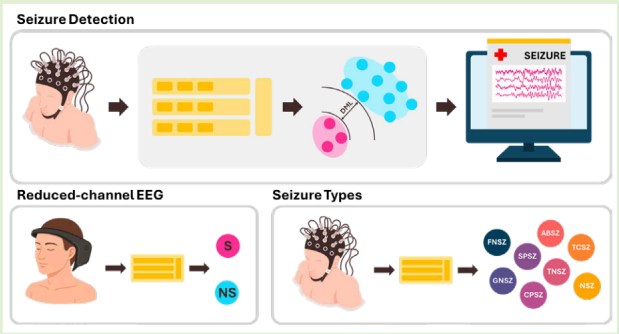
M^2-Net: A Multiscale Multitask Neural Network for EEG-Based Seizure Detection

Wipamas Polpakdee, Phairot Autthasan, Rattanaphon Chaisaen, Phattanun Thabarsa, Supavit Kongwudhikunakorn, *Graduate Student Member, IEEE*, Maarten De Vos, Theerawit Wilaiprasitporn, *Senior Member, IEEE*, Gun Bhakdisongkhram, and Wanumaidah Saengmolee

***Abstract*—Efficient and reliable automated seizure detection systems based on electroencephalography (EEG) can significantly accelerate epilepsy management and reduce the burden on clinicians. However, the complexity of EEG signals and severe class imbalance often require extensive preprocessing and data augmentation, which limit clinical scalability. To address these challenges, we propose M^2-Net, a multiscale, multitask neural network for end-to-end seizure detection. The proposed model learns hierarchical representations directly from EEG signals through a multiscale convolutional architecture, eliminating the need for manual feature engineering. Deep metric learning (DML) is integrated into the framework to enforce well-separated EEG embeddings, which are subsequently leveraged for seizure detection. Experimental results on three benchmark datasets—CHB-MIT, Siena, and EPILEPSIAE—under realistic, highly imbalanced conditions demonstrate that the proposed framework consistently outperforms existing methods, achieving $F1$-scores of 76.12% on CHB-MIT, 51.54% on Siena, and 55.77% on EPILEPSIAE, while maintaining high sensitivity. Moreover, M^2-Net attains competitive performance using only four EEG channels compared with full-scalp EEG configurations, highlighting its potential for reduced-channel monitoring. In addition, the method exhibits robust performance in comprehensive seizure type classification, including rare seizure types and nonseizure (NSZ) background signals. These findings underscore the efficiency, versatility, and clinical viability of M^2-Net for real-world seizure analysis.**



***Index Terms*—Deep metric learning (DML), electroencephalography (EEG), multiscale neural networks, reduced-channel monitoring, seizure detection, seizure type classification.**

Received 4 February 2026; accepted 30 March 2026. Date of publication 8 April 2026; date of current version 14 May 2026. This work was supported in part by PTT Public Company Ltd.; in part by Siam Commercial Bank (SCB) Public Company Ltd.; in part by the Suranaree University of Technology (SUT); in part by the Suranaree University of Technology Hospital (SUTH); and in part by Thailand Science Research and Innovation (TSRI) through the National Science, Research and Innovation Fund (NSRF) under Project 195618, Project 179275, and Project FRB690039/0457. The associate editor coordinating the review of this article and approving it for publication was Dr. Om Jee Pandey. *(Corresponding authors: Wanumaidah Saengmolee; Gun Bhakdisongkhram.)*

Wipamas Polpakdee, Phairot Autthasan, Rattanaphon Chaisaen, Phattanun Thabarsa, Supavit Kongwudhikunakorn, Theerawit Wilaiprasitporn, and Wanumaidah Saengmolee are with the Bio-Inspired Robotics and Neural Engineering Laboratory (BRAIN Lab)-INTERFACES, School of Information Science and Technology (IST), Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong 21210, Thailand (e-mail: wanumaidahs_pro@vistec.ac.th).

Maarten De Vos is with the Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium.

Gun Bhakdisongkhram is with the School of Physical Medicine and Rehabilitation, Institute of Medicine, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand (e-mail: gunbhak@sut.ac.th).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSEN.2026.3680325, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2026.3680325

I. Introduction

EPILEPSY ranks among the most prevalent neurological disorders globally, affecting 50 million across all demographics. Characterized by unprovoked and abnormal bursts of electrical activity in the brain—known as epileptic seizures—this chronic condition significantly impairs patients' cognitive, physical, and social functions [1]. Given its neurological basis, brain monitoring is key for epilepsy management. Electroencephalography (EEG) is a noninvasive technique for recording brain electrical activity, and the resulting electroencephalogram provides neurological insights for diagnosis. Seizure annotation, an initial step following EEG acquisition, relies heavily on expert visual inspection, which is both time-consuming and labor-intensive. To alleviate this burden, research has increasingly turned to automated approaches to distinguish EEG signals during seizure episodes from those during nonseizure (NSZ) stages. Despite these advances, developing generalized seizure detection models remains challenging due to interpatient variability due to age, medication, and seizure type [2]. Consequently, patient-specific EEG-

1558-1748 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

based seizure detection has become a promising direction and constitutes the central objective of this study.

Existing methods based on machine learning (ML) and deep learning (DL) commonly involve extracting seizure-indicative information prior to model input [3], [4], [5], [6], [7], [8]. In ML-based approaches, a small set of predefined features—such as variance or the Hurst exponent—is selected to minimize computational and memory requirements for hardware-constrained applications [3], [4]. However, handcrafted feature extraction requires domain expertise and leads to parameter-sensitive models. Consequently, manual parameter adjustment is needed when applied to new patients [3], [6].

Another critical limitation in the literature is the frequent use of artificially balanced datasets for evaluation. In practice, seizure EEG data are severely imbalanced due to the transient and unpredictable nature of epileptic seizures [9]. Validation on selectively balanced datasets therefore inflates performance estimates and fails to reflect clinical applicability.

The aforementioned challenges underscore the need for scalable methods that can operate effectively under realistic, imbalanced settings without manual intervention. EEG signals during epileptic seizures exhibit discriminative patterns over a wide frequency range and therefore require analysis at multiple temporal resolutions [2]. Multiscale learning, which involves analyzing features at different levels of granularity, has been proven highly effective in computer vision [10]. Inspired by this concept, EEGWaveNet [11] introduced a multiscale convolutional neural network (CNN) that learns seizure dynamics from EEG signals. The model utilizes five convolutional layers with varying receptive fields to generate multiresolution representations that highlight seizure characteristics in deeper layers. The method has shown strong performance on imbalanced datasets in several settings.

Building upon the multiscale framework, TSeizNet [12] proposed a hybrid loss function that combines triplet loss with cross-entropy (CE) loss. As a deep metric learning (DML) strategy, triplet loss encourages discriminative embeddings by enforcing similarity among samples of the same class while maximizing dissimilarity from those of different classes. Under imbalanced conditions, TSeizNet outperformed state-of-the-art methods, including EEGWaveNet, with triplet loss contributing to more flexible decision boundaries. These findings highlight the potential of combining multiscale learning with DML for clinically applicable seizure detection.

This article extends our previous conference paper [12] by systematically investigating multiscale configurations, evaluating diverse DML functions, and conducting more extensive experimental validation. In this work, we propose M^2-Net, an end-to-end multiscale, multitask neural network for EEG-based seizure detection, with a focus on enhancing the practicality of automated detection systems. In contrast to the prior study, which adopted a fixed-depth architecture and a single DML approach, this work makes the following advancements. First, we explore the optimal design of a multiscale structure that learns effective EEG representations and delivers reliable performance without manual preprocessing. Second, we examine the impact of different DML losses within a cooperative objective function. In particular, we demonstrate that the combination of ranked-list loss (RLL) and CE loss significantly enhances the discriminative quality of learned EEG embeddings.

Furthermore, we expand the evaluation to more challenging and clinically relevant scenarios. Since EEG testing involves complex setups and prolonged monitoring which lead to patient discomfort [13], we investigate reduced-channel seizure detection using only four EEG channels aiming to balance detection performance and patient experience. Lastly, recognizing the importance of seizure type identification to treatment planning, we conduct seizure type classification using a database that includes signals from rare seizure types to showcase the boarder potential of our method.

The key contributions of this study are as follows.

