UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Domain-Adaptive Transfer Learning with Recurrent Neural Networks for Cross-Subject P300 Speller Classification

Permalink

https://escholarship.org/uc/item/59j8r8z8

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Luo, Yun
Lu, Wentao
Huang, Xi
et al.

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Domain-Adaptive Transfer Learning with Recurrent Neural Networks for Cross-Subject P300 Speller Classification

Yun Luo[1], **Wentao Lu**[1], **Xi Huang**[1], **Wei Gao (gaow@m.scnu.edu.cn)**[2], **& Jiahui Pan**[1]
[1]School of Artificial Intelligence, South China Normal University, Foshan, China
[2]School of Artificial Intelligence, South China Normal University, Guangzhou, China

Abstract

Brain-computer interface (BCI) research has advanced significantly, yet cross-subject P300 decoding remains challenging due to highly variable EEG signals. To address this, we propose a novel domain adaptation framework integrating transfer learning and recurrent neural networks (RNNs). The framework incorporates a gated recurrent unit (GRU) layer to effectively capture temporal dependencies of EEG signals, thereby mitigating temporal misalignment issues. By calculating transfer loss and applying time-step weighting, the framework enhances classification. Furthermore, feature-space adaptive alignment is employed to reduce inter-subject variability, lowering subject dependency and enabling accurate cross-subject character recognition. Experimental results demonstrate that the proposed method achieves an average character recognition accuracy of 92.38%, significantly surpassing traditional P300 recognition algorithms. This indicates that the method effectively mitigates the impact of temporal misalignment in EEG data as well as the high subject-dependence of BCI systems.

Keywords: brain-computer interface (BCI); domain adaptation (DA); transfer learning (TL); recurrent neural network (RNN); P300

Introduction

In recent years, Brain-computer interfaces (BCIs) have found extensive applications in healthcare, military, and manufacturing domains. Many applications are based on event-related potentials (ERP), among which N200, P300, and N400 are the most dominant. The P300 component is characterized by a positive deflection evoked by unexpected stimuli within a sequence of distracting stimuli. It typically manifests 250–300 ms post-stimulus. Owing to its clarity, stability, and reproducibility, P300-based BCIs hold significant promise. The stimuli used to evoke P300 can be visual (Ma et al., 2024), auditory (Furdea et al., 2009), tactile (Brouwer & Van Erp, 2010), or multimodal (Jin et al., 2017). A classical P300 paradigm employs an n×n character matrix with randomly intensified rows and columns (Schalk et al., 2004; Won et al., 2022). Given that target stimuli are fewer than non-target ones, when the row and column containing the target character flash, the system captures and analyzes EEG signals to predict the intended character. This paradigm is known as the P300 speller BCI system.

Latency jitter, an inherent characteristic of P300 signals, frequently results in temporal misalignment. Although the data distributions across different sessions of the same subject exhibit similarities, they are not identical, and inter-subject variability is even more pronounced (Xu et al., 2023). Therefore, a crucial strategy to enhance classification performance is to align the data distributions between the source and target domains. To address this issue, Zanini et al. (2017) proposed Riemannian Alignment (RA), which mitigates distributional discrepancies across sessions. RA utilizes the Riemannian mean as a reference covariance matrix for each session and applies an affine transformation to align session-specific data, thereby enhancing comparability. In Xia et al. (2022), sample reweighting was employed to align the data distributions between the source and target domains, culminating in the Adaptive Source-Free Augmentation (ASFA) framework. However, this approach remains reliant on a well-trained source model. Y. Gao et al. (2023) proposed a multi-manifold distribution alignment (MMDA) method based on Riemannian and Grassmann manifolds, which not only enhances the inter-class separability of MI-EEG but also preserves the structure of EEG features. To address the challenges of temporal misalignment and inter-subject variability, we introduce a novel approach that focuses on the temporal distribution of the time series. Specifically, we propose an adaptive ADA threshold mechanism, which identifies time points that exhibit consistent discriminative patterns across both source and target domains. This mechanism effectively aligns the P300 response peaks across subjects. The resulting soft alignment strategy alleviates the information bias caused by fixed temporal windows, thereby enhancing classification performance.

