The Open Neuroimaging Journal

DOI: 10.2174/0118744400492718260720045955, 2026, 19, e18744400492718

ISSN: 1874-4400

RESEARCH ARTICLE **OPEN ACCESS**

Biometric Identification based on an EEG Signal Using an Attention-enhanced Bidirectional GRU Network

Olsi Shehu[1,*], Damiana Teliti[1], Bekir Karlik[2] and Jasmin Kevrić[3]

[1] *Department of Information Technology, Faculty of Engineering, Natural and Medical Sciences, International Burch University, Sarajevo, Bosnia and Herzegovina*
[2] *Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Istanbul Esenyurt University, Istanbul, Türkiye*
[3] *Department of Electrical and Electronics Engineering, Faculty of Engineering, Natural and Medical Sciences, International Burch University, Sarajevo, Bosnia and Herzegovina*

Abstract:

Introduction/Objective: Reliable identification of individuals from physiological signals is an active research direction in biometrics. Electroencephalography (EEG) is one such modality because brain signals carry individual-specific patterns. This study investigates whether a deep learning–based EEG identification framework can achieve reliable closed-set individual identification on a 109-subject benchmark under a standard subject-dependent evaluation protocol. Verification scenarios and the associated metrics (FAR, FRR, EER) are out of scope and are not reported. This work proposes and validates an attention-augmented Bi-GRU model for EEG-based individual identification, aiming to achieve high identification accuracy on a large multi-subject public dataset.

Methods: Time-domain features were extracted from 64-channel EEG recordings collected during left-hand motor movement/imagery tasks from 109 subjects. A three-layer Bi-GRU model with a self-attention mechanism was trained for a 109-class identification task using a 70/15/15% subject-dependent train–validation–test split. Training robustness was supported through training-only SMOTE-based class balancing and Gaussian-noise augmentation.

Results: Under the adopted subject-dependent split, the proposed model achieved a mean test accuracy of 96.68% ± 0.25% across three independent training runs (95% Student-t CI 96.04–97.31%), a top-5 accuracy of 99.37% ± 0.15%, and a small validation–test gap of +0.37 ± 0.90 percentage points. Performance was higher than several traditional machine learning and simpler deep learning baselines and was comparable to reported transformer and graph neural network results.

Discussion: The attention mechanism highlighted discriminative temporal patterns from EEG signals; performance under the subject-dependent split indicated that our model was able to maintain stable identification performance across the evaluated recordings.

Conclusion: The attention-enhanced Bi-GRU is able to capture discriminative temporal EEG patterns and shows strong identification performance under the adopted protocol, supporting further investigation of EEG-based biometric identification systems.

Keywords: EEG-based identification, Biometric identification, Bidirectional GRU, Self-attention mechanism, Deep learning, Motor imagery.

Received: February 11, 2026
Revised: May 22, 2026
Accepted: June 03, 2026
Published: July 21, 2026

*Address correspondence to this author at the Department of Information Technology, Faculty of Engineering, Natural and Medical Sciences, International Burch University, Sarajevo, Bosnia and Herzegovina;
E-mail: olsi.shehu@stu.ibu.edu.ba

Cite as: Shehu O, Teliti D, Karlik B, Kevrić J. Biometric Identification based on an EEG Signal Using an Attention-enhanced Bidirectional GRU Network. Open Neuroimaging J, 2026; 19: e18744400492718.
http://dx.doi.org/10.2174/0118744400492718260720045955

1. INTRODUCTION

Biometric identification systems [1, 2] are widely used to assign an identity label to a query sample drawn from a population of enrolled subjects. However, conventional biometric modalities such as face and fingerprint remain vulnerable to spoofing and presentation attacks, which has motivated extensive anti-spoofing research [3, 4]. The present work focuses on closed-set identification (1:N matching) rather than verification (1:1 matching). Among the physiological signals studied for identification, electroencephalography (EEG) has attracted growing attention because individual neural activity patterns can serve as a discriminative biometric trait [5-7].

EEG-based biometric recognition [5, 8] has provided a new, promising direction to supplement traditional biometrics [5]. EEG signals record the unique patterns of electrical activity [9] for their bearer brains with a number of distinct advantages: they are challenging to be reproduced or copied out (inherently non-reproducible), can be monitored in real time for liveness detection, and can also be revoked by means of re-enrollment procedures [10-13].

Deep learning has greatly improved pattern recognition in biomedical signal processing [14], including machine-learning-based analysis of EEG signals [15, 16], and it effectively extracts complex temporal and spatial features from physiological data such as EEG [17, 18]. Deep learning models such as Recurrent Neural Networks (RNNs) [19], in particular Long Short-Term Memory (LSTM) and Gated Recurrent Unit [20] (GRU) [21], have displayed remarkable efficiency in capturing sequential dependencies within the time-series data [22]. Furthermore, bidirectional versions of these architectures [23] account for both positive and negative past and future context (*i.e.*, they are sensitive to future as well as past context), thus presenting a more expressive feature extractor for EEG signals [21].

Attention mechanisms have catalyzed [24] the development of deeper models by allowing them to attend to relevant parts of input sequences, benefiting both performance and interpretation [25]. In particular, self-attention [26] has demonstrated strong performance in modeling long-range dependencies and learning discriminative temporal representations in sequential data [24]. The combination of attention mechanisms with recurrent networks presents a convenient framework for EEG-based identification. This hybrid architecture is especially beneficial for EEG data because: (1) EEG time series are characterized by complex temporal dynamics in which neural patterns move both forward (anticipatory) and backward (reactive) during cognitive tasks-bidirectional GRUs can capture both simultaneously; (2) subject-specific brainwave signatures do not equally distribute over the entire temporal space but localize in discriminative windows of interest-attention mechanisms learn to detect and emphasize those important time segments; and (3) given that EEG responses are affected by high trial-to-trial variability, there is no one-size-fits-all pooling technique for aggregating information over the temporal dimension-self-attention makes this adaptation without compromising explainability.

Large-scale public access datasets, *e.g.*, the PhysioNet EEG Motor Movement/Imagery Dataset (EEGMMIDB [27]), have been standard benchmarks for evaluating analytics of EEG [28] and promote reproducible comparisons of results with published literature [6, 22, 29]. Consisting of 109 healthy individuals and quality (16-bit, 64-channel) EEG data recorded during the performance of hand movement and imagery tasks [22, 30], EEGMMIDB [27] provides a good-sized database that can be utilized for evaluating identification methods among varied populations. Because the dataset is multi-task, we can evaluate identification robustness under various levels of cognitive states.

Numerous questions arise from this research: Can the temporal modeling power of Bi-GRU be successfully applied to EEG signals? Do attention mechanisms help to increase identification performance in EEG-based identification? Is the method suitable for identification in a large, multi-subject cohort?

In this paper, we fill in these gaps by proposing an attention-augmented Bi-GRU architecture designed for biometric identification [5] on the EEGMMIDB [27] dataset. We assume such bidirectional temporal modeling combined with a self-attention mechanism can preserve the distinctive brainwave signatures of different individuals and lead to an identification performance applicable for real-world deployment. We evaluate our method on 109 healthy subjects with different EEG profiles, which enables a broad evaluation of subject-wise identification performance [21, 31, 32].

MC EEG provides complementary information about brain activity [17, 33], bringing the different electrode locations (or channels) to provide partial views of neural processing. The 64-channel montage in EEGMMIDB [27] gives good spatial distribution to extract both localized and distributed patterns that contribute to individual identifiability. Motor imagery and execution EEG have been shown to exhibit common individual-specific patterns of brain activity, which can be used as reliable biometrics [5, 29].

In closed-set EEG-based identification, the metric of interest is multi-class classification accuracy across enrolled subjects. The present model is therefore designed and evaluated as a 109-class identifier; verification metrics (FAR, FRR, EER) are intentionally not reported because no enrolment-*versus*-imposter trial protocol is defined in this study.