1) We propose M^2-Net, an end-to-end multiscale, multitask neural network that integrates DML into multiscale learning for EEG-based seizure detection.
2) We validate the proposed method on three clinically imbalanced datasets—CHB-MIT, Siena, and EPILEPSIAE—without data augmentation, demonstrating its robustness and generalizability in real-world settings.
3) We demonstrate the versatility of M^2-Net in two challenging tasks: a) reduced-channel seizure detection to support unobstructive EEG monitoring and b) comprehensive seizure type classification.

The remainder of this article is organized as follows. Section II introduces the seizure datasets and describes the details of the proposed method. Section III outlines the experimental setup. Section IV presents the experimental results, while Section V discusses the findings, limitations, and directions for future research. Finally, Section VI concludes the article.

II. Materials and Methods

A. Data Acquisition

For seizure detection, including the reduced-channel EEG-based experiment, we employed three scalp EEG datasets, namely CHB-MIT, Siena, and EPILEPSIAE, to cover diverse patient age groups and clinical settings. The CHB-MIT dataset [14], [15], a widely used benchmark recorded at Boston Children's Hospital, consists of data from 24 pediatric patients aged 1.5–22 years. The Siena dataset [16], [17], a comparatively smaller collection from the Unit of Neurology and Neurophysiology, the University of Siena, includes recordings from 17 adult patients aged 20–71 years. The EPILEPSIAE dataset [18], acquired at the University Hospital Freiburg, is a large-scale dataset that contains recordings from 30 patients aged 5–65 years. For these three datasets, only recordings with annotated seizures were included. To ensure sufficient data for patient-specific validation, we selected patients with at least two such recordings. On average, CHB-MIT patients contributed 5.67 $\pm$ 3.00 recordings each (1.36 $\pm$ 0.92 h per recording), Siena patients contributed 3.45 $\pm$ 1.37 recordings (3.06 $\pm$ 2.78 h per recording), and EPILEPSIAE patients contributed 8.60 $\pm$ 3.02 recordings (approximately one hour each, with at least one seizure per recording).

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

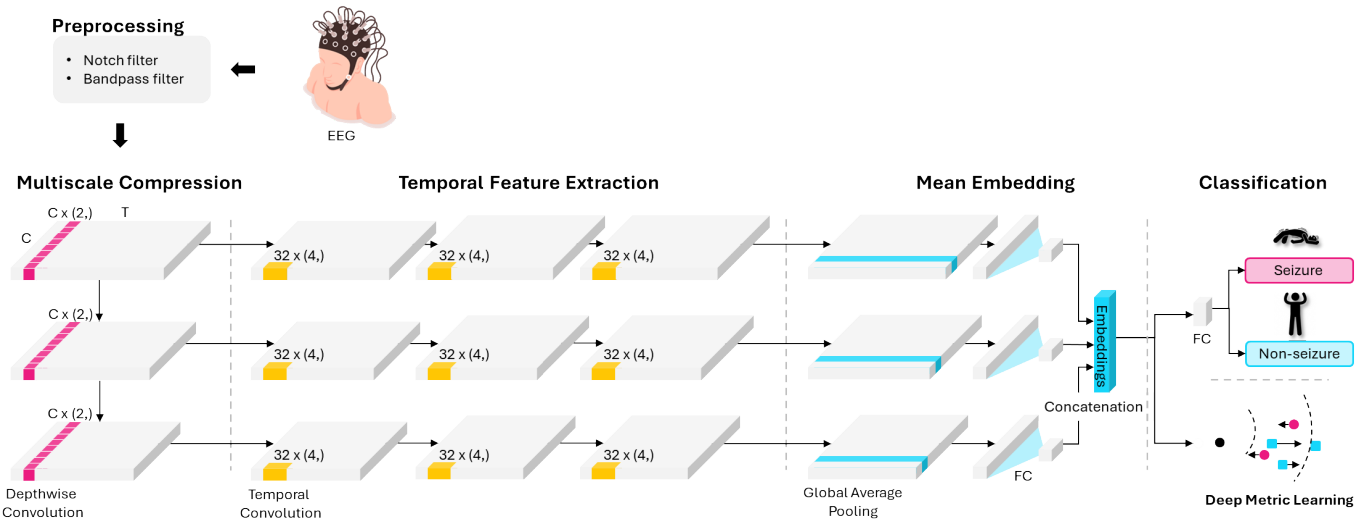


Fig. 1. Overview of the proposed method M^2-Net for seizure detection. After standard preprocessing, an EEG segment is fed to the multiscale compression, temporal feature extraction, and mean embedding modules, which together generate EEG embeddings. The embeddings are then used for classification and are simultaneously supervised by the DML component to enhance feature discrimination.

The TUH EEG Seizure Corpus (TUSZ) [19], from the TUH EEG Corpus [20], was used exclusively for the seizure type classification experiment. The dataset includes scalp EEG recordings from 116 patients with specified seizure types. Although the original dataset is partitioned into training, validation, and test sets, all sets were combined in this study. To reduce variability in recording configurations, only recordings using the averaged reference (AR) montage and containing annotated seizures were selected. NSZ data were defined as EEG segments labeled as "background." The dataset comprises eight classes: focal nonspecific seizure (FNSZ), generalized nonspecific seizure (GNSZ), simple partial seizure (SPSZ), complex partial seizure (CPSZ), absence seizure (ABSZ), tonic seizure (TNSZ), tonic clonic seizure (TCSZ), and NSZ. Class statistics are summarized in Table S1 (Supplementary Material).

All datasets followed the international 10–20 electrode placement system. For consistency, we retained only those recordings containing the most common EEG channels in each dataset. Preprocessing included a 60 Hz notch filter, bandpass filtering between 1 and 64 Hz, segmentation into 4-s windows with 25% overlap, and normalization to zero mean and unit variance. For the EPILEPSIAE dataset, however, the original data were provided with a 1–25 Hz bandpass filter and segmentation into 2-s windows with 50% overlap applied. To harmonize sampling rates, Siena (originally 512 Hz) and TUSZ (250–1000 Hz) recordings were downsampled to 256 Hz after filtering. Table S2 in the Supplementary Material summarizes the characteristics of all datasets.

B. The Proposed Multiscale Multitask M^2-Net

M^2-Net detects EEG segments during epileptic seizures in an end-to-end manner and comprises four key modules, as illustrated in Fig. 1. Additional details on the model configuration are provided in Table S3 of the Supplementary Material.

1) Multiscale Compression: Conventional multiscale CNNs employ parallel branches with heterogeneous kernel sizes to construct pyramid representations of the input [10]. In contrast, we adopt a stacked sequence of 1-D depthwise convolutional layers with small, identical kernels to hierarchically encode the EEG input. This stacking strategy substantially reduces the number of parameters compared to parallel multibranch designs. Similar to kernel factorization, the use of small kernels allows the receptive field to grow gradually across layers, while avoiding the over-smoothing and computational complexity commonly associated with large kernels. Motivated by the spatial dynamics of epileptic seizures, depthwise convolutions operate in a channel-wise manner to preserve channel-specific characteristics of EEG signals. The effectiveness of this design is empirically validated by the ablation study in Section IV-A, which shows performance improvements at an optimal scale depth. The collection of outputs from all layers constitutes the multiscale EEG representations, which retain transient details at shallow scales and global temporal context at deeper scales.

2) Temporal Feature Extraction: After multiscale encoding, high-dimensional features are extracted from the multiscale representations using parallel convolutional branches to maintain architectural simplicity. Each branch consists of three convolutional blocks, which strike a balance between learning capacity and model compactness. Each block comprises a temporal convolution that produces 32 feature maps—the minimum number found to consistently achieve robust performance—followed by batch normalization to stabilize training and a leaky ReLU activation to mitigate vanishing gradients. Branches operating on fine-scale representations emphasize high-frequency components, whereas those processing coarse-scale representations capture low-frequency contents [10]. The resulting feature maps from all branches are then forwarded to the subsequent module.

3) Mean Embedding: In this module, EEG embeddings which encapsulate seizure-indicative information are extracted

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

from the high-dimensional features. Specifically, for each scale, global average pooling is applied to the 32 feature maps to obtain representative features. The pooled features are then independently passed through a linear layer, followed by a sigmoid activation function. The sigmoid function acts as a lightweight gating mechanism that modulates the relative contribution of each feature within a scale, while enabling proportional weighting across scales. The resulting scale-specific embeddings are then concatenated to form the final EEG embeddings representing the input segment.

4) Classification: After multiscale concatenation, the EEG embeddings is fed into a linear layer, followed by a softmax activation function to convert the output into class probabilities.