Traditional BCIs often exhibit suboptimal performance in cross-subject settings due to user-specific differences and experimental environment variability. Consequently, most prior studies have concentrated on within-subject experiments. For instance, The work in Liu et al. (2018) partitions single-subject data into training and testing sets to minimize the impact of inter-subject variability. Other studies have utilized conventional machine learning methods to evaluate P300 classifiers and feature extraction techniques within the context of within-subject scenarios. However, reducing subject-dependence is essential for enhancing BCI usability. To this end, researchers have explored cross-subject classification using logistic regression (LR), linear discriminant analysis (LDA), naive Bayes (NB) (Zhang et al., 2020), support vector machines (SVM), and continuous wavelet transform (CWT) (Hwaidi & Chen, 2022). Although these methods mitigate subject-dependence to some extent, they often fail to capture the nonlinear characteristics of multichannel EEG data, which

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

are rich in temporal and spatial information. Deep learning approaches, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown considerable promise. Studies (Cecotti & Graser, 2011; W. Gao et al., 2021) have proposed CNN-based models for P300 detection, while Aljlaly et al. (2024), Eldawlatly (2024), and Maddula et al. (2017) have employed RNNs to capture the sequential nature of EEG signals. RNNs are particularly well-suited for modeling temporal dependencies and capturing ERPs like P300. Transfer learning has been increasingly adopted across various fields, and recent studies (Chen & Huang, 2022; Li et al., 2020; Liang et al., 2024) have successfully applied it to brain-computer interfaces (BCIs). For instance, Wang et al. (2024) proposed an unsupervised domain adaptation (DA) method based on Adaptive Euclidean Alignment to mitigate inter-subject variability. Building on these advances, we present a novel EEG decoding framework that integrates DA techniques with recurrent neural networks (RNNs) for cross-subject P300 speller BCI systems.

The main contributions of this study are as follows:

1. We propose a novel temporal distribution modeling strategy incorporating an ADA-based weighting threshold mechanism. By applying weighted selection over time steps, this strategy automatically identifies and retains critical time points that exhibit robust responses in both domains while filtering out non-discriminative or inconsistent temporal features.
2. We develop a general EEG decoding framework that integrates domain adaptation from transfer learning with recurrent neural networks. The GRU layer within this framework enhance the ability to extract temporal information, while the domain adaptation module mitigates the impact of inter-subject variability on decoding.
3. Experimental results demonstrate that our framework achieves remarkable performance on a P300 BCI speller dataset. It attains an average offline classification accuracy of 92.38%, surpassing traditional methods.

Proposed Method

Data Preprocessing

First, each channel was re-referenced to the common average reference (CAR). Then, the EEG data was filtered using a 4th-order Butterworth filter with cut-off frequencies between 0.5 Hz and 10 Hz. After that, the data was segmented into epochs from 0 ms to 600 ms relative to each stimulus onset. Baseline correction was performed using the average value of the data from -200 ms to 0 ms. Typically, baseline correction is applied using values recorded before stimulus onset. Next, the data was downsampled to a sampling rate of 21 Hz. After preprocessing, the final data shape is [trials×12×32], where "trials" represents the total number of trials, 12 refers to the number of time points obtained after downsampling, and 32 represents the number of EEG channels.

Model Construction

Our proposed model consists of four main modules, as illustrated in Figure 1. First, the preprocessed data is fed into Temporal Refinement Network for feature extraction. On one hand, the extracted features are input into Time-step Weighted Calculation and Normalization module for time-step weighted computation. The weighted time steps are subsequently normalized and fed back into Temporal Refinement Network. Meanwhile, the normalized time weights undergo weighted summation to obtain the final feature representation. Finally, the final feature representation is input into Feature Transformation module for feature transformation. The transformed features are then passed to the classification module for classification, and the classification results are further sent to the output layer for the final output.

On the other hand, the extracted features are fed into Transfer Loss Calculation module for transfer loss computation, which measures the feature distribution discrepancy between the source and target domains. This process further enhances the feature representation capability of the GRU, allowing the model to learn the data distribution more effectively.

