and channels exhibiting persistently high correlation with multiple others were identified as redundant and prioritized for removal.

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} \tag{5}$$

$$R_{i,j} = \rho X_i X_j \tag{6}$$

Figure 15 displays the correlation structure across EEG channels. Distinct clustering patterns emerge, revealing considerable redundancy among specific electrode locations. Utilizing this analysis, EEG channels are prioritized based on their correlation characteristics, with only the highest-ranked channels preserved at each reduction increment. Corresponding feature representations for retained electrodes are maintained to ensure consistent temporal and spectral characterization. Various reduced configurations are assessed through progressive electrode reduction from the reference 19-channel setup to a minimum of 3 channels. Table 5 presents the corresponding detection performance and computational characteristics.

Interestingly, certain reduced-channel configurations exhibit slight improvements in accuracy compared to the full-channel baseline. This can be attributed to the removal of redundant or noisy channels, which may enhance signal quality and reduce model overfitting. However, these improvements are not consistent across all configurations, suggesting sensitivity to channel selection and underlying data variability. As channel reduction becomes more aggressive, performance degradation becomes evident, confirming the importance of preserving sufficient spatial information for reliable seizure prediction.

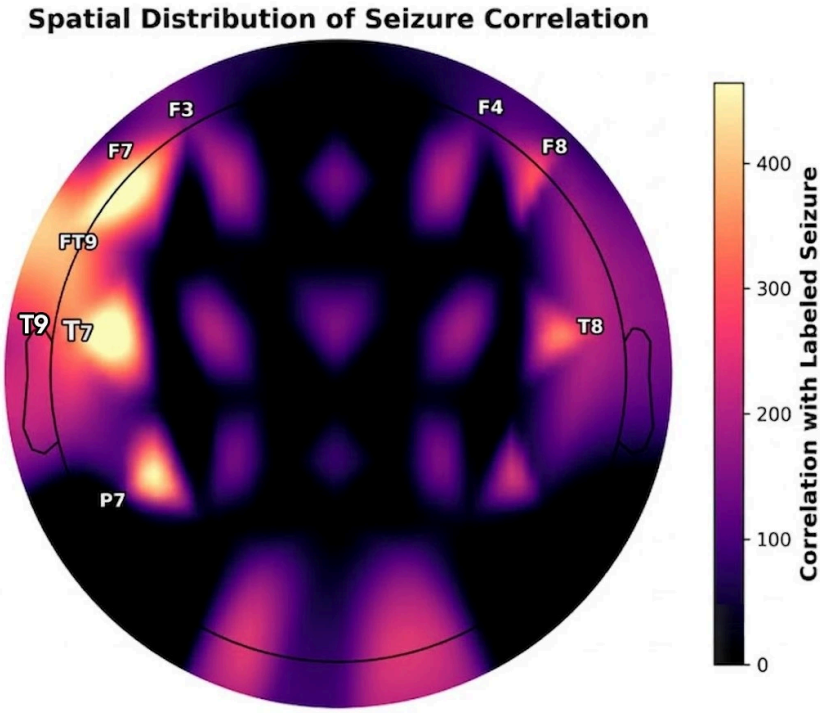



**Figure 15.** Spatial distribution of correlation with the labeled seizure. The topographic map displays the intensity of correlation across the scalp, with peak correlations highlighted by electrode labels. Interpolated using cubic spline methods to ensure a smooth anatomical gradient, the color scale represents the strength of association.

**Table 5.** Performance comparison across reduced EEG channel configurations. Channel reduction from 19 to 10 electrodes maintains accuracy above 90%, while aggressive reduction below 8 channels results in significant performance degradation.