C. Objective Function

Our multitask approach integrates DML with conventional classification through a cooperative objective function to allow the model to learn a well-structured embedding space while optimizing seizure detection performance. Specifically, the RLL is applied to the EEG embeddings and jointly optimized with the CE loss. The overall objective function is defined as a combination of the DML-based RLL and the CE loss.

For accurate classification, CE loss penalizes the discrepancy between predicted probabilities and ground-truth labels. Due to the inherent class imbalance in clinical seizure EEG data, we employ a weighted variant of CE loss, which is defined as follows:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right] \quad (1)$$

where N denotes the number of samples, $y_i \in \{0, 1\}$ is the ground-truth label for NSZ and seizure classes respectively, $\hat{y}_i$ is the predicted probability of class y_i, and w_{y_i} is the class weight assigned to balance their contribution to the CE loss.

For DML approach, RLL [21] learns an embedding function that, given a query x_i, all positive embeddings from the same class P_i are closer than all negative embeddings from different classes N_i by a margin m, while the negative embeddings $x_j \in N_i$ are pushed beyond a boundary α in the embedding space. The key advantage of RLL is updating based on all nontrivial embeddings, those that violate these conditions, only. This strategy utilizes all impactful embeddings in the mini-batch while ignoring trivial embeddings that might weaken model learning. Additionally, RLL preserves the internal similarity structure of each class cluster by pulling the positive embeddings $x_j \in P_i$ only when they are farther than class radius $\alpha - m$, rather than collapsing them to a single point [21].

Profoundly, RLL is based on a pairwise margin loss

$$\mathcal{L}_m(x_i, x_j) = (1-y_{ij})[\alpha - d_{ij}]_+ + y_{ij}[d_{ij} - (\alpha - m)]_+ \quad (2)$$

where y_{ij} is 1 if $y_i = y_j$ and 0 otherwise, and d_{ij} is the Euclidean distance between embeddings x_i and x_j in the embedding space. RLL also employs a weighting function for positive and negative embeddings

$$w_{ij} = \exp(T_p \times (d_{ij} - (\alpha - m))), \quad x_j \in P_i \quad (3)$$

$$w_{ij} = \exp(T_n \times (\alpha - d_{ij})), \quad x_j \in N_i \quad (4)$$

where T_p and T_n are positive and negative temperatures, respectively, parameters controlling the degree of weighting embeddings. Larger values increase the influence of harder samples, whereas smaller values result in more uniform weighting. The loss contributed by each class is

$$\mathcal{L}_P(x_i) = \sum_{x_j \in P_i} \frac{w_{ij}}{\sum_{x_j \in P_i} w_{ij}} \mathcal{L}_m(x_i, x_j) \quad (5)$$

$$\mathcal{L}_N(x_i) = \sum_{x_j \in N_i} \frac{w_{ij}}{\sum_{x_j \in N_i} w_{ij}} \mathcal{L}_m(x_i, x_j). \quad (6)$$

RLL is a weighted combination of these two components

$$\mathcal{L}_{\text{RLL}}(x_i) = (1-\lambda)\mathcal{L}_P(x_i) + \lambda \mathcal{L}_N(x_i) \quad (7)$$

with λ as the class balancing parameter.

During training, we set the margin and boundary as $m = 1.0$ and $\alpha = 1.5$. The positive and negative temperatures are set as $T_p = 0$ and $T_n = 1.0$, respectively, to emphasize harder negative embeddings while treating positive embeddings evenly. The parameter $\lambda = 0.5$ ensures the equal contribution of embeddings from positive and negative classes.

Finally, the cooperative objective function of M^2-Net evenly combines CE loss and RLL

$$\mathcal{L}_{M^2-Net} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{RLL}}. \quad (8)$$

III. Experiments

The primary objective of this study is to accelerate clinical workflows in epilepsy management by streamlining EEG analysis through the proposed end-to-end framework. In addition to the main seizure detection task, we demonstrate the versatility of M^2-Net through two additional experiments: seizure detection using reduced-channel EEG and seizure type classification. This section describes the experimental setups and evaluation metrics for both the main and additional experiments.

A. Seizure Detection

We evaluate the performance of M^2-Net on patient-specific seizure detection using three imbalanced datasets: CHB-MIT, Siena, and EPILEPSIAE. The evaluation protocol adopts a leave-one-recording-out strategy combined with stratified fivefold cross-validation. Specifically, for each patient, one recording is held out as the test set, and the remaining recordings are split into five stratified folds, preserving class distributions across folds. onefold is reserved for validation, and the remaining four are used for training. This procedure is repeated across all folds and recordings to compute patient-level metrics. The reported scores represent the averaged performance across all patients in each dataset.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

B. Seizure Detection Using Reduced-Channel EEG

Behind-the-ear EEG is a sparse configuration that utilizes two unilateral and two cross-head channels from four electrodes placed behind the ears in the temporal region of the brain. It has demonstrated effectiveness in detecting both temporal and extratemporal focal seizures [22]. Compared to clinical EEG, behind-the-ear EEG exhibits strong correlations in both the time and frequency domains [23], including during seizure events. In long-term home monitoring, behind-the-ear EEG combined with automated detection algorithms captures more seizures than those reported in patient diaries—particularly during episodes of impaired awareness or sleep—while minimally affecting patients' quality of life [24]. Given these advantages, we adopt behind-the-ear EEG as a representative reduced-channel EEG configuration.

To simulate reduced-channel setting, we conduct seizure detection using four temporal EEG channels selected to approximate the placement of behind-the-ear electrodes. The selected channels are *T7–FT9*, *T8–P8*, *P7–T7*, and *P8–O2* for the CHB-MIT dataset, and *T3*, *T4*, *T5*, and *T6* for the Siena and EPILEPSIAE datasets. The training and evaluation procedures follow the main experiment described in Section III-A. Although most studies report event-level results for wearable device applications in outpatient settings, we report at the segment level to better reflect the model's ability to reduce the burden on clinicians during seizure annotation.

C. Seizure Type Classification

In addition to seizure detection, EEG testing also enables seizure type classification, which can support or prompt the reassessment of a physician's diagnosis. However, automated approaches to this task remain underexplored. In this experiment, we investigate seizure type classification using M^2-Net to expand this line of research. In contrast to prior work that focuses only on major seizure types [25], [26], we perform eight-class classification, covering seven seizure types—both major and minor—as well as a NSZ class.

Evaluation on the TUSZ dataset is conducted using stratified fivefold cross-validation with an 80/20 train–test split. onefold is held out for testing, and the remaining data are further split into 80% for training and 20% for validation, with class ratios preserved. This procedure is repeated for each fold, and the results are reported as averaged performance metrics.

D. Network Training

All experiments were implemented using the PyTorch v2.1.2 framework. Model training was performed using an NVIDIA RTX 4000 Ada Generation GPU with 20 GB of memory. During training, the Adam optimizer was used with a learning rate of 0.001 and a batch size of 128, and each model was trained for up to 200 epochs. To prevent overfitting, early stopping with a patience of 10 epochs was applied; training terminated if the validation loss did not improve for 10 consecutive epochs. The model parameters corresponding to the lowest validation loss were retained for testing.

E. Evaluation Metrics

For the binary seizure detection task, model performance was evaluated using accuracy, sensitivity, specificity, $F1$-score, and the area under the receiver operating characteristic curve (AUC). Accuracy represents the overall proportion of correctly classified segments among all seizure and NSZ segments. Sensitivity measures the proportion of seizure segments correctly identified as seizures, whereas specificity measures the proportion of NSZ segments correctly identified as NSZ. The $F1$-score, defined as the harmonic mean of sensitivity and specificity, provides a balanced measure of classification performance. The AUC further quantifies the model's discriminative ability across varying decision thresholds, with values ranging from 0.5, equivalent to random guessing, to 1.0, indicating perfect discrimination.

For the multiclass seizure type classification task, we report both overall accuracy and the weighted $F1$-score. Overall accuracy is computed as the mean of per-class accuracies across all seizure types. The weighted $F1$-score accounts for class imbalance by averaging the $F1$-scores of all classes, with each class weighted according to its contribution in the dataset.

