

Temporal Convolutional Networks Enhanced RNN Models for EEG Signal Classification

Mustafa Erginli*

¹ Simav Vocational School, Kutahya Dumlupınar University,

*(mustafa.erginli@dpu.edu.tr)

(Received: 26 October 2025, Accepted: 31 October 2025)

(7th International Conference on Engineering and Applied Natural Sciences ICEANS 2025, October 25-26, 2025)

ATIF/REFERENCE: Erginli, M. (2025). Temporal Convolutional Networks Enhanced RNN Models for EEG Signal Classification, *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(10), 500-509.

Abstract – This study systematically compares four deep learning architectures for classifying EEG signals. LSTM and GRU models, as well as TCN + LSTM and TCN + GRU models, are evaluated under the same conditions of data augmentation, hyperparameterization and early stopping. The data consisted of recordings with a sampling frequency of 173.61 Hz, categorized into normal, interictal and ictal classes. Normalization and time domain data augmentation strategies were used in the preprocessing steps. The results show that hybrid structures provide significant superiority. LSTM produced 86.67 percent accuracy, 0.8646 macro F1, 0.7908 Matthews correlation coefficient, and 0.9699 AUROC. GRU achieved 81.33 percent accuracy, 0.8220 macro F1, 0.7083 Matthews correlation coefficient and 0.9664 AUROC. TCN + LSTM performed more strongly with 92.00 percent accuracy, 0.9221 macro F1, and 0.9953 AUROC. The highest success was achieved with TCN + GRU. This model achieved 96.00% accuracy, 0.9610 macro F1 score, 0.9375 Matthews correlation coefficient and 0.9993 AUROC. In probabilistic calibration, the best value was obtained for TCN + GRU with a log loss of 0.0982 and a Brier score of 0.0426. In terms of inference efficiency, GRU is the fastest model with an average of 0.39 milliseconds. The hybrid models meet the real-time usage threshold with a delay of about 7.6 milliseconds. The results show that the TCN + GRU architecture is successful in achieving a balance between accuracy, calibration and latency.

Keywords – EEG, TCN, RNN, Hybrid Deep Learning, Epileptic Seizure Classification

I. INTRODUCTION

Electroencephalogram (EEG) signals are biomedical signals that record the electrical activity of the human brain in a time-dependent manner and are widely used in areas such as the diagnosis of neurological diseases, the analysis of cognitive states, and the development of brain-computer interfaces. Classification of EEG signals has become an increasingly researched topic in the fields of machine learning and deep learning due to the complex, multidimensional, and temporally dependent nature of the signals [1]. Accurate and reliable classification of EEG signals is crucial in various applications, including epilepsy diagnosis, sleep stage identification, mood analysis, and cognitive state tracking [2].

EEG data generally have a low signal-to-noise ratio and high inter-individual variations. Moreover, their high dimensionality and temporal dependence limit the generalization capacity of classical machine learning algorithms[3]. Therefore, deep learning architectures capable of capturing the complex structure

of EEG signals have been extensively used. Recurrent Neural Networks (RNNs) offer a powerful alternative in EEG analysis thanks to their ability to learn sequential relationships in time-dependent data [4]. Long Short-Term Memory (LSTM) networks, a type of RNN, are widely used in the classification of EEG signals due to their ability to preserve long-term dependencies. The gate mechanism of LSTM enables the modeling of complex temporal patterns in signals by controlling the flow of information between past and present time. LSTM-based approaches have gained importance due to their high accuracy rates, especially in applications such as emotion recognition, motion detection and epileptic activity detection [5].

Gated Recurrent Unit (GRU) architectures have fewer parameters than LSTM and can provide similar accuracy with a simpler structure. GRU is advantageous for real-time EEG applications due to its faster training time and low inference latency. In the literature, GRU-based models have been shown to significantly reduce training time and maintain similar accuracy levels compared to LSTM in classifying EEG signals [3], [6].

Not only recurrent structures but also convolutional approaches have become prominent in EEG classification. In particular, Temporal Convolutional Networks (TCNs) can effectively model long-term dependencies using extended convolutions and residual connections and offer parallel information processing compared to RNNs. Several studies have reported that TCN architectures offer advantages in multi-scale feature extraction and temporal pattern learning for complex time series such as EEG. Furthermore, combining TCN structures with attention mechanisms improves classification accuracy by focusing on meaningful parts of the signal [7], [8].

Hybrid architectures have recently gained prominence in EEG research. In these architectures, convolutional layers extract local patterns, while recurrent layers model long-term temporal dependencies. Architectures such as LSTM-FCN and GRU-FCN have achieved significant performance improvements in EEG classification by combining temporal and convolutional representations. These approaches integrate the LSTM or GRU layer with the convolutional block to create a more balanced feature representation, thus capturing both short- and long-term dependencies. ALSTM-FCN, an extension of LSTM-FCN with an attention mechanism, makes the model's decision-making process more explainable [9], [10].