| Inputs | Accuracy | Sensitivity | Undetected Seizures | Pre-Process Time | Inference Time (s) |
|---|---|---|---|---|---|
| **19 EEG + feature** | 0.96 ± 0.83 | 0.91 ± 0.94 | 4 | 0.002 | 0.00005 |
| **15 EEG + feature** | 0.93 ± 1.01 | 0.90 ± 1.03 | 5 | 0.002 | 0.00005 |
| **10 EEG + feature** | 0.89 ± 0.99 | 0.87 ± 2.05 | 7 | 0.002 | 0.00005 |
| **9 EEG + feature** | 0.90 ± 1.04 | 0.88 ± 1.84 | 15 | 0.002 | 0.00005 |
| **8 EEG + feature** | 0.91 ± 0.98 | 0.89 ± 1.37 | 16 | 0.002 | 0.00005 |
| **7 EEG + feature** | 0.88 ± 1.73 | 0.82 ± 3.10 | 11 | 0.002 | 0.00005 |
| **6 EEG + feature** | **0.86** ± 0.91 | **0.86** ± 1.41 | **14** | **0.002** | **0.00005** |
| **5 EEG + feature** | 0.88 ± 1.22 | 0.81 ± 2.64 | 24 | 0.002 | 0.00005 |
| **4 EEG + feature** | 0.83 ± 3.01 | 0.68 ± 3.81 | 61 | 0.002 | 0.00005 |
| **3 EEG + feature** | 0.70 ± 3.48 | 0.66 ± 4.24 | 73 | 0.002 | 0.00005 |

Moderate channel reduction to 8–10 electrodes preserve strong performance, with the 8-channel configuration achieving 91.14% accuracy and 89.17% sensitivity at comparable detection latency to the reference. Below this threshold, performance deteriorates sharply, with configurations of 4 or fewer channels dropping below 70% sensitivity and exhibiting detection latency exceeding 15 s. Figure 16 illustrates the relationship between retained EEG channel count and critical performance indicators, including accuracy and sensitivity. The performance remains stable for moderate channel reductions but deteriorates significantly beyond a critical threshold.

Results reveal a clear performance plateau during moderate channel reduction, transitioning to sharp degradation below the 8-channel threshold. While inference latency differences were minimal under current hardware conditions, reductions in electrode count

substantially decrease data acquisition volume, memory footprint, and sensor-related power consumption, which are critical considerations for wearable and edge deployment.

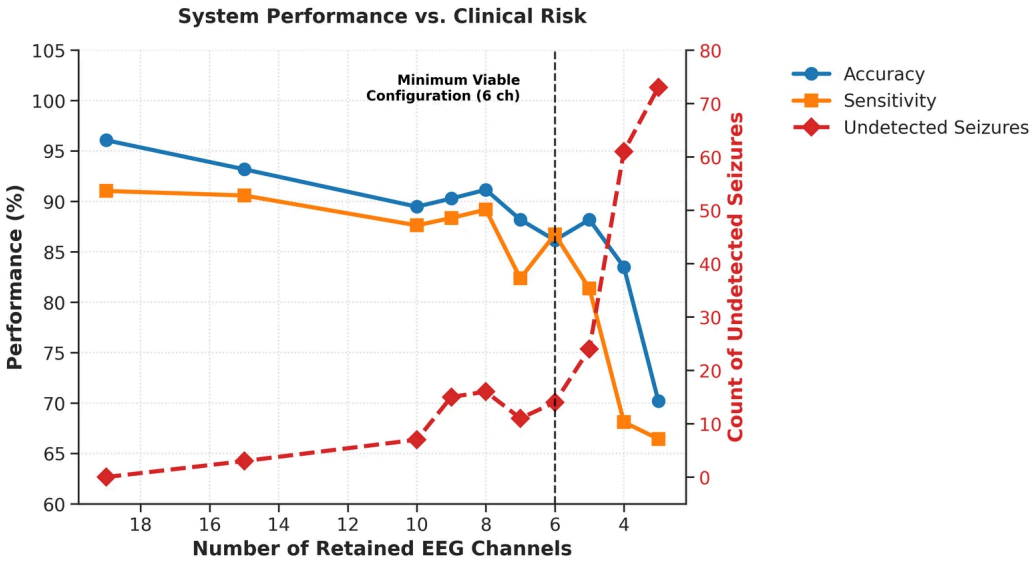



**Figure 16.** Relationship between the number of retained EEG channels and model performance, showing changes in accuracy and sensitivity as electrode count is progressively reduced.