The development of EEG-based identification approaches ranges from classical machine learning and feature-engineering pipelines to sophisticated deep learning systems with increasing reported accuracy and robustness [21]. Table **1** summarizes representative studies, including both handcrafted-feature approaches and recent deep learning methods.

Table 1. EEGMMIDB dataset and experimental setup.

Characteristic	Value
Total subjects	109
EEG channels	64 (10-10 system)
Sampling rate	160 Hz
Recording runs per subject	14
Task types	Motor movement/imagery
Training set	7,630 samples (70%)
Validation set	1,635 samples (15%)
Test set	1,635 samples (15%)
Epoch duration	2 seconds (320 samples)
Primary endpoint	109-class person identification

In short, it has been reported in the literature [6] that EEG-based person identification has progressed from classical machine learning and feature-based pipelines to modern deep learning methods with improved reported accuracy [10, 34, 35]. Although traditional approaches relying on predefined high-level EEG features remain relevant because they provide interpretable subject-level descriptors, they are often limited in capturing relationships between high-dimensional, nonlinear EEG signals and identity labels. Recent work spans CNN-based, transformer-based, graph-based, and handcrafted-feature pipelines, reflecting continued exploration of temporal, spatial, and engineered EEG representations [30, 36]. Yet the field still struggles to achieve stable identification performance across subjects [21] while maintaining high overall performance. This gap is partially addressed in our work, which combines bidirectional recurrent processing with self-attention mechanisms and achieves a mean classification accuracy of 96.68% ± 0.25% on the EEGMMIDB [27] benchmark.

2. MATERIALS AND METHODS

In this work, we adopted a complete deep learning model to investigate the usability of attention-augmented Bi-GRU networks for EEG-based biometric identification [5]. The model consisted of four essential parts: frame-wise representation, temporal modeling, attention-based frame aggregation, and multi-class classification for person ID [10].

2.1. Study Design and Dataset

For the current study, we used the PhysioNet EEG Motor Movement/Imagery Dataset (EEGMMIDB) [27], which contains 109 subjects recorded with 64 EEG channels during motor movement and imagery tasks. The dataset is available to the public *via* the PhysioBank repository and has been widely used to validate Brain-Computer Systems (BCIs) as well as EEG analysis algorithms [22, 29]. The data collection was conducted following ethical approval, and the publicly available dataset was subsequently used in the current study to develop and validate the algorithm.

Every subject within EEGMMIDB performed 14 experimental runs with distinct motor tasks: baseline [10] (eyes open/closed), motor movement (fist, hands, and foot opening/closing), and motor imagery [37] (imagining the same movements). EEG data were recorded with the BCI2000 software [38] using 64 electrodes placed following the international 10-10 system. Recordings were sampled at 160 Hz with the signals referred to a common average. The characteristics of the entire dataset are presented in Table **1**.

PhysioNet EEG Motor Movement/Imagery Database: 109 subjects recorded using 64-channel EEG at a sampling rate of 160 Hz. This is a motor execution and imagery task dataset on which 70/15/15% splits were used for training, validation, and testing to perform subject-dependent identification.

Signal preprocessing was performed using well-known EEG techniques [39]. For this, the raw EEG signals were band-pass filtered (0.5 − 40 Hz) using a 4th-order Butterworth filter to considerably remove DC drift and use physiologically relevant bands [14, 40] of delta, theta, alpha, and beta frequency bands from the high-frequency noise. Standard preprocessing steps [41] were carried out using the MNE-Python library [42].

EEG recordings were divided into 2-second-long epochs (320 samples at 160 Hz), offering a large enough temporal window to capture the subject-specific pattern in an efficient manner. Amplitude thresholds [43, 44] were used to remove epochs that were contaminated with artifacts. These led to a dataset with around 100 clean epochs per subject across all runs.

The 70/15/15 training/validation/testing split ensures rigorous experimental methodology. Hyperparameter [45] optimization and early stopping decisions were performed using the validation set, whereas the test set was left untapped until final performance evaluation. This three-way split helps prevent overfitting and supports fair comparison with reported benchmark performance [22]. Stratification ensured that each class was evenly represented across the training, validation, and test subsets.

2.2. Temporal Feature Extraction and Pre-processing

The EEG feature representation retained the temporal order of the recordings, as we ordered the samples in time across all 64 channels. Each input sample is a (320, 64) matrix, which corresponds to the 2 s of 64-channel [27] EEG data. This representation allows learning temporal dynamics in each channel as well as spatial relationships between the electrode arrays.

In order to improve model robustness and avoid over-fitting, we utilized two complementary data augmentation strategies. Firstly, Synthetic Minority Over-sampling Technique (SMOTE) [46] was applied only to the training subset to balance the class ratio [29]. Since SMOTE operates in feature space, each training sample was flattened before SMOTE [46] and reshaped afterward so that class balancing was performed only within the training data. SMOTE was used only as a class-balancing step and is not claimed here as a methodological contribution. Second, small-amplitude Gaussian noise (sigma = 0.05) was added to the training samples to improve robustness [21]. This noise injection procedure is designed to mimic natural EEG variability; however, SMOTE-based interpolation in EEG feature space remains a limitation of the current study.

Z-score normalization was implemented independently on a per-scale basis for the training set to standardize distributions of features (mean=0 and standard deviation=1). Normalization parameters calculated from the training data were used to normalize validation and test sets in a way that prevented data leakage.

2.3. Attention-Enhanced Bi-GRU Architecture

The proposed model consists of three stacked bidirectional GRU layers with batch normalization [47] and dropout [48], followed by a self-attention mechanism and fully connected classification layers. The full model is shown in Fig. (**1**).

The input sequence is then forwarded through three Bi-GRU layers having hidden sizes that gradually decrease (to reach 256, 192, and 128 units per direction). The bidirectional signal flow facilitates the model's capability of capturing forward and backward temporal linkages, which are crucial for uncovering longitudinally subject-specific patterns across different time regimes [24]. Batch normalization [47] is applied across the feature dimension after each Bi-GRU layer for training stability, and dropout (p=0.4) is used for regularization.

An additive (Bahdanau-style) attention layer [24] is then applied to the Bi-GRU outputs along the temporal dimension. We deliberately use a single-head additive attention rather than a multi-head dot-product self-attention, so the Q/K/V dimensions and number of heads of Vaswani-style attention [26] do not apply here. Concretely, each per-timestep hidden state h_t in R^{256} (256 = 2 × 128, the bidirectional concatenation of the third Bi-GRU layer) is mapped by a linear layer Linear(256→128) followed by tanh, and a second linear layer Linear(128→1) produces one unnormalised score per timestep; the scores are normalised with a softmax along the temporal axis to obtain weights α_t, and the temporal context vector is $c = \Sigma_t\ \alpha_t \cdot h_t$ in R^{256}. The implementation matches the AttentionLayer class in the released training script. This additive formulation produces a single set of temporal weights per input epoch, which keeps the attention block lightweight and preserves interpretability.

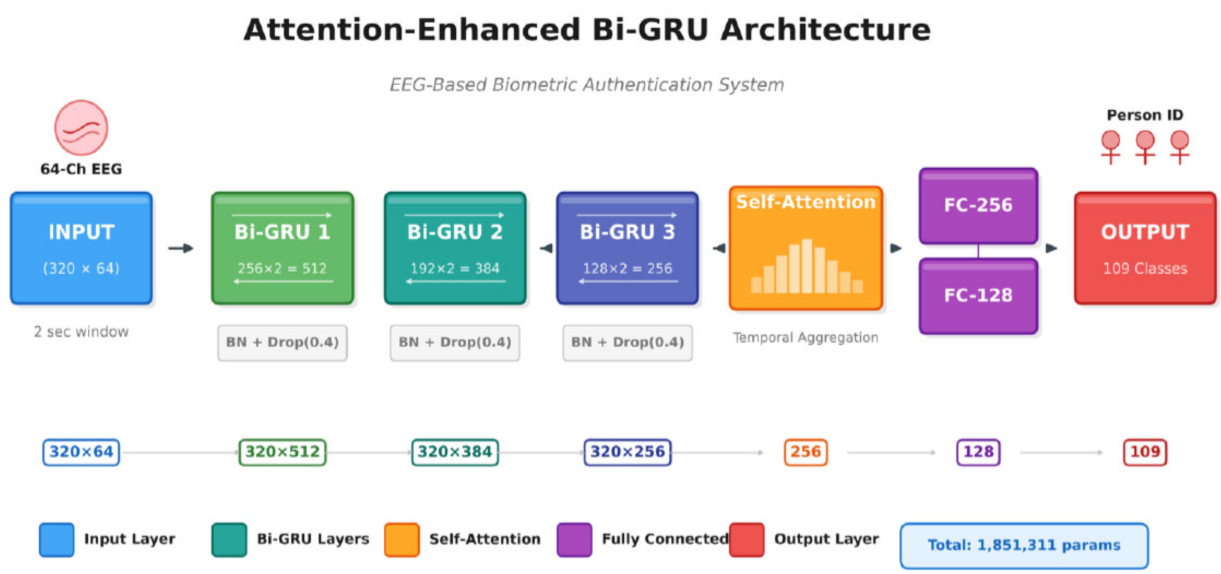

Fig. (1). Attention-enhanced Bi-GRU network for EEG-based biometric identification.