IV. Results

This section presents the experimental performance of M^2-Net for EEG-based seizure detection on the CHB-MIT, Siena, and EPILEPSIAE datasets, along with comparisons to existing methods. We additionally report results under the reduced-channel EEG setting on the three datasets, followed by seizure type classification results on the TUSZ dataset. For benchmarking, three baseline models were reproduced: EEGNet [27], a widely adopted lightweight CNN for EEG analysis; EEGWaveNet [11], the first multiscale CNN for seizure detection; and TSeizNet [12], a recent method applying DML. All baseline models were implemented under the same protocol as the proposed method. Evaluation metrics are reported as mean values with standard deviations (in percentages), with the highest scores highlighted in bold. Given variations in class contributions across studies, statistical analyses were performed only for the reproduced baseline methods. Statistical significance was assessed using a paired t-test against the best-performing model, with * indicating $p \leq 0.05$.

A. Ablation Study

While additional scales can enrich representation learning, it also increases model complexity and may introduce data redundancy, potentially degrading overall performance. We hypothesize that beyond a certain threshold, the inclusion of additional scales yields marginal—or even negative—returns. To examine this effect, we perform seizure detection using the proposed multiscale architecture with the number of scales ranging from one to five, trained using only the CE loss. Each additional scale corresponds to an extra depthwise convolutional layer connected to a dedicated branch of temporal convolutional blocks.

As reported in Table I, the AUC generally improves as the number of scales increases from one to three. However, beyond three scales, performance plateaus, supporting our

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SEIZURE DETECTION PERFORMANCE OF M^2-NET WITH DIFFERENT NUMBERS OF SCALES ON CHB-MIT, SIENA, AND EPILEPSIAE DATASETS (MEAN ± SD)

#Scales	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	AUC (%)
CHB-MIT dataset					
1	98.49 ± 1.38	82.69 ± 12.39*	98.90 ± 1.30	71.47 ± 16.90	97.23 ± 3.17*
2	98.39 ± 1.43	**84.38 ± 10.69**	98.76 ± 1.28*	72.54 ± 16.93	97.42 ± 3.14*
3	**98.57 ± 1.32**	83.93 ± 11.20	**98.97 ± 1.15**	**73.52 ± 15.63**	**97.96 ± 2.58**
4	98.31 ± 1.91	83.22 ± 10.92*	98.74 ± 1.73	73.34 ± 15.66	97.68 ± 3.35
5	98.50 ± 1.36	82.00 ± 12.34*	98.97 ± 1.12	72.65 ± 16.08	97.56 ± 3.28*
Siena dataset					
1	96.74 ± 3.57*	**63.94 ± 23.40**	97.04 ± 3.66*	47.48 ± 25.68	90.30 ± 7.89*
2	98.10 ± 1.74*	63.22 ± 22.51	98.39 ± 1.79*	49.90 ± 25.08	91.93 ± 7.60
3	98.13 ± 2.00	62.54 ± 23.81	98.43 ± 2.05*	49.28 ± 26.32	**93.10 ± 6.34**
4	98.33 ± 1.89	60.41 ± 24.21	98.64 ± 1.93	**51.20 ± 26.48**	92.59 ± 6.44
5	**98.69 ± 1.40**	58.23 ± 24.31*	**99.02 ± 1.37**	51.01 ± 25.74	92.30 ± 6.99
EPILEPSIAE dataset					
1	95.20 ± 3.34*	**67.61 ± 18.18**	95.80 ± 3.06*	46.67 ± 19.89*	90.54 ± 6.53*
2	96.09 ± 2.79*	67.36 ± 18.36	96.72 ± 2.48*	50.51 ± 20.19*	91.68 ± 6.14*
3	96.22 ± 2.77*	67.24 ± 18.86	96.86 ± 2.43*	51.54 ± 20.83	**91.87 ± 6.29**
4	96.19 ± 2.80*	65.82 ± 19.87*	96.87 ± 2.46*	52.02 ± 21.30	91.50 ± 6.69*
5	**96.40 ± 2.72**	64.73 ± 20.10*	**97.10 ± 2.37**	**52.30 ± 21.93**	91.30 ± 6.93*

TABLE II
SEIZURE DETECTION PERFORMANCE OF M^2-NET TRAINED USING CE LOSS (W/O DML) AND USING CE LOSS COMBINED WITH DIFFERENT DML LOSSES ON CHB-MIT, SIENA, AND EPILEPSIAE DATASETS (MEAN ± SD)

DML Loss	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	AUC (%)
CHB-MIT dataset					
w/o DML	98.57 ± 1.32*	**83.93 ± 11.20**	98.97 ± 1.15*	73.52 ± 15.63*	**97.96 ± 2.58**
Contrastive	98.52 ± 1.43*	82.72 ± 11.69*	98.96 ± 1.20*	73.22 ± 14.99*	97.84 ± 2.83
Triplet	98.76 ± 1.30	79.84 ± 12.34*	**99.33 ± 0.72**	74.56 ± 14.18	97.84 ± 2.84*
Center	98.64 ± 1.33	83.16 ± 11.60*	99.07 ± 1.13*	74.34 ± 15.70*	97.96 ± 2.55
Ranked-list	**98.81 ± 1.17**	82.65 ± 11.59*	99.27 ± 0.89	**76.12 ± 14.49**	97.88 ± 2.99
Siena dataset					
w/o DML	98.13 ± 2.00*	**62.54 ± 23.81**	98.43 ± 2.05*	49.28 ± 26.32*	**93.10 ± 6.34**
Contrastive	98.75 ± 0.84	59.70 ± 24.44	99.08 ± 0.80	50.17 ± 24.33	93.07 ± 6.32
Triplet	98.81 ± 1.20	59.32 ± 23.94	99.14 ± 1.21	**51.71 ± 25.47**	91.91 ± 6.99
Center	98.22 ± 2.04	62.27 ± 24.03	98.53 ± 2.09*	50.87 ± 26.51	92.92 ± 6.59
Ranked-list	**98.91 ± 0.78**	57.78 ± 25.63	**99.24 ± 0.76**	51.54 ± 27.21	92.61 ± 7.36
EPILEPSIAE dataset					
w/o DML	96.22 ± 2.77*	67.24 ± 18.86	96.86 ± 2.43*	51.54 ± 20.83*	91.87 ± 6.29*
Contrastive	96.32 ± 3.02*	**68.05 ± 18.33**	96.94 ± 2.75*	53.35 ± 20.76*	**92.38 ± 6.10**
Triplet	96.59 ± 2.80*	67.08 ± 18.63*	97.25 ± 2.45*	54.21 ± 20.30*	91.58 ± 6.30*
Center	96.40 ± 2.66*	66.61 ± 19.23*	97.06 ± 2.34*	52.18 ± 20.94*	91.81 ± 6.27*
Ranked-list	**96.99 ± 2.38**	65.20 ± 19.66*	**97.70 ± 1.98**	**55.77 ± 21.96**	91.64 ± 6.68*

hypothesis. Moreover, the results reveal a clear tradeoff: increasing the number of scales reduces sensitivity while improving specificity. This trend is particularly evident in the Siena and EPILEPSIAE datasets. Specifically, when comparing the one-scale and five-scale configurations, sensitivity decreases by 5.71% and 2.88%, respectively, whereas specificity increases by 1.98% and 1.30%. In addition, given the strong class imbalance in all datasets—where NSZ samples dominate—improved specificity contributes to higher overall accuracy. Nevertheless, the three-scale configuration consistently achieves the highest AUC while maintaining competitive performance in other metrics. Therefore, we adopt the three-scale configuration for all subsequent experiments.

Furthermore, we investigate various DML losses to evaluate their impact on seizure tasks. Specifically, we replace RLL in the cooperative objective function with contrastive loss [28], triplet loss [29], and center loss [30], and compare these variants against training using CE loss without DML as a baseline. Table II shows that although the baseline yields strong sensitivity and AUC, incorporating DML into the training process significantly improves $F1$-scores in all datasets.

Among the DML-based variants, RLL consistently delivers the best overall performance, achieving the highest $F1$-score on CHB-MIT and EPILEPSIAE, the highest specificity on Siena and EPILEPSIAE, and the highest accuracy across all datasets. Notably, the $F1$-score on Siena and the specificity on CHB-MIT achieved by RLL are comparable to the best results obtained using triplet loss. In particular, combining CE with RLL improves the $F1$-score compared to the baseline training by 2.60% on CHB-MIT, 2.26% on Siena, and 4.23% on EPILEPSIAE. To further assess subject-level robustness, Fig. 2 visualizes the per-subject $F1$-scores on the EPILEPSIAE dataset for each loss configuration. The result indicates that integrating RLL outperforms the other configurations for the majority of patients.