#### 4.6.3. Model Optimization

Three CNN-LSTM configurations (AlexNet, MobileNet v2, and SqueezeNet, each with 128 LSTM nodes) were evaluated under pruning and quantization using MATLAB's model compression framework. Pruning removes low-contribution parameters to reduce computational cost, while quantization reduces numerical precision to decrease memory consumption, together enabling significant reductions in model size and inference time while preserving detection reliability. Table 6 summarizes the performance and computational characteristics of both baseline and compressed models across the three evaluated architectures. The results demonstrate that model compression consistently achieves substantial reductions in network size, parameter count, and training time, with minimal impact on detection accuracy and sensitivity.

**Table 6.** Model Compression Performance Comparison.

| Model | Accuracy | Sensitivity | Undetected Seizures | Total Parameters | Training Time (Min) | Network Size (MB) |
|---|---|---|---|---|---|---|
| **AlexNet** | 96% ± 0.83% | 91% ± 0.94% | 4 | 63,403,512 | 268 | 227.64 |
| **Compressed AlexNet** | 95% ± 0.87% | 92% ± 1.03% | 5 | 57,157,209 | 261 | 212.35 |
| **Mobilenet v2** | 94% ± 1.22% | 90% ± 1.37% | 5 | 3,521,928 | 175 | 14.24 |
| **Compressed MobileNet v2** | 94% ± 1.66% | 91% ± 1.08% | 7 | 3,105,422 | 170 | 13.5 |
| **Squeeze net** | 95% ± 1.94% | 90% ± 2.75% | 8 | 1,235,496 | 138 | 5.02 |
| **Compressed squeezenet** | 93% | 91% | 17 | 1,210,843 | 136 | 4.7 |

The optimized models achieve a substantial reduction in trainable parameters and training time, while maintaining comparable accuracy and sensitivity. This demonstrates that the proposed framework can be adapted for deployment in resource-constrained environments.

### *4.7. Summary of Findings*

Table 7 contextualizes the proposed framework against recent state-of-the-art methods, noting that direct comparison is limited by heterogeneity in datasets, evaluation protocols,

and prediction horizon definitions. The generalized CNN-LSTM achieves 96.30% accuracy and 91.62% sensitivity at 5 min, with accuracy remaining above 91% at 60 min, competitive with or exceeding recent deep learning approaches on comparable datasets. Uniquely, the framework extends evaluation to 300 min prior to onset, and the DTW-based personalization strategy achieves 89.68% accuracy using only 70 morphologically similar training patients, demonstrating effective adaptation without extensive patient-specific data.

**Table 7.** Performance Comparison with Recent State-of-the-Art EEG Seizure Prediction Methods.

| Study | Year | Dataset | Model | Horizon (Min) | Acc % | Sens % |
|---|---|---|---|---|---|---|
| This Work | **2026** | **EPILEPSIAE (161 patients)** | **CNN-LSTM** | **5** | **96.3** ± 2.8 | **91.6** ± 2.1 |
| This Work (Optimized) | **2026** | **EPILEPSIAE (161 patients)** | **CNN-LSTM** | **5** | **89.7** ± 3.7 | **86.7** ± 2.6 |
| This Work (generalized) | **2026** | **EPILEPSIAE (161 patients)** | **CNN-LSTM** | **60** | **91.3** ± 1.9 | **83.0** ± 4.2 |
| **Pontes et al. [41]** | 2024 | EPILEPSIAE (37 patients) | SVM | 50 | NA | 75.0 ± 33 |
| **Batista et al. [28]** | 2024 | EPILEPSIAE (37 patients) | SVM | 55 | NA | 49.0 |
| **Jiang et al. [29]** | 2023 | Siena Scalp EEG | PAC feature extraction + RandomForest | 5–15 | 85.71% | NA |
| **Lopes et al. [42]** | 2024 | EPILEPSIAE (41 patients) | DCAE + BiLSTM | 40 | NA | 0.16 ± 0.23 |

## 5. Discussions

This study investigated personalized EEG-based seizure prediction through a unified framework that integrates generalized learning, incremental transfer learning, and similarity-based data selection. The results demonstrate that while a well-trained generalized model can achieve strong overall performance, meaningful and consistent improvements in seizure prediction require patient-aware adaptation strategies.