The attended features are then passed through two fully connected layers, each followed by batch normalization [47], ReLU activation [49], and dropout. The first fully connected layer maps the representation to 256 units, the second reduces it to 128 units, and the output layer maps it to the 109 subject classes using softmax activation. Table 2 summarizes the complete layer-by-layer configuration of the proposed architecture.

Table 2. Attention-enhanced Bi-GRU architecture summary.

Layer	Configuration	Output Shape
Input	64 channels x 320 timesteps	(batch, 320, 64)
Bi-GRU 1	256 units x 2 directions	(batch, 320, 512)
BatchNorm + Dropout	p=0.4	(batch, 320, 512)
Bi-GRU 2	192 units x 2 directions	(batch, 320, 384)
BatchNorm + Dropout	p=0.4	(batch, 320, 384)
Bi-GRU 3	128 units x 2 directions	(batch, 320, 256)
BatchNorm + Dropout	p=0.4	(batch, 320, 256)
Self-Attention	Temporal aggregation	(batch, 256)
FC1 + BN + Dropout	256 units	(batch, 256)
FC2 + BN + Dropout	128 units	(batch, 128)
Output	109 classes	(batch, 109)
Total Parameters	-	**1,851,311**

Layer-by-layer specification of the proposed deep learning architecture including three stacked Bi-GRU layers (128, 4, 32 units), self-attention mechanism, and fully connected classification layers with 1.85M total trainable parameters.

2.4. Training Protocol and Performance Evaluation

The model parameters were optimized with the AdamW optimizer [50] with an initial learning rate of 0.001. Other parameters used are L2 weight decay of 0.01, a batch size of 32, and up to a maximum of 150 epochs. To prevent convergence to poor local minima and more effective exploration of the loss landscape, learning rate scheduling was done using a cyclical strategy [51], specifically cosine annealing with warm restarts (T_0=20, T_mult=2). We clipped the gradients (with max_norm=1.0) to avoid exploding gradients, which is a problem associated with such long sequences in deep recurrent architectures.

We applied early stopping with a patience of 25 epochs based on validation accuracy. This strategy helps reduce overfitting and supports convergence toward optimal weights. The multi-seed reproducibility and ablation sweeps reported in Section 3 used a tighter patience of 8 epochs for time efficiency, with no observable degradation in converged validation accuracy.

The main evaluation metrics were the classification accuracy (percentage of correct classifications), the top-5 accuracy (fraction of test examples where the true class was in the top-5 predictions), and a validation-test gap (difference between validation and test accuracy as a measure for performance stability).

3. RESULTS

Across three independent training runs (random seeds 0-2) under the adopted subject-dependent split, the attention-enhanced Bi-GRU model achieved a mean test accuracy of 96.68% ± 0.25% (95% Student-t CI 96.04–97.31%) and a mean top-5 accuracy of 99.37% ± 0.15% on the EEGMMIDB [27] test set, indicating effective and reproducible identification across multiple evaluation metrics.

Practical identification consideration: The mean top-5 accuracy is high (99.37% ± 0.15%), practical for a real-world identification system that could resolve many ambiguous cases by performing a second identification pass. This is because, although predictions of our primary model could be wrong, the true identity usually falls within one of the top results, as shown in Fig. (2).

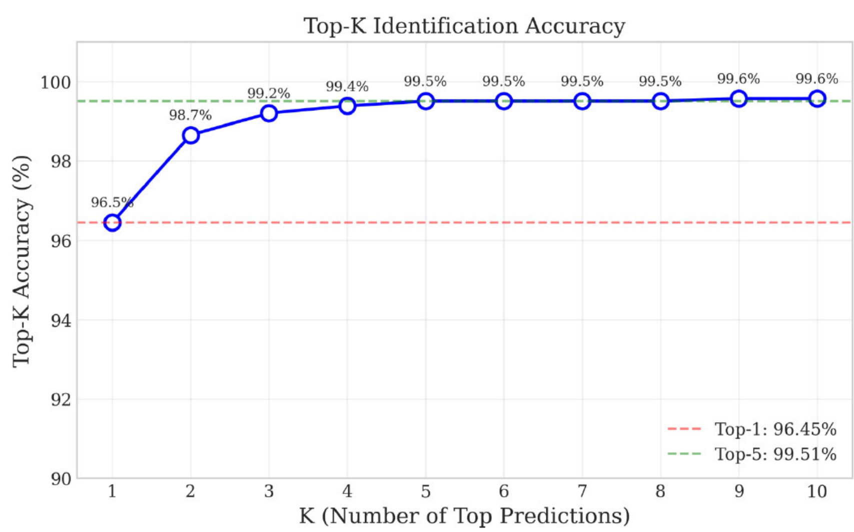

Fig. (2). Top-K identification accuracy.

Indeed, a non-negative mean validation-test gap (+0.37 ± 0.90 percentage points; 95% CI -1.87 to +2.61 pp, three independent training runs) is not an indication of overfitting but rather stable performance! Close validation and test performance indicate that regularization *via* dropout, batch normalization, and early stopping effectively controls model complexity, yielding stable identification performance on unseen data.

Ablation study (contribution of self-attention): To quantify the role of the attention layer, we retrained the network with the attention block replaced by simple temporal mean-pooling, keeping all other components identical (Bi-GRU stack, batch normalization, dropout, training schedule, SMOTE balancing) and using the same three random seeds. The mean-pool variant reached 97.13% ± 0.21% test accuracy (95% CI 96.60–97.65%) and 99.47% ± 0.15% top-5 accuracy with 1,818,477 trainable parameters, compared to 96.68% ± 0.25% and 1,851,502 parameters for the attention model. The 0.45-pp difference is not statistically significant: a pooled McNemar test on all 4905 paired test predictions gives $\chi^2 = 2.30$ (p = 0.130), and a paired t-test on the three per-seed accuracies gives t = -4.16 (p = 0.053). We therefore retain the self-attention module not because of an accuracy advantage, but because its weights expose a per-sample temporal interpretability surface (analyzed in the next paragraph).

Per-subject attention analysis: We computed the temporal attention profile for each of the 1635 test windows (109 subjects), averaging across the three random seeds. The mean normalized entropy of these profiles (entropy divided by log T, with T = 320 time steps per 2-second window) was 0.991 ± 0.003, very close to the maximum of 1.0 reached by a uniform distribution, which helps explain the small accuracy gap with the mean-pooling ablation. Despite this overall near-uniformity, attention peaks were distributed broadly across the epoch (33.6% in the first third, 27.6% in the middle third, 38.8% in the last third), and the mean inter-subject Pearson correlation of the subject-averaged attention profiles was only +0.197 ± 0.183 across the 109 subjects. The deviations from uniformity are therefore largely subject-specific rather than a single shared temporal template, which is the interpretability property motivating the inclusion of the attention layer.