Temporal Refinement Network

This module extracts EEG data features using a two-layer GRU. The first GRU layer captures primary temporal features, while the second GRU layer further refines feature representations, making the temporal dependencies in EEG sequences clearer. The GRU layers process EEG data step by step, incorporating the current input state and updating information adaptively. The GRU mainly relies on a gating mechanism to extract EEG features, particularly the update gate and reset gate, which dynamically control the flow of information.

Reset Gate Calculation The reset gate determines how much of the previous hidden state should be forgotten in the current computation. Let the output at each time step be denoted as x_T, where $T = \{1, 2, 3, \ldots, t\}$, and let the hidden state at each time step be denoted as h_T, where $T = \{1, 2, 3, \ldots, t\}$, with t representing the total number of time steps. The reset gate is computed as:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t] + b_r\right) \tag{1}$$

where x_t is the input at the current time step, h_{t-1} is the hidden state from the previous time step, W_r and b_r are the weight matrix and bias of the reset gate, and σ is the activation function.

The reset gate determines how much of the previous hidden state h_{t-1} should be forgotten in the current computation. If r_t is close to 0, the model ignores past information and relies mainly on the current input for feature extraction.

Update Gate Calculation The update gate determines how much of the previous time step's information should be retained in the current state. Let the time step information be denoted as N_j, where $j = \{1, 2, 3, \ldots, t\}$, and let the output

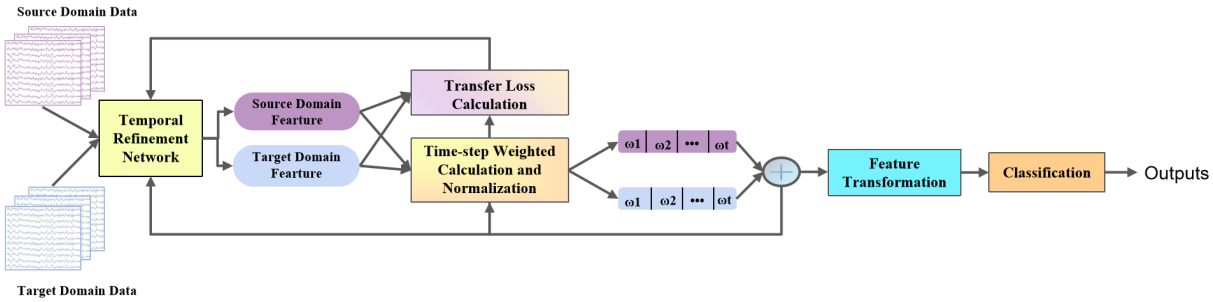

Figure 1: Model Architecture: The proposed framework integrates four core modules—Temporal Refinement Network for sequential feature extraction using stacked GRUs; Transfer Loss Calculation via MMD for domain discrepancy quantification and feature adaptation; Time-step Weighted Calculation and Normalization for noise suppression and subject-specific feature alignment; and Feature Transformation for enhanced adaptability and overfitting mitigation—collectively optimizing EEG signal classification across subjects.

at time step j be denoted as x_j, with t representing the total number of time steps. The update gate is computed as:

$$z_j = \sigma\left(W_z \cdot [N_{j-1}, x_j] + b_z\right) \tag{2}$$

where W_z and b_z are the weight matrix and bias of the update gate, and σ is the activation function.

The update gate controls how much of the previous time step's information is retained. If z_j is close to 1, more past information is preserved.

Feature Enhancement Through Gating Mechanisms The reset gate helps capture local information, while the update gate facilitates long-term information retention. Through the dynamic control of these two gating mechanisms, the model maximizes the retention of both short-term and long-term EEG features. The features extracted by the GRU layers are then passed to Time-step Weighted Calculation and Normalization module for time-step weighting. The weighted time steps are fed back into the GRU layers, adjusting the hidden state of the GRU to enhance classification performance.Unlike conventional RNN approaches that rely on a single temporal modeling layer, our method integrates multiple GRU layers and incorporates their hidden states into the transfer loss computation. This design enables the model to capture both low-level temporal patterns and high-level semantic representations. Such an innovation allows for more comprehensive feature extraction and contributes to improved model performance.