### *5.1. Generalized Learning as a Strong Foundation*

The generalized CNN–LSTM model achieved high accuracy and sensitivity at the 5 min prediction horizon, confirming that population-level pre-ictal patterns can be learned effectively. Evaluation on unseen patients further validated the model's generalization capability, indicating that the extracted representations capture shared seizure dynamics across individuals. However, despite strong average performance, notable inter-patient variability was observed, particularly in sensitivity and missed seizure counts. This variability highlights a fundamental limitation of generalized seizure prediction systems, as population-level optimization cannot fully account for the heterogeneous nature of epileptic EEG patterns.

### *5.2. Effectiveness of Incremental Personalization via Transfer Learning*

Incremental personalization using the generalized model as a base and progressively fine-tuning it with patient-specific data produced consistent and monotonic performance gains. Accuracy, sensitivity, detection delay, and composite score all improved as more patient data became available, with the most pronounced gains occurring between the low-resource and moderate stages. These findings suggest that a minimum quantity of patient data is required before effective personalization can occur.

Importantly, detection delay decreased substantially with increased personalization, indicating earlier seizure warnings. This result is clinically significant, as reduced detection delay directly translates to increased reaction time for intervention. The transfer learning approach also proved computationally efficient, as it preserved population-level represen-

tations while adapting selectively to individual EEG characteristics, avoiding the need for training models from scratch.

### 5.3. Personalization Strategies: DTW-Based and Incremental Approaches

The DTW-based similarity personalization approach addressed a different but complementary challenge: personalization under extreme data scarcity. By using only a single seizure from a blind patient to identify morphologically similar seizures in the training set, the method enabled rapid construction of a patient-specific dataset without requiring longitudinal patient data.

Results show that similarity-based personalization significantly improves performance over blind generalization, particularly when expanding the similarity pool from 30 to 70 seizures. The Top-70 configuration consistently achieved higher accuracy, improved sensitivity, and reduced detection delay, indicating that a moderate increase in dataset diversity stabilizes learning while maintaining patient relevance. This confirms that seizure morphology similarity is a meaningful criterion for personalization and that data-centric personalization can be as effective as model-centric adaptation.

While both personalization strategies improved performance, they operate under different assumptions and constraints. Incremental personalization excels when patient-specific data are progressively available, offering the highest overall performance in later stages. In contrast, DTW-based similarity personalization is particularly effective in cold-start scenarios, where little or no patient data exist. Together, these approaches form a complementary personalization pipeline: similarity-based selection enables early personalization, while incremental transfer learning refines performance over time.

### 5.4. Long-Horizon Prediction and Clinical Implications

As the prediction horizon extends, task complexity increases and personalization becomes increasingly important, as generalized models alone prove insufficient for reliable long-horizon forecasting. Clinically, short horizons of 5–15 min are most suitable for real-time intervention such as neurostimulation and patient alerting, intermediate horizons of 30–60 min enable preventive actions such as medication intake or activity adjustment, while horizons beyond 60 min are better suited to risk monitoring than direct intervention. It is important to clarify that the inclusion of prediction horizons extending up to 300 min was not intended to claim reliable seizure prediction at these extended temporal distances. Rather, the primary objective was to systematically investigate the temporal boundaries of EEG-based seizure predictability and identify the point beyond which forecasting becomes non-viable. Our results consistently confirm that this boundary lies at approximately 60–75 min prior to seizure onset, beyond which model performance converges toward chance level across all evaluated architectures and input representations. We therefore consider the identification of this temporal limit to be a meaningful contribution in itself, as it defines a realistic and evidence-based operational window for EEG-based seizure forecasting and guards against overoptimistic claims regarding long-horizon prediction capability. From a deployment perspective, the framework is compatible with edge and wearable devices through its emphasis on optimization and compact personalized datasets, with false positives per hour and undetected seizure counts included to ensure performance gains reflect practical clinical usability.