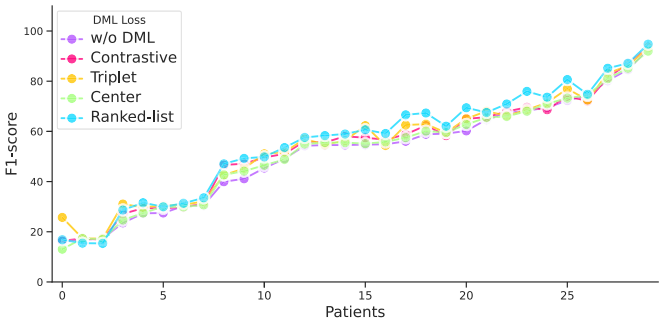


Fig. 2. Patient-wise $F1$-score for seizure detection using M^2-Net trained using CE loss (w/o DML) and using CE loss combined with different DML losses on EPILEPSIAE dataset. Patients are sorted in ascending order of performance to facilitate visual comparison.

B. Seizure Detection

Table III presents a comparative evaluation of the proposed method—configured with three scales and a cooperative objective function combining CE loss and RLL—against existing and reproduced baseline methods for patient-specific seizure detection at the segment level. Although prior studies report near-perfect performance, such results rely on feature extraction and evaluation under balanced datasets, which can lead to overly optimistic results. In contrast, when evaluated under clinically imbalanced conditions, M^2-Net consistently achieves superior accuracy, $F1$-score, and AUC on all datasets.

Specifically, the proposed method achieves an $F1$-score of 76.12% with 82.56% sensitivity on the pediatric CHB-MIT dataset, a 51.54% $F1$-score with 57.78% sensitivity on the adult Siena dataset, and a 55.77% $F1$-score with 65.20% sensitivity on the large-scale EPILEPSIAE dataset. Per-patient detection results are provided in Tables S4–S6 in the Supplementary Material. Notably, on the CHB-MIT dataset, 15 out of 24 patients achieve sensitivity exceeding 80%, underscoring the robustness of the proposed approach across diverse patient cases. Although EEGNet attains higher sensitivity, M^2-Net delivers substantially stronger $F1$-scores. For example, on CHB-MIT, the $F1$-score increases to 76.1% compared

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

TABLE III
SEIZURE DETECTION PERFORMANCE OF EXISTING METHODS AND M^2-NET ON CHB-MIT, SIENA, AND EPILEPSIAE DATASETS (MEAN ± SD)

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	AUC (%)	Class Ratio	Imbalanced Handling	Validation Method
CHB-MIT dataset								
EMD-CSP + SVMs [3]	97.49	97.34	97.50	–	–	1:1	SW	5-fold CV
EMA-GHE + RF [4]	99.9	95.2	99.9	96.3	97.6	1:1	SMOTE	5-fold CV
Persistent Homology + GoogLeNet [5]	97.05 ± 3.89	96.71 ± 4.59	97.38 ± 3.65	97.02 ± 3.95	–	1:1	–	70/30 train-test
Path Signature + Bi-LSTM [6]	99.09	99.28	98.95	98.89	–	–	SW	5-fold CV
Group CosCNN [8]	97.54	97.70	97.54	–	97.67	1:1	SW	Hold-out
EEGNet [27]	95.85 ± 2.69*	**86.26 ± 11.45**	96.19 ± 2.72*	52.97 ± 21.10*	96.66 ± 3.80*	1:60	CW	LOOCV + 5-fold CV
EEGWaveNet [11]	96.68 ± 3.39*	78.17 ± 16.00*	97.20 ± 3.25*	60.05 ± 20.62*	96.03 ± 5.25*	1:60	CW	LOOCV + 5-fold CV
TSeizNet [12]	98.74 ± 1.32	76.28 ± 18.06*	**99.38 ± 0.67**	72.04 ± 17.87*	97.46 ± 2.99	1:60	DML	LOOCV + 5-fold CV
M^2-Net (Ours)	**98.81 ± 1.17**	82.65 ± 11.59*	99.27 ± 0.89	**76.12 ± 14.49**	**97.88 ± 2.99**	1:60	DML	LOOCV + 5-fold CV
Siena dataset								
Persistent Homology + GoogLeNet [5]	96.41 ± 2.91	95.23 ± 4.22	97.60 ± 2.61	96.34 ± 3.01	–	1:1	–	70/30 train-test
EEGNet [27]	93.79 ± 4.19*	**69.80 ± 16.62**	94.02 ± 4.27*	30.42 ± 19.71*	90.72 ± 5.86*	1:190	CW	LOOCV + 5-fold CV
EEGWaveNet [11]	97.17 ± 1.90*	54.38 ± 25.34*	97.50 ± 1.92*	33.88 ± 21.52*	89.16 ± 9.65*	1:190	CW	LOOCV + 5-fold CV
TSeizNet [12]	98.83 ± 1.53	51.09 ± 26.25*	99.23 ± 1.56	48.47 ± 27.61	**92.65 ± 6.52**	1:190	DML	LOOCV + 5-fold CV
M^2-Net (Ours)	**98.91 ± 0.78**	57.78 ± 25.63*	**99.24 ± 0.76**	**51.54 ± 27.21**	92.61 ± 7.36	1:190	DML	LOOCV + 5-fold CV
EPILEPSIAE dataset								
EEGNet [27]	94.97 ± 3.69*	**65.65 ± 20.41**	95.59 ± 3.58*	46.47 ± 21.57*	90.84 ± 6.57*	1:44	CW	LOOCV + 5-fold CV
EEGWaveNet [11]	95.27 ± 3.25*	64.77 ± 19.69	95.93 ± 2.95*	45.81 ± 21.35*	88.99 ± 7.92*	1:44	CW	LOOCV + 5-fold CV
TSeizNet [12]	96.89 ± 2.79	62.66 ± 21.23*	97.64 ± 2.44	54.24 ± 22.22*	91.00 ± 7.04*	1:44	DML	LOOCV + 5-fold CV
M^2-Net (Ours)	**96.99 ± 2.38**	65.20 ± 19.66	**97.70 ± 1.98**	**55.77 ± 21.96**	**91.64 ± 6.68**	1:44	DML	LOOCV + 5-fold CV

SW: Sliding window; SMOTE: synthetic minority over-sampling technique; CW: class weighting; DML: deep metric learning.

with 52.9% for EEGNet, while simultaneously improving all other evaluation metrics. Similar trends are observed on the Siena and EPILEPSIAE datasets, further confirming the superiority of the proposed method.

Additionally, Table III highlights the effectiveness of incorporating DML. Both TSeizNet and M^2-Net, which explicitly employ embedding-level supervision, demonstrate marked improvements over traditional class-weighting strategies such as EEGNet and EEGWaveNet. These findings suggest that embedding-level regularization yields more discriminative representations, particularly under the highly skewed class distributions encountered in practice (e.g., seizure-to-NSZ ratios of approximately 1:60 in CHB-MIT, 1:190 in Siena, and 1:44 in EPILEPSIAE).

C. Seizure Detection Using Reduced-Channel EEG

We further evaluate M^2-Net for seizure detection using reduced-channel EEG, simulating resource-constrained settings. The results are summarized in Table IV. Compared with other models, the proposed method achieves the highest $F1$-score, indicating more reliable detection when only a limited number of EEG channels are available. The Deep SHAP + LSTM model [31], which employs cost-sensitive learning (CSL) to address class imbalance, reports a higher sensitivity of 92.42% on the CHB-MIT dataset; however, it requires explicit feature extraction in time, frequency, and nonlinear domains.

Overall, the performance trend closely mirrors that observed when using full-scalp EEG, as reported in Section IV-B. Transitioning from full-channel to reduced-channel EEG results in a modest decline in all evaluation metrics for all models. The reduction in $F1$-score is primarily attributable to decreases in both sensitivity and specificity. Notably, the sensitivity degradation is less pronounced on the Siena dataset, whereas the reductions in accuracy and specificity are relatively consistent across datasets.