Reliable assessment of model performance is crucial for time series classification. Especially with unbalanced datasets, accuracy alone is not a sufficient criterion for evaluation. Therefore, comprehensive measures such as Matthews Correlation Coefficient (MCC), Cohen's Kappa, Log Loss and Brier Score are preferred. MCC evaluates whether both positive and negative classes are correctly predicted, providing a reliable assessment, especially in unbalanced datasets. Brier Score and Log Loss measure the quality of probabilistic estimates and are important for analyzing uncertain signals such as EEG [11].

The limited amount of data due to the nature of biomedical signals makes data augmentation techniques important. Methods such as time axis shifting, random noise addition, scaling and time masking increase the generalization power of the model and reduce overfitting [12].

The literature shows that LSTM, GRU, TCN and their hybrid versions have achieved remarkable success in classifying EEG signals. However, most of the existing studies have not conducted comprehensive comparisons of hybrid architectures, especially in terms of calibration metrics. Moreover, practical factors such as inference latency of the models have been evaluated to a limited extent.

This study aims to comprehensively compare RNN-based models (LSTM and GRU) and TCN+RNN hybrid models (TCN+LSTM and TCN+GRU) for EEG signal classification. In this study, four different deep learning models, namely LSTM, GRU, TCN+LSTM and TCN+GRU, are comprehensively compared for the classification of EEG signals. The models are trained using the same data augmentation methods, hyperparameters and early stopping conditions. The EEG dataset used in the experiments contained multichannel recordings with a sampling frequency of 173.61 Hz. The results show that hybrid architectures provide significant performance improvements compared to classical RNN-based models. The LSTM model achieved 86.67% accuracy (Macro F1=0.8646, MCC=0.7908, AUROC=0.9699), while the GRU model achieved 81.33% accuracy (Macro F1=0.8220, MCC=0.7083, AUROC=0.9664). The TCN+LSTM model produced 92.00% accuracy (Macro F1=0.9221, MCC=0.8810, AUROC=0.9953),

while the highest performance was obtained with the TCN+GRU model (96.00% accuracy, Macro F1=0.9610, MCC=0.9375, AUROC=0.9993). Furthermore, the Log Loss and Brier Score values of the TCN+GRU model are 0.0982 and 0.0426, respectively, indicating high calibration accuracy of the probabilistic forecasts. In terms of inference time, GRU was the fastest model (0.39 ms), while TCN-based hybrids showed a delay of about 7 ms. These findings suggest that TCN+RNN structures, which jointly model short- and long-term dependencies of EEG signals, exhibit superior performance in terms of accuracy, stability and calibration.

The original contributions of this study are as follows:

- The TCN+GRU hybrid architecture is tested on EEG signals to analyze the interaction of both convolutional and recurrent representations.
- LSTM, GRU, TCN+LSTM, and TCN+GRU models are trained under the same data augmentation strategies, hyperparameters, and early stopping conditions to provide a fair comparison.
- A multidimensional performance analysis is presented using both classical and calibration metrics such as accuracy, F1, MCC, AUROC, log loss, Brier Score, and inference delay.
- The findings show that hybrid deep learning architectures can produce reliable probabilistic predictions with high accuracy in biomedical time series such as EEG and provide guidance for future EEG-based clinical systems.

II. MATERIALS AND METHOD

A. Dataset

In this study, we used the EEG dataset "Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity" published by Universitat Pompeu Fabra (Barcelona, Spain) [1]. The dataset comprises EEG recordings with a sampling frequency of 173.61 Hz, containing multichannel signals from various brain regions, including the frontal, temporal, and parietal lobes. The raw signals represent time-dependent changes of electrical activity in the brain.

In this study, the signals were categorized into three classes based on their clinical content: normal (healthy brain activity), interictal (inter-seizure activity), and ictal (epileptic seizure moment) recordings. The models' ability to distinguish between epileptic and non-epileptic states was examined. Access to the dataset is provided through the official data sharing portal of Universitat Pompeu Fabra[13].

B. Preprocessing

EEG data were transformed, normalized and prepared for model input prior to analysis. To eliminate inconsistencies due to amplitude variations of signals at different scales, all samples were scaled to the 0-1 range using min-max normalization.

Data augmentation techniques such as time masking, truncation, Gaussian jitter, amplitude scaling and circular temporal shifting were applied in the time domain to increase the generalizability of the model and reduce the risk of overfitting. Thus, overfitting of the models to the training data was prevented and the generalization capability was significantly improved.

C. Deep Learning Models

In this study, four different deep learning architectures are designed and compared for the classification of EEG signals. The first two are LSTM and GRU models based on RNN structure only. The last two models are TCN+LSTM and TCN+GRU hybrid architectures that combine a Temporal Convolutional Network (TCN) with recurrent layers. The LSTM model is an effective structure for capturing long-term dependencies in sequential time series such as EEG. LSTM cells selectively retain or forget information through a control mechanism consisting of input, output and forget gates, and learn sensitively to past signal patterns. The LSTM model used in this study has a single layer structure with 64 hidden units. Similar to the LSTM, the GRU model models sequential dependencies but with fewer parameters. GRU cells organize the flow of information through update and reset gates. This simplified structure results in faster training and lower inference latency. The TCN+LSTM hybrid model aims to learn both short- and long-term dependencies by combining convolutional and recurrent layers. The TCN layers capture local

patterns in the temporal structure of the signal through extended convolutions. The LSTM layer then processes these convolutional features in temporal context to produce higher-level representations. The TCN+GRU hybrid model uses GRU cells in the recurrent part. In this way, TCN's multi-scale feature extraction is combined with GRU's low-parameter sequential information processing capability. The model achieved high accuracy rates while maintaining efficiency in extraction time. In all models, dropout regularization was applied to avoid overlearning and the number of parameters was kept balanced to ensure a fair comparison. The same data augmentation methods, hyperparameters and early stopping conditions were used for each model during the training process.