### 5.5. Limitations

Several limitations should be acknowledged when interpreting the findings. First, the dataset comprises recordings from a single clinical database (EPILEPSIAE), predominantly featuring patients with focal epilepsy. While the cohort is relatively large (161 patients, 1023 seizures), single-database recruitment may introduce biases related to acquisition

protocols and annotation practices, and the learned patterns may not generalize to other epilepsy syndromes or mixed seizure populations. Future work should evaluate the framework on multi-center datasets spanning diverse epilepsy types, and we note that related work from our group on the CHB-MIT dataset provides initial evidence of cross-dataset applicability.

Second, although the optimization pipeline significantly reduces model complexity, personalization procedures introduce additional computational overhead that may require dedicated hardware for real-time deployment. The DTW-based similarity approach also requires at least one representative seizure from the target patient, limiting applicability for newly diagnosed patients, and while the reference seizure was randomly selected and averaged across five independent runs to avoid bias, a formal sensitivity analysis of reference seizure selection remains an important direction for future work. Alternative strategies based on interictal EEG characteristics or demographic features may be necessary in cold-start scenarios. Additionally, while heterogeneous input representations reflect common practice, they introduce a confounding variable in architectural comparisons; future work will investigate unified input pipelines for more controlled benchmarking. Real-device benchmarking on embedded platforms such as Raspberry Pi or NVIDIA Jetson is also identified as a priority to validate practical deployability beyond the desktop hardware conditions reported here.

Third, the current study relies entirely on retrospective analysis, and prospective clinical validation is essential before real-world deployment can be considered. At extended horizons beyond 90 min, prediction accuracy declines toward chance level, confirming that clinically actionable windows realistically lie within the 60 to 75 min range. While false positive per hour metrics were computed for all training configurations and prediction horizons, editorial length constraints required that reporting be limited to the most clinically important models and primary prediction horizon; a comprehensive false alarm analysis across all horizons and personalization settings is identified as a priority for future work. Finally, the current framework does not include a dedicated explainability analysis identifying which EEG channels, temporal regions, or frequency components contribute most strongly to seizure prediction. This is a planned direction within the broader doctoral research, where techniques such as SHAP, Grad-CAM, and attention-based visualization will be systematically investigated and reported in future publications.

## 6. Conclusions

This study presented a unified framework for generalized and personalized EEG-based seizure prediction, jointly addressing personalization, model optimization, and long-horizon prediction feasibility. Beginning from a population-level CNN-LSTM baseline, systematic personalization through incremental retraining and DTW-based similarity adaptation consistently improved performance across patients. Through architecture simplification, channel reduction, and feature compression, the framework achieved substantial reductions in model complexity, reducing electrode count from 19 to as few as 6 channels with only marginal performance degradation, while the optimized personalized models retained accuracy up to 91.0% and sensitivity up to 89.2%, with configurations suitable for edge and wearable deployment. Analysis across ten prediction horizons confirmed that reliable prediction is achievable within clinically actionable windows of 5 to 60 min, while performance converges toward a chance level beyond 90 min, defining a practical temporal boundary for EEG-based forecasting. Together, these findings provide a scalable pathway for translating seizure prediction systems from research environments to real-world clinical and ambulatory settings.

**Author Contributions:** Conceptualization, K.A.; methodology, K.A.; software, K.A.; validation, K.A.; formal analysis, K.A.; investigation, K.A.; resources, K.A.; data curation, K.A.; writing—original draft preparation, K.A.; writing—review and editing, M.E.B. and C.R.; visualization, K.A.; supervision, M.E.B. and C.R.; project administration, M.E.B. and C.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this study were obtained from the EPILEPSIAE Dataset. Due to licensing and copyright restrictions, the dataset cannot be publicly shared by the authors. Access to the data can be obtained directly from the dataset providers under the appropriate licensing terms at https://www.epilepsy.uni-freiburg.de/database (accessed on 24 January 2023).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| CNN | Convolutional Neural Network |
| DL | Deep Learning |
| DTW | Dynamic Time Warping |
| EEG | Electroencephalography |
| ICA | Independent Component Analysis |
| LSTM | Long Short-Term Memory |
| ML | Machine Learning |
| PCA | Principal Component Analysis |
| SHAP | SHapley Additive exPlanations |
| SVM | Support Vector Machine |
| XGBoost | eXtreme Gradient Boosting |