TABLE IV
SEIZURE DETECTION PERFORMANCE USING REDUCED-CHANNEL EEG OF EXISTING METHODS AND M^2-NET ON CHB-MIT, SIENA, AND EPILEPSIAE DATASETS (MEAN ± SD)

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	AUC (%)
CHB-MIT dataset					
Deep SHAP + LSTM [31]	95.31	92.42	95.32	–	93.87
EEGNet [27]	90.41 ± 5.54*	**80.41 ± 16.41**	90.66 ± 5.43*	34.73 ± 17.80*	92.91 ± 7.14*
EEGWaveNet [11]	93.30 ± 4.81*	71.39 ± 18.66*	93.67 ± 4.66*	40.35 ± 20.08*	91.67 ± 9.19*
TSeizNet [12]	**97.39 ± 2.40**	66.12 ± 22.70*	**98.08 ± 1.92**	54.46 ± 21.81*	94.53 ± 6.32*
M^2-Net (Ours)	96.63 ± 4.58	74.12 ± 17.27*	97.06 ± 4.49*	**58.20 ± 20.37**	**95.24 ± 5.42**
Siena dataset					
EEGNet [27]	89.13 ± 7.40*	**72.30 ± 21.98**	89.26 ± 7.47*	21.30 ± 14.35*	87.37 ± 12.78
EEGWaveNet [11]	94.12 ± 3.91*	52.38 ± 21.53*	94.45 ± 3.89*	23.12 ± 13.32*	85.11 ± 15.23*
TSeizNet [12]	98.07 ± 1.71	45.47 ± 23.37*	**98.53 ± 1.64**	39.54 ± 23.11*	**89.69 ± 10.00**
M^2-Net (Ours)	**98.07 ± 1.63**	56.53 ± 27.13*	98.39 ± 1.60	**45.65 ± 25.13**	88.98 ± 12.06
EPILEPSIAE dataset					
EEGNet [27]	93.58 ± 4.55*	**64.02 ± 21.89**	94.12 ± 4.44*	40.68 ± 21.33*	**90.15 ± 7.16**
EEGWaveNet [11]	94.68 ± 3.66*	61.89 ± 21.10*	95.33 ± 3.33*	42.32 ± 22.02*	88.48 ± 9.31*
TSeizNet [12]	**96.51 ± 3.05**	57.75 ± 23.23*	**97.31 ± 2.54**	48.54 ± 23.38*	89.20 ± 8.73
M^2-Net (Ours)	96.12 ± 3.29*	61.97 ± 21.39	96.83 ± 2.80*	**49.88 ± 23.33**	89.92 ± 8.06

D. Seizure Type Classification

Lastly, we performed multiclass seizure type classification using M^2-Net on the TUSZ dataset, with results presented in Table V. Despite the increased complexity of this task, our framework achieves an average accuracy of 97.65% and a weighted $F1$-score of 90.61%. While the hybrid bilinear model [25] reports a higher $F1$-score of 97.4%, it excludes NSZ data from classification. Given that epileptic seizures are typically brief in duration [32], NSZ segments constitute a considerable portion of clinical EEG recordings; excluding the NSZ class, therefore limits the practical deployment of the model. Similarly, the HDV + CNN-LSTM model [26] reports higher accuracy but is evaluated on a reduced set of six seizure classes, omitting ABSZs and TNSZs—the two least

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

TABLE V
SEIZURE TYPE CLASSIFICATION PERFORMANCE OF EXISTING METHODS AND M^2-NET ON TUSZ DATASET (MEAN ± SD)

Method	Accuracy (%)	F1-score (Weighted) (%)	Classes	Contain Non-seizure
TUSZ dataset				
Hybrid Bilinear [25]	–	97.4	FNSZ, GNSZ, SPSZ, CPSZ, ABSZ, TNSZ, TCSZ, MYSZ	
HDV + CNN-LSTM [26]	98.8	98.8	FNSZ, GNSZ, SPSZ, CPSZ, TCSZ, NSZ	✓
M^2-Net (Ours)	97.65 ± 2.88	90.61 ± 1.13	FNSZ, GNSZ, SPSZ, CPSZ, ABSZ, TNSZ, TCSZ, NSZ	✓

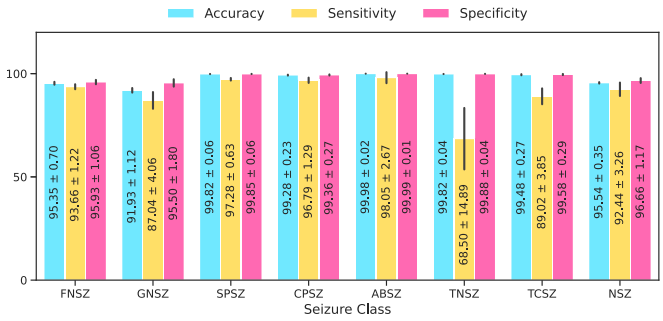


Fig. 3. Class-wise accuracy, sensitivity, and specificity for seizure type classification using M^2-Net on TUSZ dataset.

represented classes in the dataset, with approximately 5 and 4 min of EEG data, respectively.

Furthermore, Fig. 3 illustrates the class-wise performance of the proposed method. The model achieves over 87% accuracy, sensitivity, and specificity for most seizure types. The only exception is the smallest class, TNSZ, which exhibits a sensitivity of 68.50%, due to the limited amount of available training data. Overall, M^2-Net attains an average sensitivity of 90.35% across all classes, demonstrating its effectiveness in handling diverse seizure types.

V. DISCUSSION

This study introduces M^2-Net, an end-to-end framework for patient-specific EEG-based seizure detection. Our approach adheres to rigorous validation protocols on three benchmark datasets without data selection to verify its robustness. Without feature engineering, the multiscale compression and temporal feature extraction modules capture EEG representations at different temporal scales and learn high-dimensional features from EEG signals to simplify the pipeline for clinical integration. The mean embedding module compacts the learned features into EEG embeddings. These embeddings are trained under both classification and DML objectives to enhance discriminability for seizure detection. The proposed framework is evaluated using leave-one-recording-out and stratified fivefold cross-validation strategies to emulate practical use cases, where detection is performed on new recordings after training on all available data from a patient. Under realistic data distributions, the model yields excellent results, with $F1$-scores of 76.12% on CHB-MIT, 51.54% on Siena, and 55.77% on EPILEPSIAE, and sensitivities of 82.65%, 57.78%, and 65.20%, respectively.

A. Realistic Seizure Detection

The goal of automated seizure detection is to replace manual annotation with an efficient and reliable system suitable for clinical use. However, many EEG-based studies attempt to boost model performance by exploiting advanced preprocessing pipelines [3], [4], [5], [6], [7], [8], which often require expert knowledge to fine-tune parameters. Furthermore, reported results are often overly optimistic because of evaluations on balanced datasets—constructed through selective data inclusion [3], [6], [8], biased segmentation strategies [3], [6], [8], or synthetic augmentation [4]. In addition, many validation protocols risk data leakage, where training and test samples are taken from the same recording; this introduces bias into model learning and fails to reflect real-world scenarios [33].

Among the methods evaluated under the imbalanced settings, EEGNet reaches high sensitivity but suffers from low specificity and a reduced $F1$-score. This suggests EEGNet frequently misclassifies NSZ segments as seizure segments, resulting in a high false positive rate. Such errors undermine clinical utility by increasing the operational burden on clinicians and reducing trust in the automated system.

Prior multiscale study has suggested that seizure patterns are emphasized at deeper scales [11]. However, compared to EEGWaveNet and TSeizNet—which both employ five scales—M^2-Net reaches distinguished performance using only three scales. Our investigation into the impact of multiscale depth (Section IV-B) reveals that increasing the number of scales improves AUC, confirming that multiscale representations enhance the model's discriminative power. Nevertheless, we also observe diminishing returns beyond a certain depth. A tradeoff between sensitivity and specificity emerges, likely because deeper scales capture global temporal patterns that introduce redundancy and give less informative embeddings. For our architecture, a three-scale configuration offers the best balance of sensitivity and specificity while avoiding unnecessary complexity.

To further illustrate the multiscale representations, Fig. 4 depicts EEG embeddings extracted from five-scale configured M^2-Net for seizure and NSZ segments. At scales 1–3, seizure embeddings are highly distinguishable, reflecting their ability to capture meaningful features from spectrally rich representations. In contrast, NSZ embeddings at shallow scales—dominated by background brain activity—show poor separability. At scales 4–5, which are derived from low-resolution representations, the discriminative power of seizure embeddings diminishes, indicating the limited utility of deeper scales. Meanwhile, NSZ embeddings become more separable, suggesting that background brain activity is more strongly represented at deep scales. This shift biases the model toward specificity at the expense of sensitivity. Taken together, these observations support our design choice of using three scales and highlight the efficiency of its architecture.