D. Training Configuration

The models were trained using the PyTorch deep learning library. All experiments were performed under the same conditions to ensure a fair comparison. During the training process, an early stopping mechanism was implemented to improve the overall performance of the models and control the loss of validation.

Each model was trained for 300 epochs; however, the training process was automatically terminated when no improvement in verification loss was observed for three consecutive epochs. This avoids overlearning and increases the generalizability of the model. Training was performed using mini-batches of 16 samples.

The AdamW algorithm was used for weight optimization. The learning rate was set to 0.0001 and the weight reduction coefficient to 0.005. During training, a four-step gradient accumulation strategy was used to improve memory efficiency.

Dropout was applied for regularization in all models. In the LSTM and GRU layers, the dropout rate was set to 50 percent, while in the TCN layers a dropout rate of 40 percent was used. These rates were optimized to minimize the risk of overlearning while maintaining the learning capacity of the model.

The training and validation datasets were separated by randomly mixing them with a fixed seed value of 42 for all applications. This ensured the repeatability of the experiments and the consistency of the results. Furthermore, all training was performed on a GPU and the process was accelerated using the CUDA library.

After training each model, comprehensive performance metrics such as accuracy, macro F1 score, Matthews correlation coefficient, Cohen's Kappa, AUROC, log loss, Brier score and average inference delay were calculated during the validation and testing phases. The combined evaluation of these metrics revealed the classification success of the models as well as their reliability and calibration levels.

III. RESULTS

In this study, we compared the performance of four different models (LSTM, GRU, TCN+LSTM and TCN+GRU) in classifying EEG signals. The models are evaluated using multifaceted metrics such as accuracy, F1 score, Matthews correlation coefficient (MCC), AUROC, log loss and inference delay. The results show that hybrid structures exhibit significant advantages over classical RNN-based models.

Table 1 summarizes the test performance of all models. The LSTM model shows the baseline performance with an accuracy of 86.67%, while the GRU model achieves a lower accuracy of 81.33%. In contrast, hybrid models that combine convolutional time coding with RNN structure show a significant increase in performance. The TCN+LSTM model achieved 92% accuracy and an F1 score of 0.9221, while the TCN+GRU model gave the best results with 96% accuracy, 0.961 F1 and 0.9375 MCC values.

Table 1. Classification of performance metrics

Model	Accuracy (%)	Macro F1	MCC	AUROC	Log Loss	Brier Score	Mean Latency (ms)
LSTM	86.67	0.8646	0.7908	0.9699	0.3931	0.2074	0.409
GRU	81.33	0.8220	0.7083	0.9664	0.3698	0.2161	0.391
TCN + LSTM	92.00	0.9221	0.8810	0.9953	0.2903	0.1362	7.234
TCN + GRU	96.00	0.9610	0.9375	0.9993	0.0982	0.0426	7.641