## References

1. Fisher, R.S.; Acevedo, C.; Arzimanoglou, A.; Bogacz, A.; Cross, J.H.; Elger, C.E.; Engel, J.; Forsgren, L.; French, J.A.; Glynn, M.; et al. ILAE Official Report: A Practical Clinical Definition of Epilepsy. *Epilepsia* **2014**, *55*, 475–482. [CrossRef] [PubMed]
2. Nunez, P.L.; Srinivasan, R. *Electric Fields of the Brain: The Neurophysics of EEG*; Oxford University Press: Oxford, UK, 2006.
3. Afsari, K.; El Barachi, M.; Fasciani, S.; Belqasmi, F. A Deep Learning Approach for Real-Time Detection of Epileptic Seizures Using EEG. In *Proceedings of the 2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*; IEEE: Split/Bol, Croatia, 2022; pp. 1–7.
4. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep Learning for Electroencephalogram (EEG) Classification Tasks: A Review. *J. Neural Eng.* **2019**, *16*, 031001. [CrossRef]
5. Beniczky, S.; Wiebe, S.; Jeppesen, J.; Tatum, W.O.; Brazdil, M.; Wang, Y.; Herman, S.T.; Ryvlin, P. Automated Seizure Detection Using Wearable Devices: A Clinical Practice Guideline of the International League Against Epilepsy and the International Federation of Clinical Neurophysiology. *Clin. Neurophysiol.* **2021**, *132*, 1173–1184. [CrossRef] [PubMed]
6. Munch Nielsen, J.; Zibrandtsen, I.C.; Masulli, P.; Lykke Sørensen, T.; Andersen, T.S.; Wesenberg Kjær, T. Towards a Wearable Multi-Modal Seizure Detection System in Epilepsy: A Pilot Study. *Clin. Neurophysiol.* **2022**, *136*, 40–48. [CrossRef] [PubMed]
7. Afsari, K.; El Barachi, M.; Ritz, C. Near-Real-Time Epileptic Seizure Detection with Reduced EEG Electrodes: A BiLSTM-Wavelet Approach on the EPILEPSIAE Dataset. *Brain Sci.* **2026**, *16*, 119. [CrossRef]
8. Birjandtalab, J.; Baran Pouyan, M.; Cogan, D.; Nourani, M.; Harvey, J. Automated Seizure Detection Using Limited-Channel EEG and Non-Linear Dimension Reduction. *Comput. Biol. Med.* **2017**, *82*, 49–58. [CrossRef]
9. Moctezuma, L.A.; Molinas, M. EEG Channel-Selection Method for Epileptic-Seizure Classification Based on Multi-Objective Optimization. *Front. Neurosci.* **2020**, *14*, 593. [CrossRef]