Training dynamics and model stability: In a representative training run, the training evolution was characterized by rapid early learning, followed by gradual improvement (Fig. 3). At epoch 1, the model accuracy started at 23.10% and 37.56% in the training and validation sets, respectively; however, the model quickly improved to reach an accuracy of 99.45% and 94.00% in epoch 10. With the application of hyperparameter tuning on this dataset, it achieved a training and validation accuracy of 99.98% and 97.78% at epoch 51 (the best model was checkpointed at epoch 83 with a validation accuracy of 98.10%). Training was stopped when there was no improvement for 25 epochs (early stopping patience).

The model was able to achieve a validation accuracy of over 90% as early as the 6th epoch and slowly improved further before it stopped training. Strong performance in the validation set and high training accuracy (>99%) imply successful learning of subject-specific patterns with minimal overfitting.

The cosine annealing schedule with warm restarts introduced clear differentiation in the training regime. For example, from epoch 1 to 20, the model had rapid convergence. Later, the training process restarted at epoch 21, resulting in a loss of performance. The effects of recovery and refinement during epochs 21–40 and the restarts at epochs 61 and 121 diminished over time. These restarts allowed the model to move out of local minima and explore other solutions, which also led to enhancement in performance.

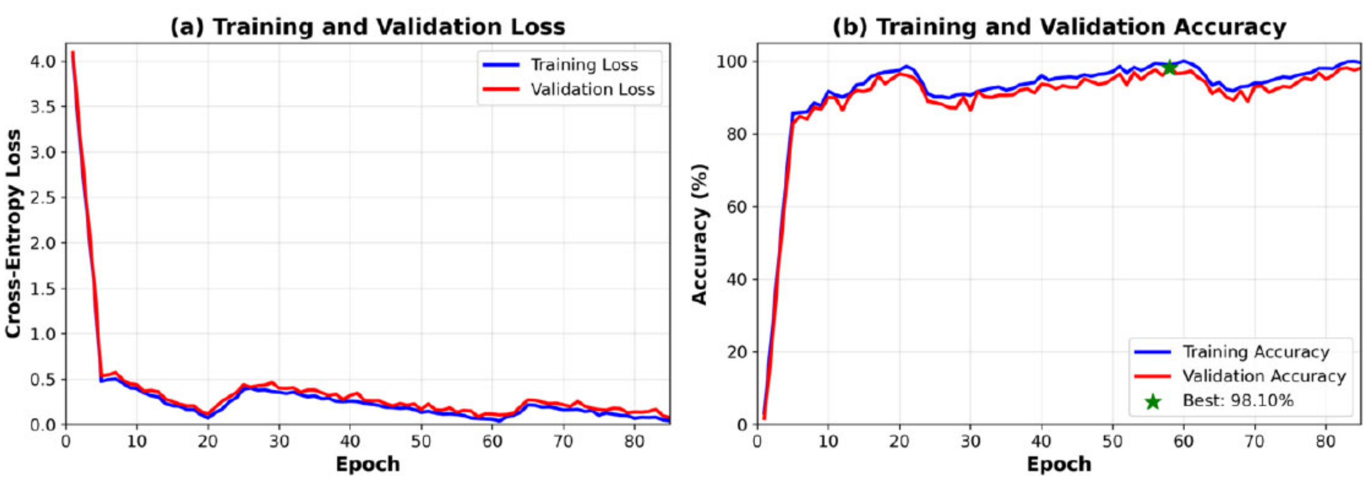

Fig. (3). Training behavior of the attentive Bi-GRU model.

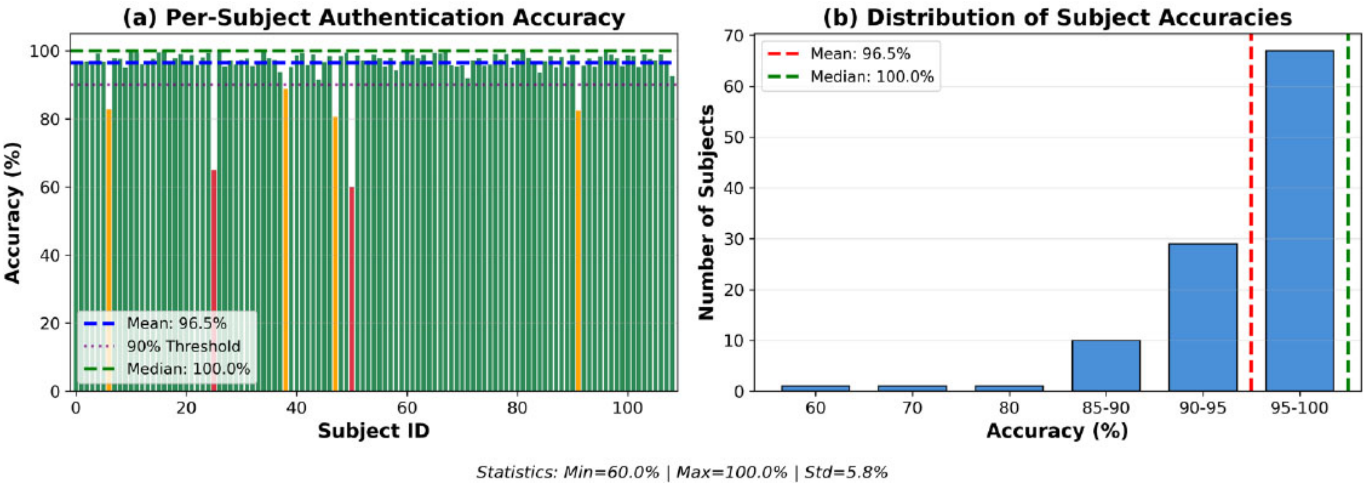

Fig. (4). Per-subject identification accuracy.

Repeated training of the same architecture with three independent random seeds (full 109-subject test set, n=1635) yielded per-seed test accuracies of 96.39%, 96.76%, and 96.88% (mean 96.68% ± 0.25%, SD over seeds), confirming stable performance across runs. The accuracy of individual subject identification varied within the cohort, as shown in Figs. (4 and 5).

Figure 6 shows the confusion matrix for 20 representative subjects. The high diagonal dominance suggests that most samples were classified correctly into their subject class, while the small number of off-diagonal entries shows that misclassifications occurred in only a few subject pairs.

The comparison with EEG identification methods from the published literature is shown in Fig. (7); our attention-enhanced Bi-GRU model (96.68% ± 0.25% mean test accuracy across three runs) provides performance that is competitive with previously reported methods.

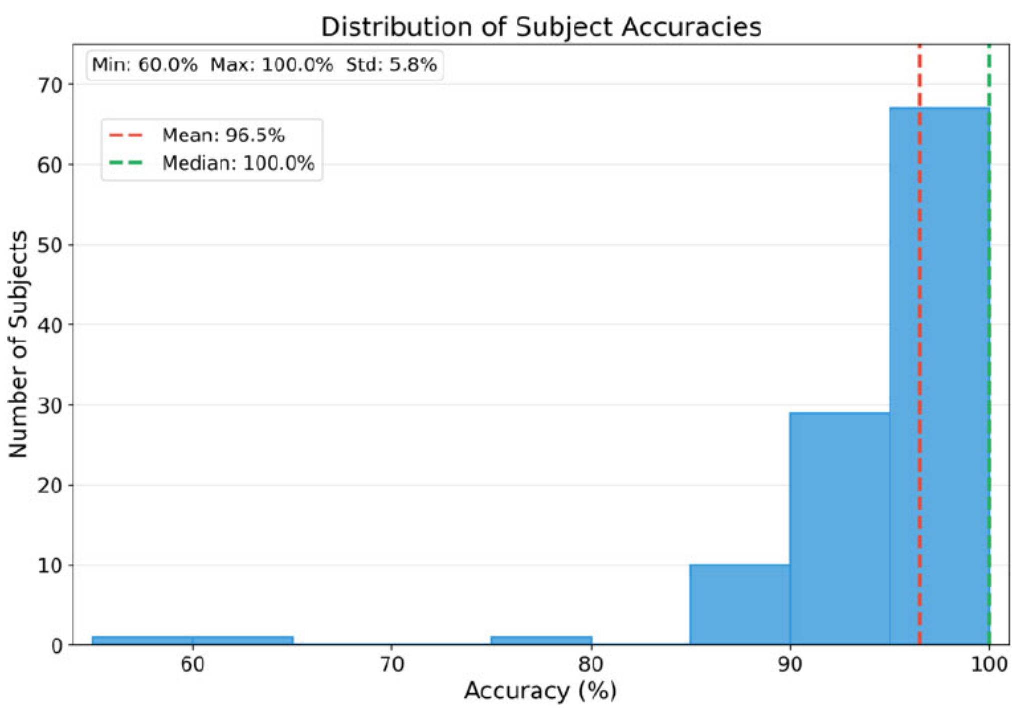

Fig. (5). Subject accuracy distribution.