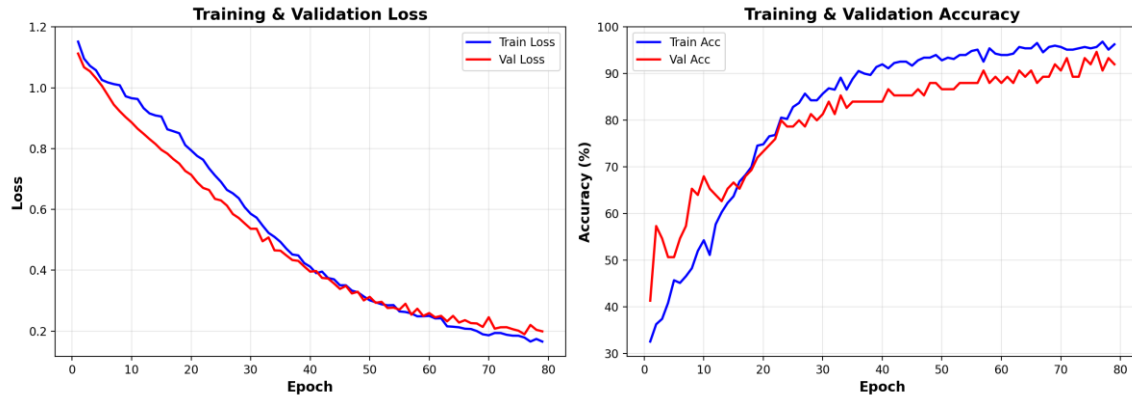


Fig. 1 TCN+GRU loss and accuracy curves

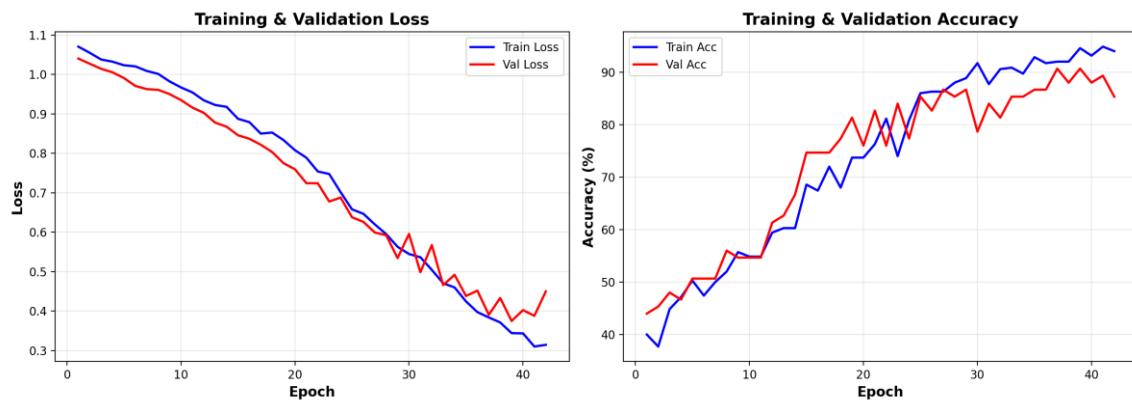


Fig. 2 TCN+LSTM loss and accuracy curves

Fig. 1 and Fig. 2 show the training and validation losses and accuracy curves of the TCN+GRU and TCN+LSTM models, respectively. In both models, the loss function decreases steadily over the epochs, and the verification accuracy reaches a saturation point. The faster convergence trend in the TCN+GRU model indicates that the model undergoes a more stable learning process.

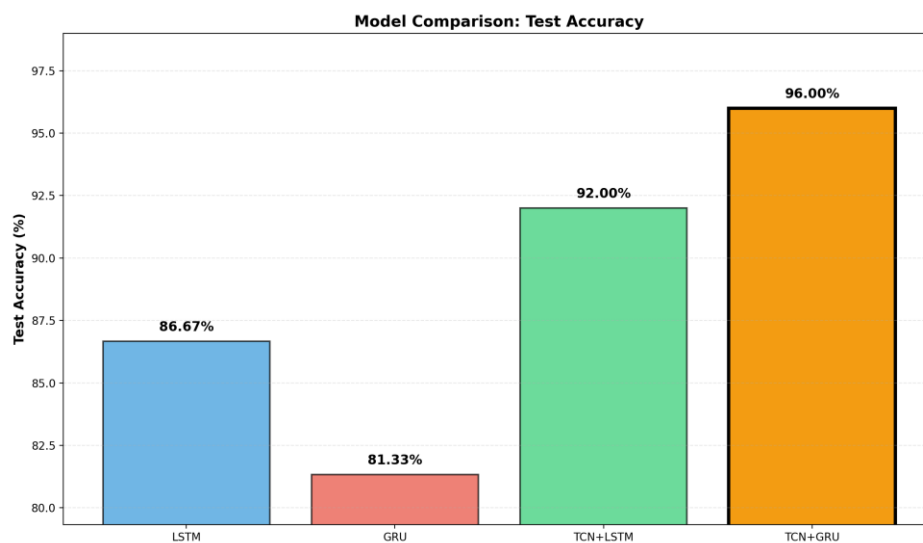


Fig. 3 Models' test accuracy comparison

Fig. 3 shows the comparative test accuracy of all models. This graph clearly shows that hybrid structures provide an accuracy improvement of about 10% over traditional RNN models.