10. Baghersalimi, S.; Teijeiro, T.; Atienza, D.; Aminifar, A. Personalized Real-Time Federated Learning for Epileptic Seizure Detection. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 898–909. [CrossRef]
11. Aldahr, R.S.; Alanazi, M.; Ilyas, M. Addressing Inter-Patient Variability in EEG: Diversity-Enhanced Data Augmentation and Few-Shot Learning-Based Epilepsy Detection. In Proceedings of the 2022 International Conference on Healthcare Engineering (ICHE), Johor, Malaysia, 23–25 September 2022; pp. 1–7.
12. Afsari, K.; Ritz, C.; ElBarachi, M. Deep Learning for EEG Seizure Prediction: Impact of Feature Engineering and Prediction Window. In *Proceedings of the 2025 IEEE International Conference on Signals and Systems (ICSigSys)*; IEEE: New York, NY, USA, 2025; pp. 48–53.
13. Cao, X.; Zheng, S.; Zhang, J.; Chen, W.; Du, G. A Hybrid CNN-Bi-LSTM Model with Feature Fusion for Accurate Epilepsy Seizure Detection. *BMC Med. Inf. Decis. Mak.* **2025**, *25*, 6. [CrossRef]
14. Liu, S.; Wang, J.; Li, S.; Cai, L. Multi-Dimensional Hybrid Bilinear CNN-LSTM Models for Epileptic Seizure Detection and Prediction Using EEG Signals. *J. Neural Eng.* **2024**, *21*, 066045. [CrossRef] [PubMed]
15. Costa, G.; Teixeira, C.; Pinto, M.F. Comparison between Epileptic Seizure Prediction and Forecasting Based on Machine Learning. *Sci. Rep.* **2024**, *14*, 5653. [CrossRef]
16. Jaishankar, B.; Ashwini, A.M.; Vidyabharathi, D.; Raja, L. A Novel Epilepsy Seizure Prediction Model Using Deep Learning and Classification. *Healthc. Anal.* **2023**, *4*, 100222. [CrossRef]
17. Altaf, Z.; Unar, M.A.; Narejo, S.; Zaki, M.A.; Naseer-u-Din. Generalized Epileptic Seizure Prediction Using Machine Learning Method. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2023**, *14*, 502–510. [CrossRef]
18. West, J.; Dasht Bozorgi, Z.; Herron, J.; Chizeck, H.J.; Chambers, J.D.; Li, L. Machine Learning Seizure Prediction: One Problematic but Accepted Practice. *J. Neural Eng.* **2023**, *20*, 016008. [CrossRef] [PubMed]
19. Dong, X.; He, L.; Li, H.; Liu, Z.; Shang, W.; Zhou, W. Deep Learning Based Automatic Seizure Prediction with EEG Time-Frequency Representation. *Biomed. Signal Process. Control* **2024**, *95*, 106447. [CrossRef]
20. Georgis-Yap, Z.; Popovic, M.R.; Khan, S.S. Supervised and Unsupervised Deep Learning Approaches for EEG Seizure Prediction. *J. Healthc. Inf. Res.* **2024**, *8*, 286–312. [CrossRef]
21. Esmaeilpour, A.; Tabarestani, S.S.; Niazi, A. Deep Learning-Based Seizure Prediction Using EEG Signals: A Comparative Analysis of Classification Methods on the CHB-MIT Dataset. *Eng. Rep.* **2024**, *6*, e12918. [CrossRef]
22. Zhu, R.; Pan, W.; Liu, J.; Shang, J. Epileptic Seizure Prediction via Multidimensional Transformer and Recurrent Neural Network Fusion. *J. Transl. Med.* **2024**, *22*, 895. [CrossRef]
23. Parani, P.; Mohammad, U.; Saeed, F. Utilizing Pretrained Vision Transformers and Large Language Models for Epileptic Seizure Prediction. In Proceedings of the 2025 8th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 16–17 February 2025; pp. 132–137.
24. Yuan, S.; Yan, K.; Wang, S.; Liu, J.-X.; Wang, J. EEG-Based Seizure Prediction Using Hybrid DenseNet–ViT Network with Attention Fusion. *Brain Sci.* **2024**, *14*, 839. [CrossRef]
25. Khansari, H.S.; Abbaszadeh, M.; Joonaghany, G.H.; Mohagerani, H.; Faraji, F. Epileptic Seizure Prediction Using a Combination of Deep Learning, Time–Frequency Fusion Methods, and Discrete Wavelet Analysis. *Algorithms* **2025**, *18*, 492. [CrossRef]
26. Upadhyay, P.K.; Kumar, P.; Panda, M.K.; Samantaray, A.K. Seamless Integration for Enhanced Seizure Prediction Using Hybrid-ConvMobileNet on Typhoon HIL. *Sci. Rep.* **2025**, *15*, 45480. [CrossRef]
27. Li, Z.; Yeo, K.; Gifford, W.; Marcuse, L.; Fields, M.; Yener, B. Adversarial Spatio-Temporal Attention Networks for Epileptic Seizure Forecasting. *arXiv* **2025**, arXiv:2511.01275. [CrossRef]
28. Batista, J.; Pinto, M.F.; Tavares, M.; Lopes, F.; Oliveira, A.; Teixeira, C. EEG Epilepsy Seizure Prediction: The Post-Processing Stage as a Chronology. *Sci. Rep.* **2024**, *14*, 407. [CrossRef]
29. Jiang, X.; Liu, X.; Liu, Y.; Wang, Q.; Li, B.; Zhang, L. Epileptic Seizures Detection and the Analysis of Optimal Seizure Prediction Horizon Based on Frequency and Phase Analysis. *Front. Neurosci.* **2023**, *17*, 1191683. [CrossRef]
30. Zhang, X.; Zhang, X.; Huang, Q.; Chen, F. A Review of Epilepsy Detection and Prediction Methods Based on EEG Signal Processing and Deep Learning. *Front. Neurosci.* **2024**, *18*, 1468967. [CrossRef]
31. Shafiezadeh, S.; Marco Duma, G.; Pozza, M.; Testolin, A. A Systematic Review of Cross-Patient Approaches for EEG Epileptic Seizure Prediction. *J. Neural Eng.* **2024**, *21*, 061004. [CrossRef] [PubMed]
32. Jana, R.; Mukherjee, I. Efficient Seizure Prediction and EEG Channel Selection Based on Multi-Objective Optimization. *IEEE Access* **2023**, *11*, 54112–54121. [CrossRef]
33. Saadoon, Y.A.; Khalil, M.; Battikh, D. Machine and Deep Learning-Based Seizure Prediction: A Scoping Review on the Use of Temporal and Spectral Features. *Appl. Sci.* **2025**, *15*, 6279. [CrossRef]
34. Wang, N.; Wang, M.; Zhou, Y.; Liu, H.; Wei, L.; Fei, X.; Chen, H. Sequential Data-Based Patient Similarity Framework for Patient Outcome Prediction: Algorithm Development. *J. Med. Internet Res.* **2022**, *24*, e30720. [CrossRef]
35. Suo, Q.; Ma, F.; Yuan, Y.; Huai, M.; Zhong, W.; Gao, J.; Zhang, A. Deep Patient Similarity Learning for Personalized Healthcare. *IEEE Trans. Nanobiosci.* **2018**, *17*, 219–227. [CrossRef] [PubMed]