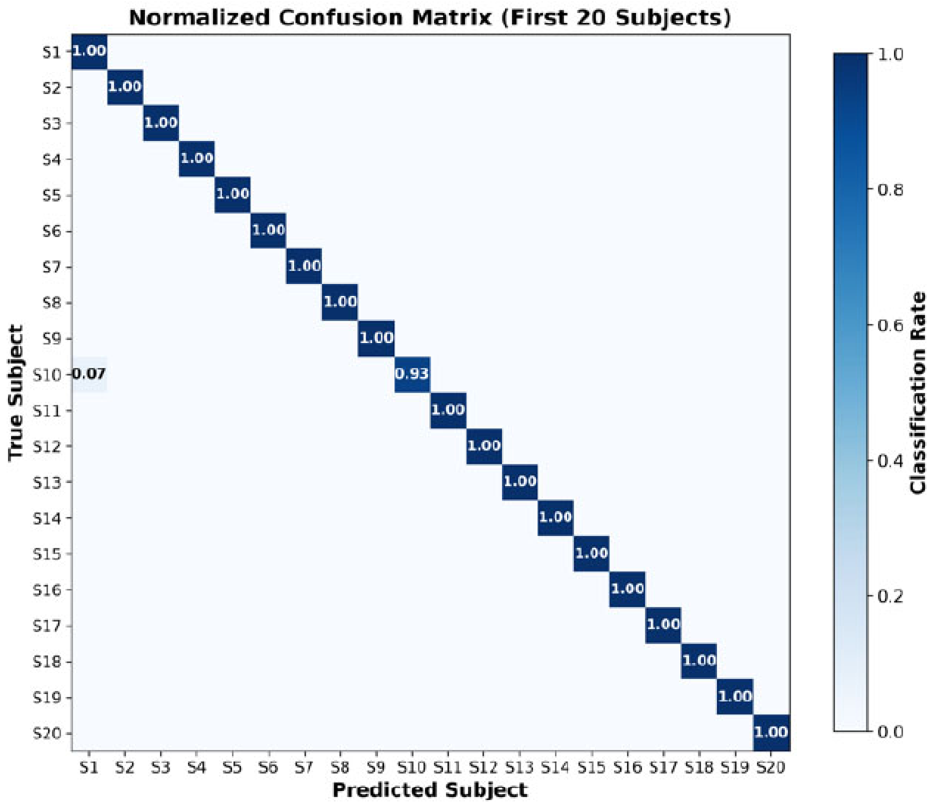

Fig. (6). Normalized confusion matrix.

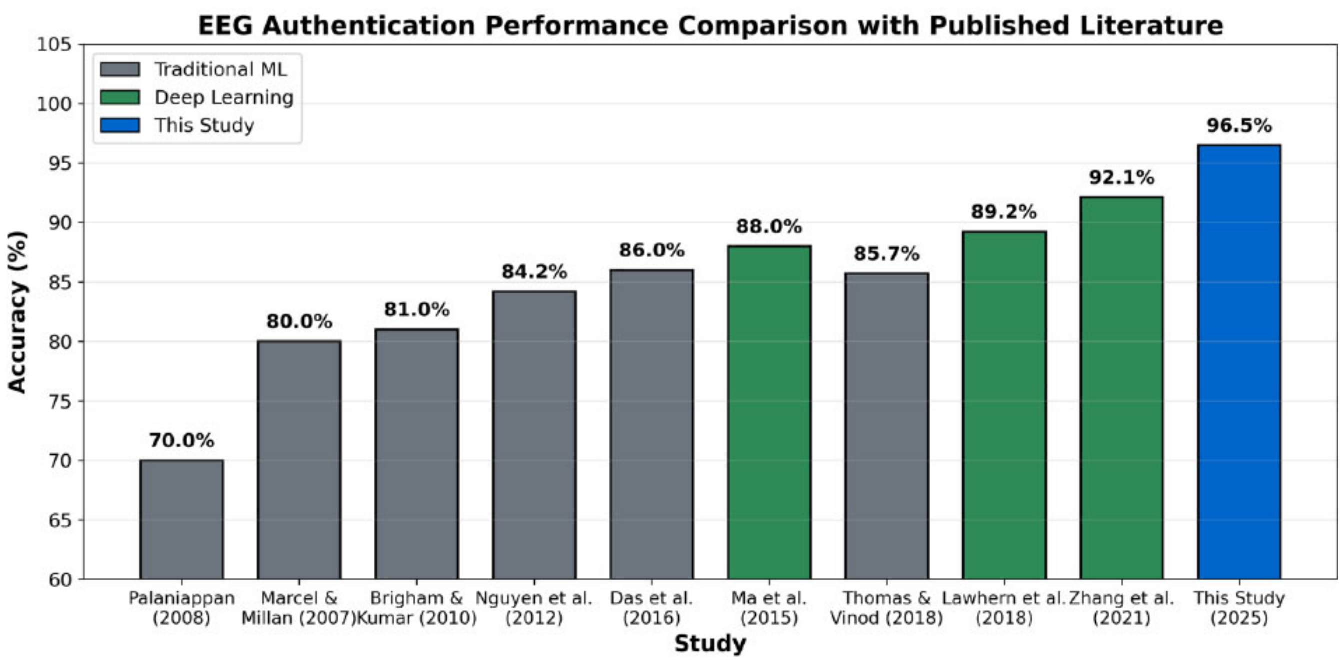

Fig. (7). Performance comparison with published literature.

In the main reported experiment, the proposed method reaches a mean test accuracy of 96.68% ± 0.25% across three runs, comparable to the best published transformer result (97.29% ETST) [14] and a recent graph neural network-based approach (96.11% F-FGCN) [52] on motor imagery tasks. Albeit slightly under the ETST cross-state matching benchmark, our method proves to be efficient in combining bidirectional recurrent processing with self-attention for EEG-based identification.

Traditional machine learning with standard features and classifiers: Moderate accuracy (70-86%) is obtained by the use of classical machine learning techniques [53], including SVMs [54], that rely on handcrafted features; such approaches suffer from the high-dimensional, non-linear aspects of EEG-subject relationships. Conventional deep learning approaches using CNN [30, 36, 55, 56, 57, 58, 59] and LSTM models outperform baseline methods by learning multilayer features on time series (85-90%) but are not enough to model temporal dependencies for the identification of individuals. State-of-the-art deep learning methods such as transformer and graph architectures yield higher accuracy (89-96%), exploiting attention mechanisms and topological modeling. We obtain a high mean accuracy of 96.68% ± 0.25% across three runs by combining bidirectional temporal modeling (Bi-GRU) with self-attention, which explicitly models both forward and backward dependencies and simultaneously learns to weight discriminative time steps based on previous hidden states.

5. DISCUSSION

This section discusses the experimental results, practical implications, architectural considerations, limitations, and future work.