Moreover, the experiment with multitask objective functions shows that combining DML with the classification objective, CE loss, enhances the discriminative capacity of the proposed framework. Among the DML strategies evaluated, RLL achieves the best overall performance. Unlike contrastive and

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

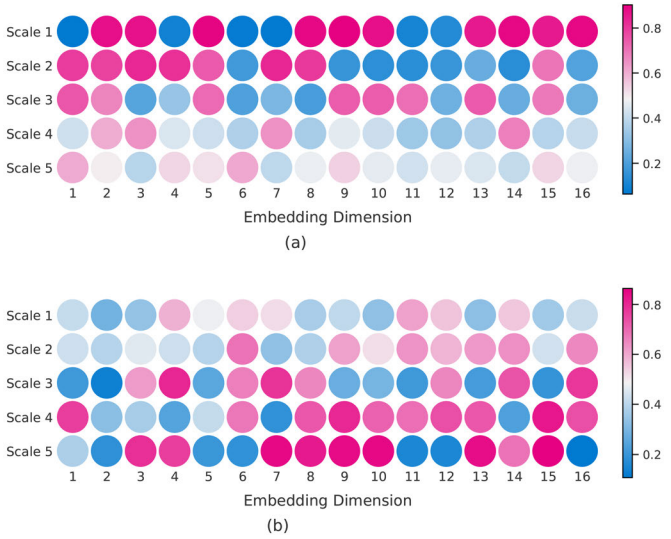


Fig. 4. Heatmap of EEG embeddings extracted from each scale of M^2-Net with five-scale configuration. The embeddings are averaged from (a) seizure and (b) NSZ segments in recording 6 of patient S01 in EPILEPSIAE dataset.

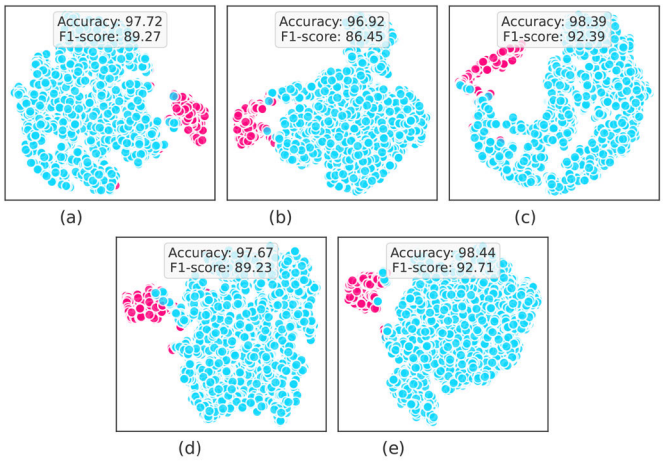


Fig. 5. 2-D t-SNE plots of EEG embeddings for seizure (pink) and NSZ (blue) segments generated by M^2-Net trained using a cooperative function that employs (a) CE loss (w/o DML) and using CE loss combined with (b) contrastive loss, (c) triplet loss, (d) center loss, and (e) RLL. The results are from recording 1 of patient S05 in EPILEPSIAE dataset.

triplet losses, which rely on a limited number of samples and may neglect informative ones, RLL leverages all impactful instances within a mini-batch. This comprehensive use of training data leads to more robust embedding learning.

To visualize the embedding space learned under the under different cooperative objective configurations, we project the high-dimensional embeddings into two dimensions using t-SNE, as shown in Fig. 5. Embeddings trained with RLL form well-separated clusters of seizure and NSZ samples, whereas alternative loss combinations result in more scattered or overlapping distributions. An additional strength of RLL lies in its ability to preserve intraclass similarity, allowing the model to benefit from seizure embedding structures that are specific to individual patients. This advantage is further evidenced in Fig. 2, where RLL consistently achieves higher $F1$-scores across patients, demonstrating generalizability despite inter-subject variability. Our approach surpasses TSeizNet—which, to our knowledge, is the first seizure framework to incorporate DML. Ultimately, the improvements achieved by both DML-based models, TSeizNet and M^2-Net, underscore the effectiveness of DML for seizure detection.

Fig. S1 in Supplementary Material presents the training and validation loss curves of the baseline models and M^2-Net. The figure shows that our model not only converges more rapidly but also exhibits the most stable training compared to EEGNet, EEGWaveNet, and TSeizNet. Furthermore, the close alignment between training and validation losses suggests minimal overfitting.

B. Multipurpose Capabilities of M^2-Net

Reduced-channel EEG systems promote patient compliance and unobtrusive monitoring, making them suitable for both inpatient and outpatient settings. These systems can rely on predefined configurations [22], [23], [24], [34], [35] or employ channel selection techniques [31]. We evaluated seizure detection using four temporal EEG channels positioned to approximate behind-the-ear placement. Using a fixed channel set, M^2-Net outperformed the baseline methods across all three datasets. While the Deep SHAP + LSTM [31] method reports higher $F1$-scores on CHB-MIT, it uses an average of five channels whose selection process relies on full-channel data for per-patient calibration. In contrast, our approach requires only reduced-channel input and does not depend on channel selection procedures.

Compared to full-channel EEG, using four channels resulted in moderate performance degradation. Although behind-the-ear configurations are proven useful for detecting focal seizures, they may fail to capture seizure activity originating outside of their coverage, for instance, generalized or nontemporal focal seizures. Nevertheless, the observed sensitivity reductions—8.53% on CHB-MIT, 3.23% on Siena, and 1.25% on EPILEPSIAE—remain relatively small, indicating that essential seizure-related features are still retained. These results support our model's robustness to input sparsity and its viability for unobstructive EEG systems for continuous seizure monitoring.

Epilepsy classification is based on seizure duration, motor symptoms, and consciousness impairment [36], all of which are reflected in the temporal and spectral characteristics of EEG signals. Consequently, existing studies transform EEG signals into image-based representations using methods such as fast Fourier transform (FFT) or continuous wavelet transform (CWT), followed by convolutional and recurrent neural networks for classification [25], [26]. The common limitations of prior work are the reliance on complex detection pipelines and the seizure class selection, both of which contribute to incomplete clinical representation.

In this study, M^2-Net was evaluated on seizure type classification that included both rare seizure types and NSZ data from the benchmark TUSZ dataset. The model achieved an accuracy of 97.65%, a weighted $F1$-score of 90.61%, and an average sensitivity of 90.35% within an end-to-end

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

framework. Notably, it detected the underrepresented TNSZs with a sensitivity of 68.5%, despite being trained on only 4 minutes of data. Although the inclusion of NSZ data increases task complexity, it helps reduce false positives and may eliminate the need for a separate seizure detection stage. A similar objective was pursued using variational mode decomposition with a Random Forest classifier [33], which reported a 97.7% $F1$-score for seizure detection and a 96.5% $F1$-score for seizure type classification on the TUH dataset.

C. Computational Complexity and Edge Deployment

M^2-Net is designed to be computationally efficient and lightweight. For full-scalp EEG configurations, the model contains between 33 803 and 35 375 parameters, depending on the number of channels, while in the reduced 4-channel setting the parameter count decreases to 28 694. For all configurations, the corresponding weight memory footprint is approximately 100–150 kB in FP32 precision.

Inference efficiency is evaluated on a CPU platform (AMD EPYC 7282). For a 4-s EEG window sampled at 256 Hz with 21 channels, the average inference latency of a single forward pass is 1.72 ms. This latency is substantially smaller than the typical 1–2 s sliding-window duration used in continuous EEG monitoring, indicating that the proposed model can support real-time seizure detection with negligible computational delay. Due to its compact architecture and minimal computational requirements, the model is suitable for deployment on standard clinical computing platforms, such as bedside monitoring devices.

In addition, the proposed framework facilitates patient-specific fine-tuning. An initial model may be trained using EEG recordings available during hospital admission before subsequent overnight seizure screening. As additional EEG recordings become available, the model can be incrementally fine-tuned to adapt to patient-specific seizure characteristics with limited computational overhead.

The strong performance of M^2-Net under extensive evaluation highlights its potential as an efficient and streamlined solution for seizure EEG analysis in healthcare settings. Nevertheless, a key limitation lies in sensitivity, as misclassifying seizure segments as NSZ—especially in cases with limited seizure data—may result in loss of data utility and missed diagnoses. Future work could explore adaptive loss-weighting strategies to dynamically balance the contribution of DML components during training for further optimization. In addition, developing more generalized models for patient-independent detection and extending toward multipurpose frameworks for diverse seizure-related tasks are important directions for advancing seizure analysis.