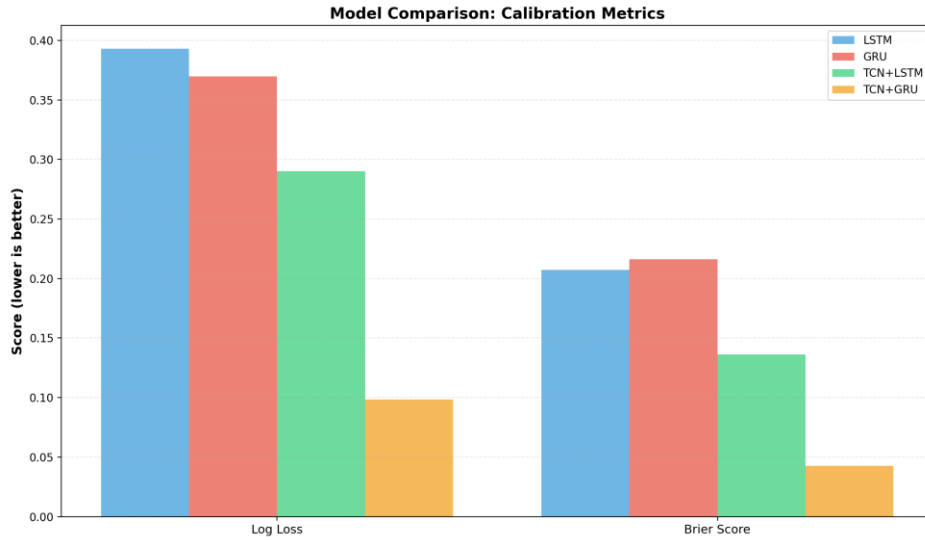


Fig. 4 Models' calibration metrics comparison

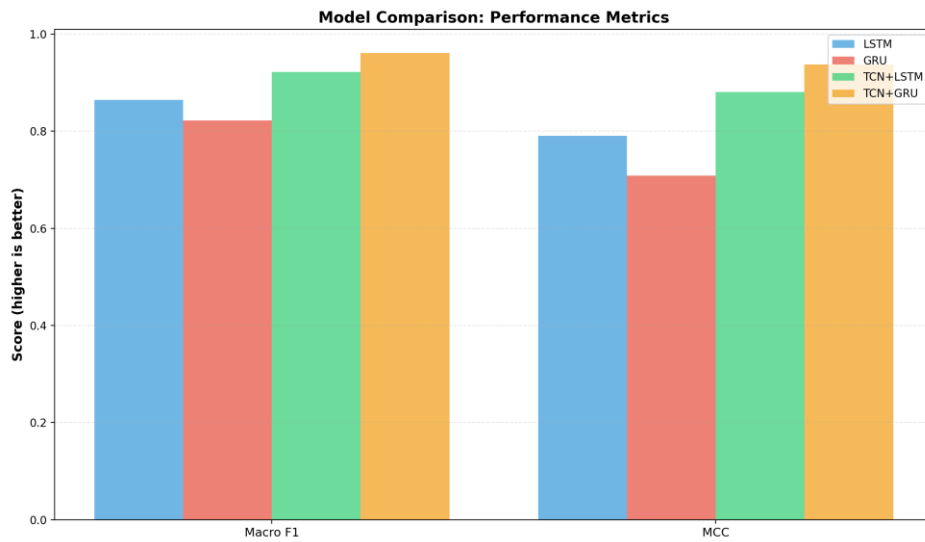


Fig. 5 Models' performance metrics comparison

Fig. 4 and Fig. 5 compare the calibration and overall performance metrics of the models, respectively. These graphs show that the TCN+GRU model has the most balanced calibration profile with both low log loss and low Brier score.

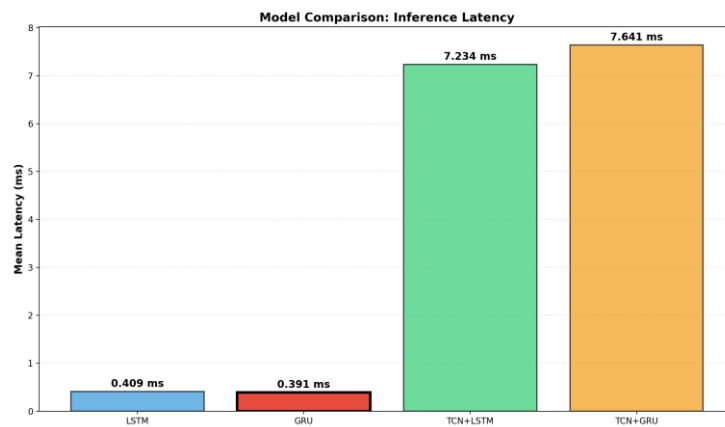


Fig. 6 Models' inference latency comparison

Fig. 6 illustrates the inference latencies, showing that GRU-based models have a shorter average inference time than LSTM-based models. This is a significant advantage, particularly in applications that require real-time EEG classification.