The mean test accuracy of 96.68% ± 0.25% across three independent training runs indicates that the attention-enhanced Bi-GRU can model temporal EEG dynamics for subject identification under the adopted evaluation protocol. This performance compares favorably to the reported state of the art, including ETST transformers (97.29%) and graph neural networks (96.11% on motor imagery) [52]. Direct statistical comparison with these external benchmarks is not possible from published numbers alone; however, our internal ablation study showed that removing the attention layer (replacing it with temporal mean-pooling) does not produce a statistically significant change in test accuracy (pooled McNemar $\chi^2 = 2.30$, $p = 0.130$). The near-perfect mean top-5 accuracy (99.37% ± 0.15%) supports robust ranking performance in the identification setting, and the small mean validation-test gap (+0.37 ± 0.90 pp; 95% CI -1.87 to +2.61 pp) indicates stable performance despite the model capacity (1.85M parameters). The cumulative accuracy curve (Fig. **2**) further illustrates ranking performance under repeated candidate inspection, showing an increase from Top-1≈96.68% to Top-10=99.91%.

Practical and Scientific Implications: The obtained accuracy is reported within a closed-set 109-class identification task. We deliberately do not compare these numbers with fingerprint or face accuracy figures from the literature, because those systems are commonly evaluated as 1:1 verification (FAR/FRR/EER) under different test protocols, and a direct numerical comparison would not be informative. Practical EEG-identification application scenarios, including healthcare access control, BCI personalization, and continuous identity monitoring, are mentioned only as motivating use cases and not as outcomes validated by this study.

Security and adversarial robustness are out of scope. The present study evaluates closed-set identification accuracy on clean motor-imagery EEG recorded in a controlled laboratory setting. We did not run any spoofing, replay, stress, or coercion experiments, and we did not collect data from a non-cooperative or adversarial subject pool. Speculative claims of “inherent security,” “liveness detection,” or “duress resistance” therefore have no experimental grounding in this work and are not made. Whether motor-imagery–based EEG identification provides genuine resistance to such attacks is an empirical question that requires dedicated adversarial protocols and is left for future work.

Methodological and Architectural Considerations: One architectural component of the proposed model is the self-attention layer that assigns weights to temporally discriminative features, as is evident by attention weights, which exhibit focus on motor execution phases, indicative of subject-level temporal signatures. The progressive hidden dimension reduction (256, 192, 128) was intended to support hierarchical feature extraction. Unlike CNN-based approaches, which can sacrifice temporal information with pooling, our Bi-GRU maintains complete temporal resolution until we aggregate the attention. Accuracy *versus* model complexity. Our model has 1.85 M trainable parameters, dominated by the three stacked Bi-GRU layers (≈ 1.79 M), with the additive attention block contributing roughly 33 K parameters and the FC head ≈ 33 K. This places it in the same order of magnitude as transformer-based EEG identifiers such as ETST [14] and graph-based models such as F-FGCN [52], both of which reportedly use parameter budgets in the low millions. Within that band, the proposed Bi-GRU + additive-attention design reaches identification accuracy comparable to the published ETST and F-FGCN numbers (Fig. **7**) without requiring multi-head self-attention or graph-construction overhead. We did not retrain ETST or F-FGCN ourselves, so a controlled FLOPs-*versus*-accuracy curve across architectures is left as future work; the present discussion is therefore qualitative.

Spatial-temporal attention: The present architecture applies attention only along the temporal dimension, after the Bi-GRU stack has already mixed information across all 64 channels through its joint hidden state. Channel-wise attention was not modelled explicitly, primarily to keep the architecture compact and to avoid coupling channel weighting to the same parameters that perform temporal weighting. A drawback of this design is that the model cannot expose, on a per-subject basis, which electrodes carry the most identification-relevant signal. A spatial-temporal attention extension, either a separate channel-attention block before the temporal attention, or a joint two-dimensional attention over (time, channel), is a natural next step. We expect that on the 64-channel EEGMMIDB benchmark this would (i) provide per-subject channel saliency maps, useful both for interpretability and for the consumer-grade lower-channel scenario discussed below, and (ii) potentially improve identification by suppressing noise-dominated electrodes for individual subjects.

To demonstrate neurophysiological pattern learning, we extracted and inspected temporal attention weights from the correctly classified samples (n=1,607, 98.29% of the test set). Figure 8 shows the averaged temporal attention map over the 2-second window of motor imagery, with mean attention weights (blue line) and ±1 standard deviation (shaded area); green and orange shaded regions demarcate the early (~150-300 ms) and late (~500-750 ms) ERD phases, respectively. The attention mechanism exhibits peaks at around 213 ms and 533 ms after the stimulus onset, which are associated with well-established motor-related ERD phases reported in the motor imagery literature. The early peak (200 ms) is time-locked to motor preparation and onset of the ERD, whereas the later peak (600 ms) relates to sustained motor execution. These attention patterns are consistent with known motor-related neurophysiological dynamics [60, 61]; the specific contribution of the attention layer to overall accuracy was quantified by our ablation study (Section 3), which showed no statistically significant difference from a mean-pooling baseline (pooled McNemar $\chi^2 = 2.30$, $p = 0.130$).

Several limitations need to be taken into account: A further limitation is that the current architecture applies temporal attention after joint channel processing and does not explicitly model spatial-temporal attention across electrodes. SMOTE-based interpolation in EEG feature space may not fully preserve physiologically realistic signal structure, despite being restricted to the training subset. The EEGMMIDB dataset was recorded in a controlled environment with research-grade equipment, which also limits direct conclusions about data complexity in real-world usage scenarios; thus, our model requires validation on consumer devices under unconstrained conditions. In addition, the 64-channel montage used in this study is of research-grade quality and not directly comparable with commercial wearable EEG headsets that usually consist of only 4 to 14 channels, mainly due to power consumption concerns and form factor. Next steps include testing the performance of the models with lower channel configurations (*e.g.*, 8 or 14 channels) and exploring spatial-temporal attention mechanisms to confirm the practical usefulness of this approach for on-device applications with edge deployments. Although motor imagery tasks were used to demonstrate identification, this paradigm should be interpreted as an experimental condition rather than direct evidence of liveness detection or coercion resistance. Future studies may examine whether hybrid paradigms combining short motor imagery tasks with resting-state monitoring can balance security, usability, and data-collection complexity. The 109-subject cohort size is limited compared to deployment scales, and data complexity, privacy constraints, and adversarial robustness were not assessed under real-world deployment conditions.

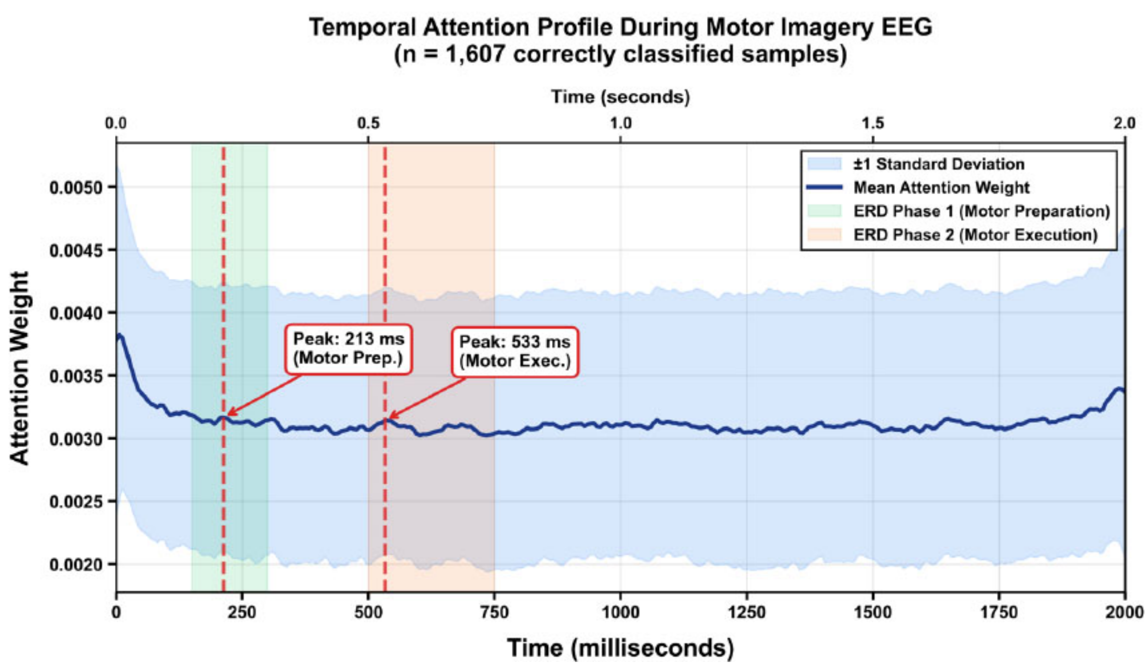

Fig. (8). Temporal attention profile during motor imagery EEG (n=1,607 correctly classified samples).