VI. Conclusion

This article proposes M^2-Net, an end-to-end multiscale, multitask framework for EEG-based seizure detection. The proposed method employs multiscale convolution to generate multiresolution EEG representations, enabling the extraction of seizure-relevant features without manual feature engineering. To enhance the discriminability of the learned representations, the model integrates DML-based RLL with CE classification loss, encouraging well-separated class embeddings. Experimental results on three datasets demonstrate that M^2-Net achieves strong $F1$-scores with competitive sensitivity, confirming its robustness to interpatient and clinical variability under realistic, imbalanced conditions. In addition, the proposed framework delivers reliable performance in reduced-channel EEG configurations using only four temporal channels and generalizes effectively to comprehensive seizure type classification, including rare and NSZ classes. These results indicate the clinical readiness of M^2-Net for real-world seizure applications.

References

[1] A. Strzelczyk et al., "The impact of epilepsy on quality of life: Findings from a European survey," *Epilepsy Behav.*, vol. 142, May 2023, Art. no. 109179.

[2] N. F. Moran et al., "Epilepsy in the United Kingdom: Seizure frequency and severity, anti-epileptic drug utilization and impact on life in 1652 people with epilepsy," *Seizure*, vol. 13, no. 6, pp. 425–433, Sep. 2004.

[3] C. Li et al., "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 458–467, 2021.

[4] W. Zhao, Y. Wang, X. Sun, S. Zhang, and X. Li, "IoMT-based seizure detection system leveraging edge machine learning," *IEEE Sensors J.*, vol. 23, no. 18, pp. 21474–21483, Sep. 2023.

[5] Z. Wang et al., "Automatic epileptic seizure detection based on persistent homology," *Frontiers Physiol.*, vol. 14, Dec. 2023.

[6] Y. Tang, Q. Wu, H. Mao, and L. Guo, "Epileptic seizure detection based on path signature and bi-LSTM network with attention mechanism," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 304–313, 2024.

[7] C. Lei et al., "Epileptic seizure detection in EEG signals using discriminative stein kernel-based sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[8] G. Liu, S. Ren, J. Wang, and W. Zhou, "Efficient group cosine convolutional neural network for EEG-based seizure identification," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–14, 2025.

[9] S. Wong et al., "EEG datasets for seizure detection and prediction—A review," *Epilepsia Open*, vol. 8, no. 2, pp. 252–267, 2023.

[10] L. Jiao et al., "Multiscale deep learning for detection and recognition: A comprehensive survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 4, pp. 5900–5920, Apr. 2025.

[11] P. Thuwajit et al., "EEGWaveNet: Multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5547–5557, Aug. 2022.

[12] W. Polpakdee, P. Autthasan, and T. Wilaiprasitporn, "TSeizNet: Triplet loss empowered multi-scale CNN for superior EEG seizure detection," in *Proc. IEEE SENSORS*, Oct. 2024, pp. 1–4.

[13] E. Bruno et al., "Wearable technology in epilepsy: The views of patients, caregivers, and healthcare professionals," *Epilepsy Behav.*, vol. 85, pp. 141–149, Aug. 2018.

[14] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Harvard Univ.-MIT Division Health Sci. Technol., Cambridge, MA, USA, 2009.

[15] J. Guttag, "CHB-MIT scalp EEG database (version 1.0.0)," PhysioNet, Tech. Rep., 2010.

[16] P. Detti, G. Vatti, and G. Z. M. de Lara, "EEG synchronization analysis for seizure prediction: A study on data of noninvasive recordings," *Processes*, vol. 8, no. 7, 2020, Art. no. 846.

[17] P. Detti, "Siena scalp EEG database (version 1.0.0)," PhysioNet, Tech. Rep., 2020.

[18] J. Klatt et al., "The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients," *Epilepsia*, vol. 53, no. 9, pp. 1669–1676, Sep. 2012.

[19] V. Shah et al., "The temple university hospital seizure detection corpus," *Frontiers Neuroinform.*, vol. 12, Nov. 2018.

[20] I. Obeid and J. Picone, "The temple university hospital EEG data corpus," *Frontiers Neurosci.*, vol. 10, May 2016.

[21] X. Wang, Y. Hua, E. Kodirov, and N. M. Robertson, "Ranked list loss for deep metric learning," 2019, *arXiv:1903.03238*.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.

[22] M. Bhagubai et al., "The power of ECG in semi-automated seizure detection in addition to two-channel behind-the-ear EEG," *Bioengineering*, vol. 10, no. 4, p. 491, Apr. 2023.
[23] Y. Gu et al., "Comparison between scalp EEG and behind-the-ear EEG for development of a wearable seizure detection system for patients with focal epilepsy," *Sensors*, vol. 18, no. 1, p. 29, Dec. 2017.
[24] J. Macea, M. Bhagubai, V. Broux, M. De Vos, and W. Van Paesschen, "In-hospital and home-based long-term monitoring of focal epilepsy with a wearable electroencephalographic device: Diagnostic yield and user experience," *Epilepsia*, vol. 64, no. 4, pp. 937–950, Apr. 2023.
[25] T. Liu, N. D. Truong, A. Nikpour, L. Zhou, and O. Kavehei, "Epileptic seizure classification with symmetric and hybrid bilinear models," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2844–2851, Oct. 2020.
[26] A. Shankar, S. Dandapat, and S. Barma, "Seizure types classification by generating input images with in-depth features from decomposed EEG signals for deep learning pipeline," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 10, pp. 4903–4912, Oct. 2022.
[27] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 056013.
[28] P. Khosla et al., "Supervised contrastive learning," 2020, *arXiv:2004.11362*.
[29] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[30] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2016, pp. 499–515.
[31] Y. Ding and W. Zhao, "Channel selection for seizure detection based on explainable AI with Shapley values," *IEEE Sensors J.*, vol. 24, no. 16, pp. 26126–26135, Aug. 2024.
[32] P. M. Larsen, S. Wüstenhagen, D. Terney, E. Gardella, H. Aurlien, and S. Beniczky, "Duration of epileptic seizure types: A data-driven approach," *Epilepsia*, vol. 64, no. 2, pp. 469–478, Feb. 2023.
[33] J. Mathew, N. Sivakumaran, and P. A. Karthick, "Variational mode decomposition-based moment fusion for the detection of seizure types from the scalp EEG measurements," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
[34] K. Vandecasteele et al., "The power of ECG in multimodal patient-specific seizure monitoring: Added value to an EEG-based detector using limited channels," *Epilepsia*, vol. 62, no. 10, pp. 2333–2343, Oct. 2021.
[35] K. Vandecasteele et al., "Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels," *Epilepsia*, vol. 61, no. 4, pp. 766–775, Apr. 2020.
[36] R. S. Fisher et al., "Operational classification of seizure types by the international league against epilepsy: Position paper of the ILAE commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, Apr. 2017.



Wipamas Polpakdee received the B.Sc. degree in physics from Mahidol University, Bangkok, Thailand, in 2022. She is currently pursuing the M.Eng. degree in Information Science and Technology with the Bio-Inspired Robotics and Neural Engineering Laboratory (BRAIN Lab), School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand.

Her research interests include AI for healthcare and physiological signal processing and analysis.

Phairot Autthasan, photograph and biography not available at the time of publication.

Rattanaphon Chaisaen, photograph and biography not available at the time of publication.

Phattanun Thabarsa, photograph and biography not available at the time of publication.

Supavit Kongwudhikunakorn (Graduate Student Member, IEEE), photograph and biography not available at the time of publication.

Maarten De Vos, photograph and biography not available at the time of publication.

Theerawit Wilaiprasitporn (Senior Member, IEEE), photograph and biography not available at the time of publication.

Gun Bhakdisongkhram, photograph and biography not available at the time of publication.

Wanumaidah Saengmolee received the Ph.D. degree in physiology program, division of health and applied sciences from the Faculty of Science, Prince of Songkla University, Hat Yai, Thailand, in 2022, with a focus on biomedical signal processing and applied machine learning in Kratom users.

Her research interests include neurophysiology, biomedical signal processing, applied deep learning (DL), and computer vision.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:47:48 UTC from IEEE Xplore. Restrictions apply.