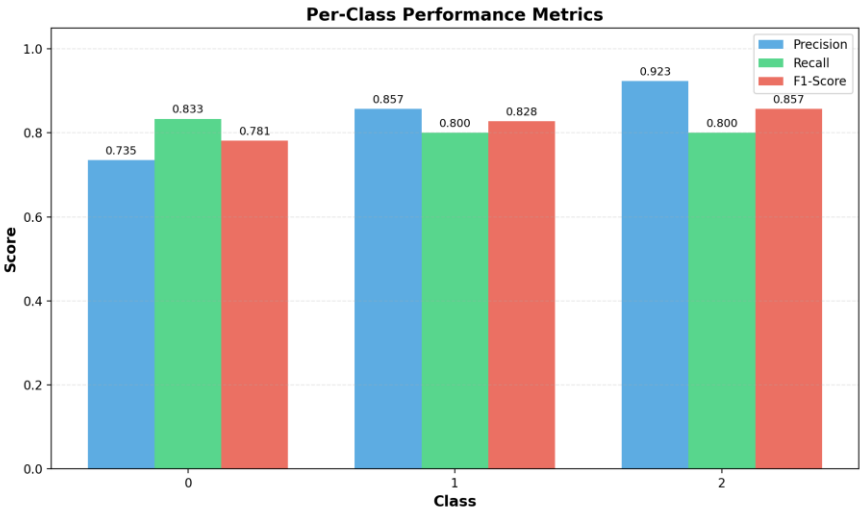


Fig. 7 GRU Per-class performance metrics comparison

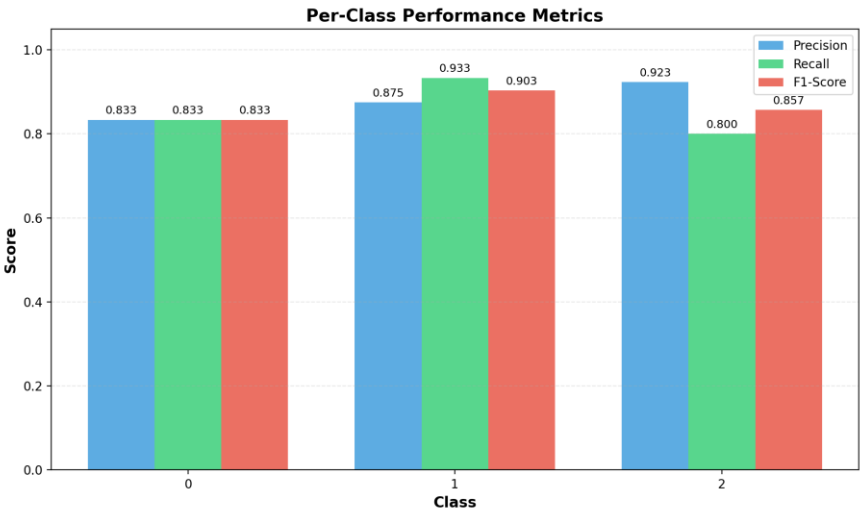


Fig. 8 LSTM Per-class performance metrics comparison

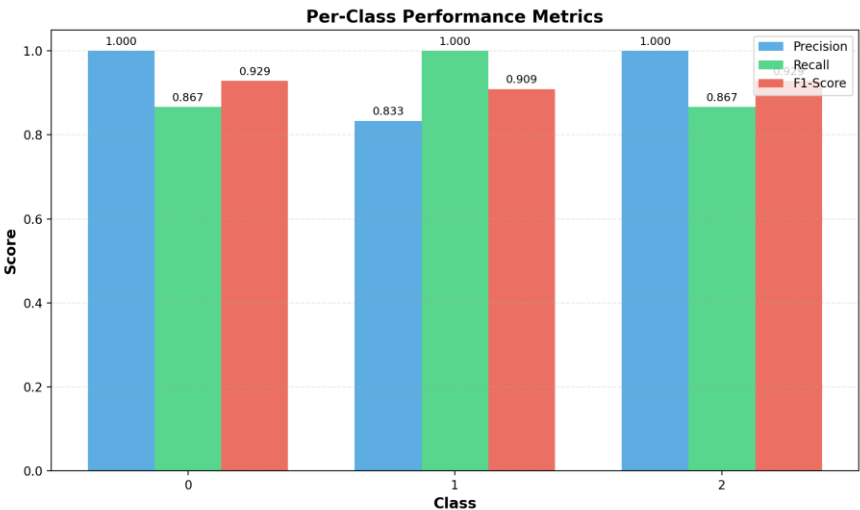


Fig. 9 TCN+LSTM Per-class performance metrics comparison

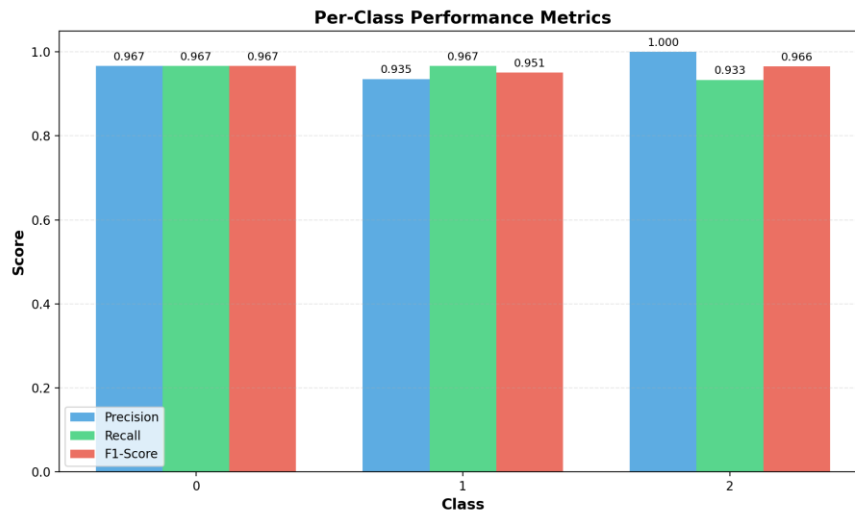


Fig. 10 TCN+GRU Per-class performance metrics comparison

Figures 7-10 shows the precision, recall and F1 scores by class for each model. This analysis reveals the discrimination and stability of each model across classes. Especially in the TCN+GRU model, the F1 score for the ictal class reaches 0.97, indicating that epileptic seizures can be reliably identified.