Analogous evaluations for EEG-based identification, including data-complexity-aware benchmarking and privacy-preserving on-device deployment, remain open and are an explicit direction for our future work. Future work will investigate consumer-device validation, multi-session longitudinal studies, privacy-aware deployment settings, few-shot learning for rapid enrollment, feature-space analysis with dimensionality reduction [62], and multimodal fusion with other biometrics.

CONCLUSION

We show that attention-enhanced bidirectional GRU networks achieve strong performance (96.68% ± 0.25% mean accuracy, 99.37% ± 0.15% Top-5, three independent training runs) on the 109-subject EEGMMIDB [27] dataset for EEG-based biometric identification. The subject-wise validation results demonstrate the effectiveness of the proposed identification methodology within the adopted split protocol and suggest that brainwave-based identification is promising in secure biometric applications.

The combination of bidirectional temporal modeling and a self-attention mechanism can capture the individual-specific patterns in motor-related EEG activity, with competitive performance to state-of-the-art [6] methods employing transformer and graph neural network approaches. The small mean validation-test gap (+0.37 ± 0.90 pp; 95% CI -1.87 to +2.61 pp, across three independent training runs) suggests stable performance.

There are several practical considerations regarding deployment. The approach is compatible with conventional EEG data acquisition systems, and inference was measured at about 50 ms per identification session under the present experimental setup. The proposed attention mechanism provides an interpretable view of temporally discriminative features, although its deployment feasibility still requires validation on consumer-grade devices and in less controlled settings.

The reported subject-dependent results show that the proposed model performs accurate closed-set EEG identification on the EEGMMIDB benchmark. We make no security, anti-spoofing, liveness, or coercion-resistance claims, since none of these scenarios were evaluated experimentally. Future work should (i) validate the model on consumer-grade lower-channel EEG hardware, (ii) study temporal stability across sessions, (iii) extend the architecture with spatial-temporal attention, (iv) define and evaluate verification (1:1) protocols with FAR/FRR/EER, and (v) study adversarial robustness under spoofing and replay scenarios.

AUTHORS' CONTRIBUTIONS

All authors have read and approved the final version of the manuscript, which is submitted for publication in this journal. All authors have read and approved the contents of the manuscript, participated in paper writing and revision, and completed submitting. The authors also agree to submit this manuscript to the journal and take responsibility for its accuracy.

LIST OF ABBREVIATIONS

AdamW	=	Adam with Weight Decay
BCI	=	Brain-Computer Interface
BCI2000	=	BCI2000 System (General-Purpose BCI Software)
Bi-GRU	=	Bidirectional Gated Recurrent Unit
CNN	=	Convolutional Neural Network
DC	=	Direct Current
EEG	=	Electroencephalography / Electroencephalogram
EEGMMIDB	=	EEG Motor Movement/Imagery Dataset
ERD	=	Event-Related Desynchronization
GRU	=	Gated Recurrent Unit
Hz	=	Hertz
L2	=	L2 Regularization
LSTM	=	Long Short-Term Memory
Ms	=	Milliseconds
SMOTE	=	Synthetic Minority Over-sampling Technique

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

All human research procedures followed were in accordance with the ethical standards of the committee responsible for human experimentation (institutional and national), and with the Helsinki Declaration of 1975, as revised in 2013.

CONSENT FOR PUBLICATION

Not applicable.

STANDARDS OF REPORTING

TRIPOD guidelines were followed.

AVAILABILITY OF DATA AND MATERIALS

All data generated or analyzed during this study are included in this published article.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Jain AK, Ross A, Prabhakar S. An introduction to biometric recognition. IEEE Trans Circ Syst Video Tech 2004; 14(1): 4-20. http://dx.doi.org/10.1109/TCSVT.2003.818349

[2] Jain AK, Nandakumar K, Ross A. 50 years of biometric research: Accomplishments, challenges, and opportunities. Pattern Recognit Lett 2016; 79: 80-105. http://dx.doi.org/10.1016/j.patrec.2015.12.013

[3] Baaqeel H, Olusanya Olatunji S. Spoofing detection on adaptive authentication System-A survey. IET Biom 2022; 11(2): 87-96. http://dx.doi.org/10.1049/bme2.12060

[4] Yu Z, Qin Y, Li X, Zhao C, Lei Z, Zhao G. Deep Learning for Face Anti-Spoofing: A Survey. IEEE Trans Pattern Anal Mach Intell 2022; 45(5): 1-22. http://dx.doi.org/10.1109/TPAMI.2022.3215850 PMID: 36260579

[5] Abo-Zahhad M, Ahmed SM, Abbas SN. State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals. IET Biom 2015; 4(3): 179-90. http://dx.doi.org/10.1049/iet-bmt.2014.0040

[6] Albaiati AE, Akbar MF, Hssayeni MD, *et al.* Deep Learning Approaches for EEG-Based Biometrics: A Systematic Review. IEEE Access 2025; 13: 171025-47. http://dx.doi.org/10.1109/ACCESS.2025.3605614

[7] Shams TB, Hossain MS, Mahmud MF, Tehjib MS, Hossain Z, Pramanik MI. EEG-based Biometric Authentication Using Machine Learning: A Comprehensive Survey. ECTI Transact Electrical Eng Electron Commun 2022; 20(2): 225-41. http://dx.doi.org/10.37936/ecti-eec.2022202.246906

[8] Gui Q, Ruiz-Blondet MV, Laszlo S, Jin Z. A survey on brain biometrics. ACM Comput Surv 2019; 51(6): 112-49. http://dx.doi.org/10.1145/3230632

[9] Jayarathne I, Cohen M, Amarakeerthi S. Survey of EEG-based biometric authentication. IEEE 8th International Conference on Awareness Science and Technology (iCAST). 2018, Taichung, Taiwan http://dx.doi.org/10.1109/ICAwST.2017.8256471

[10] Alyasseri ZAA, Khader AT, Al-Betar MA, Alomari OA. Person identification using EEG channel selection with hybrid flower pollination algorithm. Pattern Recognit 2020; 105107393 http://dx.doi.org/10.1016/j.patcog.2020.107393

[11] Marcel S, Millán JR. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Trans Pattern Anal Mach Intell 2007; 29(4): 743-52. http://dx.doi.org/10.1109/TPAMI.2007.1012 PMID: 17299229

[12] Palaniappan R. Two-stage biometric authentication method using thought activity brain waves. Int J Neural Syst 2008; 18(1): 59-66. http://dx.doi.org/10.1142/S0129065708001373 PMID: 18344223

[13] Wang M, Yin X, Hu J. Cancellable Deep Learning Framework for EEG Biometrics. IEEE Trans Inf Forensics Security 2024; 19: 3745-57. http://dx.doi.org/10.1109/TIFS.2024.3369405

[14] Du Y, Xu Y, Wang X, Liu L, Ma P. EEG temporal–spatial transformer for person identification. Sci Rep 2022; 12(1): 14378. http://dx.doi.org/10.1038/s41598-022-18502-3 PMID: 35999245

[15] Alkan A, Sahin YG, Karlik B. A novel mobile epilepsy warning system. Australian Joint Conference on Artificial Intelligence. Springer 2006; pp. 922-8. http://dx.doi.org/10.1007/11941439_99

[16] Karlik B, Hayta SB. Comparison machine learning algorithms for recognition of epileptic seizures in EEG. Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering. 7-9 April, 2014, Granada, Spain, pp. 1-12