36. Wang, N.; Huang, Y.; Liu, H.; Fei, X.; Wei, L.; Zhao, X.; Chen, H. Measurement and Application of Patient Similarity in Personalized Predictive Modeling Based on Electronic Medical Records. *BioMed. Eng. Online* **2019**, *18*, 98. [CrossRef] [PubMed]
37. Li, D.; Yu, A.S.L.; Liu, M. A Comparative Analysis of Patient Similarity Measures for Outcome Prediction. *AMIA Summits Transl. Sci. Proc.* **2025**, *2025*, 270–279.
38. Aljohani, A. Optimizing Patient Stratification in Healthcare: A Comparative Analysis of Clustering Algorithms for EHR Data. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 173. [CrossRef]
39. Parimbelli, E.; Marini, S.; Sacchi, L.; Bellazzi, R. Patient Similarity for Precision Medicine: A Systematic Review. *J. Biomed. Inform.* **2018**, *83*, 87–96. [CrossRef] [PubMed]
40. Aljalal, M.; Aldosari, S.A.; Molinas, M.; Alturki, F.A. Selecting EEG Channels and Features Using Multi-Objective Optimization for Accurate MCI Detection: Validation Using Leave-One-Subject-out Strategy. *Sci. Rep.* **2024**, *14*, 12483. [CrossRef]
41. Pontes, E.D.; Pinto, M.; Lopes, F.; Teixeira, C. Concept-Drifts Adaptation for Machine Learning EEG Epilepsy Seizure Prediction. *Sci. Rep.* **2024**, *14*, 8204. [CrossRef]
42. Lopes, F.; Pinto, M.F.; Dourado, A.; Schulze-Bonhage, A.; Dümpelmann, M.; Teixeira, C. Addressing Data Limitations in Seizure Prediction through Transfer Learning. *Sci. Rep.* **2024**, *14*, 14169. [CrossRef] [PubMed]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.