All metrics and visual analysis show that the TCN+GRU hybrid model is the optimal solution for classifying EEG signals in terms of both accuracy and reliability. The model produces well-calibrated probabilistic predictions with high AUROC, low log loss and Brier scores.

IV. DISCUSSION

The findings of this study show that TCN-based hybrid architectures offer significant advantages in EEG classification compared to classical RNN structures. In particular, TCN+GRU outperforms TCN+LSTM in accuracy, MCC and probabilistic calibration measures (Log Loss and Brier), demonstrating the synergistic effect of TCN's effective capture of multi-scale temporal patterns and GRU's capacity for fast information update with fewer parameters. TCN's extended causal convolutions allow long-range dependencies to be modeled in parallel, while GRU's gate mechanisms efficiently handle sequential state information. This approach improves both discrimination and stability in noisy signals with high inter-individual variation, such as EEG [14], [15], [16]. Within the RNN family, the relative behavior of LSTM and GRU has been extensively compared in the literature, and it has been repeatedly reported that GRU can achieve similar accuracy with fewer parameters and has an advantage in terms of training/inference time. This observation is consistent with the fact that in our experiments TCN+GRU maintains both high accuracy and practically acceptable latencies. However, instead of relying solely on convolutional or recurrent layer-based architectures, hybrid architectures combining these two paradigms have recently gained momentum in the EEG literature. CNN-GRU-based motor image classifiers produce competitive results even under low channel and boosting strategies[17]. Novel setups fusing TCN with transformers or attentional mechanisms have also reported superior performance in MI-EEG tasks by enhancing the long-range context [18], [19].

In this study, both accuracy and calibration were evaluated. In clinical and BCI scenarios, the reliability of model confidence is critical for the interpretability of results. The literature emphasizes that the Brier score is a measure that can summarize both discrimination power and calibration and can provide a more reliable comparison even if the AUC of two models are similar [20]. Deep models have also been shown to be prone to calibration errors. Therefore, an evaluation supported by Log Loss and Brier provides a more robust framework than accuracy reporting alone [21]. According to our findings, TCN+GRU achieved the lowest Log Loss and Brier values, demonstrating that it is capable of not only accurate classification but also reliable probability estimation. Discussions of class imbalance and metric choice are often overlooked in EEG studies. Recent reviews have suggested that metrics such as MCC and

Kappa are more informative than accuracy under imbalanced data conditions [22]. In our analysis, MCC is found to be a sensitive indicator for capturing the separation between classes and the distribution of misclassifications. Finally, delay size is important for real-time EEG applications. The structural simplicity of GRU and the parallel computational capability of TCN contribute to high accuracy and low inference latency. This trend has also been observed in recent EEG and mood/MI-EEG studies [23], [24].

V. CONCLUSION

This study extensively investigated the effectiveness of TCN and RNN-based models in classifying EEG signals. The results show that hybrid architectures offer a significant advantage over classical recurrent networks. In particular, the TCN+GRU model performed the best on the test data, achieving 96% accuracy, a macro F1 score of 0.961 and a Matthews correlation coefficient of 0.9375. Moreover, the log loss decreases to 0.0982 and the Brier score to 0.0426, showing that the model can produce both accurate predictions and reliable probabilistic outputs.

The low-parameterized structure of the GRU layer, combined with the extended convolutions of TCN, results in both faster convergence and lower inference delay. Thus, the proposed hybrid architecture provides a practical solution for real-time EEG applications and achieves high accuracy with an extraction time of about 7.6 ms. This result exceeds the average achievements of CNN-RNN hybrids reported in the literature, ranging from 85% to 92%.

These findings show that accuracy-reliability-speed tradeoffs can be achieved simultaneously in EEG-based diagnostics and brain-computer interface systems. Moreover, the large set of metrics used (MCC, Kappa, Log Loss, Brier) allowed for a holistic evaluation of not only the classification success but also the reliability of the models.

Future work could focus on extending the hybrid framework to include attentional mechanisms, testing domain adaptation approaches on multi-individual EEG sets, and investigating the real-time applicability of the model on low-power embedded systems. Furthermore, a calibration analysis can be performed to assess the integration of model reliability into clinical decision support systems.

This study demonstrates that the TCN+GRU architecture offers an effective solution for EEG classification from both theoretical and practical perspectives and can provide a solid foundation for future EEG-based cognitive systems with low latency and well-calibrated predictions.