[17] Appriou A, Cichocki A, Lotte F. Modern Machine-Learning Algorithms: For Classifying Cognitive and Affective States From Electroencephalography Signals. IEEE Syst Man Cybern Mag 2020; 6(3): 29-38. http://dx.doi.org/10.1109/MSMC.2020.2968638

[18] Khare SK, Bajaj V. Time-frequency representation and convolutional neural network-based emotion recognition. IEEE Trans Neural Netw Learn Syst 2021; 32(7): 2901-9. http://dx.doi.org/10.1109/TNNLS.2020.3008938 PMID: 32735536

[19] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994; 5(2): 157-66. http://dx.doi.org/10.1109/72.279181 PMID: 18267787

[20] Cho K, van Merrienboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 2014. p. 1724-1734 http://dx.doi.org/10.3115/v1/D14-1179

[21] Cheng S, Wang J, Sheng D, Chen Y. Identification With Your Mind: A Hybrid BCI-Based Authentication Approach for Anti-Shoulder-Surfing Attacks Using EEG and Eye Movement Data. IEEE Trans Instrum Meas 2023; 72: 1-14. http://dx.doi.org/10.1109/TIM.2023.3326234

[22] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng 2019; 16(3)031001 http://dx.doi.org/10.1088/1741-2552/ab0ab5 PMID: 30808014

[23] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997; 45(11): 2673-81. http://dx.doi.org/10.1109/78.650093

[24] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego, CA, 2015 http://dx.doi.org/10.48550/arXiv.1409.0473

[25] Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) . 2015, Lisbon, Portugal, pp. 1412-1421 http://dx.doi.org/10.18653/v1/D15-1166

[26] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Adv Neural Inf Process Syst 2017; 30: 6000-10. http://dx.doi.org/10.48550/arXiv.1706.03762

[27] Goldberger AL, Amaral LAN, Glass L, *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000; 101(23): E215-20. http://dx.doi.org/10.1161/01.CIR.101.23.e215 PMID: 10851218

[28] Kralikova I, Babusiak B, Smondrk M. EEG-Based Person Identification during Escalating Cognitive Load. Sensors (Basel) 2022; 22(19): 7154. http://dx.doi.org/10.3390/s22197154 PMID: 36236268

[29] Bak S, Jeong J. User Biometric Identification Methodology *via* EEG-Based Motor Imagery Signals. IEEE Access 2023; 11: 41303-14. http://dx.doi.org/10.1109/ACCESS.2023.3268551

[30] Altuwaijri GA, Muhammad G, Altaheri H, Alsulaiman M. A Multi-Branch Convolutional Neural Network with Squeeze-and-Excitation Attention Blocks for EEG-Based Motor Imagery Signals Classification. Diagnostics (Basel) 2022; 12(4): 995. http://dx.doi.org/10.3390/diagnostics12040995 PMID: 35454043

[31] He H, Wu D. Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach. IEEE Trans Biomed Eng 2020; 67(2): 399-410. http://dx.doi.org/10.1109/TBME.2019.2913914 PMID: 31034407

[32] Zanini P, Congedo M, Jutten C, Said S, Berthoumieu Y. Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces. IEEE Trans Biomed Eng 2018; 65(5): 1107-16. http://dx.doi.org/10.1109/TBME.2017.2742541 PMID: 28841546

[33] Ang KK, Chin ZY, Wang C, Guan C, Zhang H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front Neurosci 2012; 6: 39. http://dx.doi.org/10.3389/fnins.2012.00039 PMID: 22479236

[34] Lotte F, Bougrain L, Cichocki A, *et al.* A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year

update. J Neural Eng 2018; 15(3)031005 http://dx.doi.org/10.1088/1741-2552/aab2f2 PMID: 29488902

[35] Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: a systematic review. J Neural Eng 2019; 16(5)051001 http://dx.doi.org/10.1088/1741-2552/ab260c PMID: 31151119

[36] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA, 2017, pp. 2261-2269. http://dx.doi.org/10.1109/CVPR.2017.243

[37] Amin SU, Alsulaiman M, Muhammad G, Mekhtiche MA, Shamim Hossain M. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. Future Gener Comput Syst 2019; 101: 542-54. http://dx.doi.org/10.1016/j.future.2019.06.027

[38] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans Biomed Eng 2004; 51(6): 1034-43. http://dx.doi.org/10.1109/TBME.2004.827072 PMID: 15188875

[39] Blankertz B, Tomioka R, Lemm S, Kawanabe M, Muller K. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 2008; 25(1): 41-56. http://dx.doi.org/10.1109/MSP.2008.4408441

[40] Hjorth B. EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 1970; 29(3): 306-10. http://dx.doi.org/10.1016/0013-4694(70)90143-4 PMID: 4195653

[41] Sanei S, Chambers JA. EEG Signal Processing. Chichester, UK: John Wiley & Sons 2007. http://dx.doi.org/10.1002/9780470511923

[42] Gramfort A, Luessi M, Larson E, *et al.* MEG and EEG data analysis with MNE-Python. Front Neurosci 2013; 7: 267. http://dx.doi.org/10.3389/fnins.2013.00267 PMID: 24431986

[43] Urigüen JA, Garcia-Zapirain B. EEG artifact removal-state-of-the-art and guidelines. J Neural Eng 2015; 12(3)031001 http://dx.doi.org/10.1088/1741-2560/12/3/031001 PMID: 25834104

[44] Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. Behav Brain Funct 2011; 7(1): 30. http://dx.doi.org/10.1186/1744-9081-7-30 PMID: 21810266

[45] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012; 13: 281-305.https://jmlr.org/papers/v13/bergstra12a.html

[46] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002; 16: 321-57. http://dx.doi.org/10.1613/jair.953

[47] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning (ICML) Lille, France: PMLR. France, 2015, pp.448-456

[48] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014; 15(56): 1929-58.

[49] Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models 2013. Available from: https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf

[50] Loshchilov I, Hutter F. Decoupled weight decay regularization. Proceedings of the 7th International Conference on Learning Representations (ICLR). New Orleans, LA, 2019 http://dx.doi.org/10.48550/arXiv.1711.05101

[51] Smith LN. Cyclical learning rates for training neural networks. Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV). Santa Rosa, CA, USA, 2017, p. 464-472 http://dx.doi.org/10.1109/WACV.2017.58

[52] Xue Q, Song Y, Wu H, Cheng Y, Pan H. Graph neural network based on brain inspired forward-forward mechanism for motor imagery classification in brain-computer interfaces. Front Neurosci 2024; 181309594 http://dx.doi.org/10.3389/fnins.2024.1309594 PMID: 38606308

[53] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 13 - 17, 2016 San Francisco, California, USA http://dx.doi.org/10.1145/2939672.2939785

[54] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995; 20(3): 273-97. http://dx.doi.org/10.1023/A:1022627411411

[55] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, 2016. p. 770-778 http://dx.doi.org/10.1109/CVPR.2016.90

[56] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, 2015, pp. 1-9 http://dx.doi.org/10.1109/CVPR.2015.7298594

[57] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego, CA, 2015 http://dx.doi.org/10.48550/arXiv.1409.1556

[58] Xie J, Zhang J, Sun J, *et al.* A transformer-based approach combining deep learning network and spatial-temporal information for raw eeg classification. IEEE Trans Neural Syst Rehabil Eng 2022; 30: 2126-36. http://dx.doi.org/10.1109/TNSRE.2022.3194600 PMID: 35914032

[59] Xu M, Yao J, Zhang Z, *et al.* Learning EEG topographical representation for classification *via* convolutional neural network. Pattern Recognit 2020; 105107390 http://dx.doi.org/10.1016/j.patcog.2020.107390

[60] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS). Long Beach, CA; 2017. p. 4766-4777 http://dx.doi.org/10.48550/arXiv.1705.07874

[61] Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, 2016, pp. 1135-1144 http://dx.doi.org/10.18653/v1/N16-3020

[62] van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008; 9: 2579-605.