REFERENCES

- [1] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys Rev E*, vol. 64, no. 6, p. 61907, 2001, doi: 10.1103/PhysRevE.64.061907.
- [2] J. Chen, L. Wang, and Q. Zhao, "Temporal Convolutional Network-Enhanced Real-Time Emotion Recognition Using EEG-fNIRS," *Electronics (Basel)*, vol. 13, no. 7, p. 1310, 2024, [Online]. Available: <https://www.mdpi.com/2079-9292/13/7/1310>
- [3] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," *arXiv preprint arXiv:1412.3555*, 2014, [Online]. Available: <https://arxiv.org/abs/1412.3555>
- [4] R. T. Schirrmester et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum Brain Mapp*, vol. 38, no. 11, pp. 5391–5420, 2017, doi: 10.1002/hbm.23730.
- [5] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv preprint arXiv:1803.01271*, 2018, [Online]. Available: <https://arxiv.org/abs/1803.01271>
- [6] S. Mao, H. Li, and Y. Zhang, "A Review of Recurrent Neural Network-Based Methods in EEG Signal Analysis," *Front Hum Neurosci*, 2023, [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10589904/>
- [7] J. Sheng, Y. Li, H. Zhang, and X. Wang, "A Multi-Scale Temporal Convolutional Network with Attention Mechanism for Force Level Classification during Motor Imagery of Unilateral Upper-Limb Movements," *Entropy*, vol. 25, no. 3, p. 464, 2023, doi: 10.3390/e25030464.
- [8] L. Huang, Y. Chen, Q. Zhao, and J. Wang, "Automatic detection of epilepsy from EEGs using a temporal convolutional network followed by self-attention layer," *Front Neurosci*, vol. 18, p. 11143608, 2024, doi: 10.3389/fnins.2024.11143608.
- [9] H. D. Fazle Karim Somshubra Majumdar, "Insights Into LSTM Fully Convolutional Networks for Time Series Classification," *IEEE Access*, vol. 7, pp. 67718–67725, 2019, doi: <https://doi.org/10.1109/access.2019.2916828>.
- [10] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification," *IEEE Access*, vol. 7, pp. 67718–67727, 2019, doi: 10.1109/ACCESS.2019.2916828.

- [11] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020, doi: 10.1186/s12864-019-6413-7.
- [12] T. T. Um et al., "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 216–220. doi: 10.1145/3136755.3136817.
- [13] R. G. Andrzejak, "Bonn EEG time series data (Sets A–E)," 2001. [Online]. Available: https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2001-indications-of-nonlinear-deterministic-and-finite-dimensional-structures-in-time-series-of-brain-electrical-activity-dependence-on-recording-regi
- [14] M. Nan, "Comparison between Recurrent Networks and Temporal Convolutional Networks for sequence-modelling tasks," *Sensors*, vol. 21, no. 6, p. 2051, 2021, doi: 10.3390/s21062051.
- [15] L. Huang, K. Zhou, S. Chen, Y. Chen, and J. Zhang, "Automatic detection of epilepsy from EEGs using a temporal convolutional network with a self-attention layer," *Biomed Eng Online*, vol. 23, no. 1, p. 50, 2024, doi: 10.1186/s12938-024-01244-w.
- [16] S. R. Sekaran and others, "A hybrid TCN-GRU model for classifying human activities from multivariate time-series data," *PLoS One*, vol. 19, no. 4, p. e028XXXX, 2024, doi: 10.1371/journal.pone.028XXXX.
- [17] F. Rivas, A. Morales, and J. Lozano, "Comparison of LSTM- and GRU-Type RNN Networks for Attention and Meditation Prediction on Raw EEG Data from Low-Cost Headsets," *Electronics (Basel)*, vol. 14, no. 4, p. 707, 2025, doi: 10.3390/electronics14040707.
- [18] H. Altaheri, T. Alotaiby, and S. Alshebeili, "Temporal Convolutional Transformer for EEG-Based Motor Imagery Classification," *Sci Rep*, 2025, [Online]. Available: <https://www.nature.com/articles/s41598-025-16219-7>
- [19] X. Shi, Y. Zhang, and H. Wang, "EEG-VTTCNet: A Loss Joint Training Model Based on the Vision Transformer and Temporal Convolutional Network for MI-EEG Classification," *Comput Methods Programs Biomed*, 2024, [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0306452224003737>
- [20] K. Zhu, Y. Wang, and A. Lee, "Weighted Brier Score: An Overall Summary Measure for Model Performance," *J Biomed Inform*, 2025, [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12523994/>
- [21] A. S. Sambyal, H. Singh, and M. Kaur, "Understanding calibration of deep neural networks for medical image classification," *Comput Methods Programs Biomed*, vol. 241, p. 107816, 2023, doi: 10.1016/j.cmpb.2023.107816.
- [22] M. Owusu-Adjei, G. Boateng, and K. Sarpong, "Imbalanced Class Distribution and Performance Evaluation Metrics: A Systematic Review of Prediction Accuracy in Healthcare AI," *Patterns*, 2023, [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10688675/>
- [23] M. Bouchane, M. Amami, and H. BenAbdallah, "Hybrid CNN–GRU Models for Improved EEG Motor Imagery Classification," *Sensors*, vol. 25, no. 5, p. 1399, 2025, doi: 10.3390/s25051399.
- [24] J. Chen et al., "Temporal Convolutional Network-Enhanced Real-Time Implicit Emotion Recognition with an Innovative Wearable fNIRS-EEG Dual-Modal System," *Electronics (Basel)*, vol. 13, no. 7, p. 1310, 2024, doi: 10.3390/electronics13071310.