Time-step Weighted Calculation and Normalization

In EEG data, different time-steps have different levels of importance for the final classification. For example, in the P300 task, target stimuli typically produce significant neural responses at specific time-points, while at other time-points, there might be noise or other interfering information. Therefore, calculating the weight for each time-step can effectively help the model suppress noise and focus on important time-steps. For the feature sequence extracted by the GRU layer, we denote it as T, where $T = \{T_1, T_2, T_3, \ldots, T_K\}$, with $K \in \mathbb{N}^+$. The weight for the time-step w_t is computed as:

$$w_t = \sigma(W_w^T T_k) \tag{3}$$

where W_w belongs to $\mathbb{R}^{m \times n}$ and is a learnable parameter, σ represents the sigmoid activation function.

After computing the temporal step weights using the above formula, we use SoftMax function to normalize them to prevent certain time steps from having excessively large or small weights. Finally, the normalized temporal step weights are fed back into the GRU layer to enhance its ability to extract important features. Simultaneously, the normalized temporal step weights, after filtering by the ADA weight threshold, are summed to obtain the final feature representation, which is then input into the Feature Transformation module. In contrast to global distribution alignment methods that use fixed temporal windows, our approach introduces a dynamic weighted alignment strategy that automatically focuses on key time intervals. This mechanism effectively compensates for subject-specific variations in P300 latency and suppresses noise interference from non-P300 segments.

Innovatively, we adopt the average of the summed time-step weights from the source and target domain weighted features as the ADA weight threshold. For a given feature sequence, if a time-step weight exceeds the ADA weight threshold, that time step is selected as a critical time step and included in the weighted summation of the feature sequence. This helps mitigate information bias caused by temporal misalignment between the source and target domains, thereby improving classification performance. The ADA weight threshold can adaptively adjust according to the weight distribution of critical time steps within the feature sequence. This ensures that the threshold captures the most salient time steps shared across both domains, effectively aligning the P300 response peaks across subjects. As a result, this mechanism alleviates the challenge of feature alignment caused by temporal discrepancies in cross-domain learning. The calculation formula

of the ADA weight is as follows:

$$\theta_{ADA} = \frac{1}{T}\sum_{t=1}^{T} w_t \tag{4}$$

where $W_t = (w_1, w_2, w_3, \ldots, w_T)$, and T is the number of key time steps.

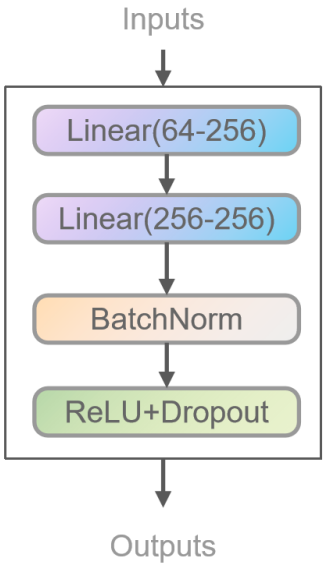

Figure 2: Feature transformation module. The module comprises two linear layers, a BatchNorm layer, a Dropout layer, and a ReLU activation layer, designed to map the input data to a higher-dimensional space. This architecture enhances model adaptability, improves feature representation, and prevents overfitting.

Feature Transformation

As illustrated in Figure 2,the proposed module consists of two fully connected layers followed by Dropout, non-linear activation, and Batch Normalization. The weighted feature vector is initially expanded from 64 to 256 dimensions by the fully connected layers. A Dropout layer is then applied to mitigate overfitting, while the subsequent activation function introduces non-linearity to improve representation. Finally, Batch Normalization is implemented to ensure distribution stability, address gradient issues, and accelerate training. This design aims to improve both the generalization ability and stability of the model.

Transfer Loss Calculation

In cross-subject experiments, EEG data distributions vary among different subjects. Therefore, this module is used to compute the feature distribution differences between the source and target domains, thereby optimizing the feature extraction capability of the GRU layer and improving transfer learning performance. In this module, we primarily employ Maximum Mean Discrepancy (MMD) to measure the feature distribution differences between different EEG subjects, enhancing the model's generalization ability. We assume that the GRU sample features in the source domain are denoted as H_S, where $H_S = \{h_1^s, h_2^s, h_3^s, ..., h_i^s\}$, and i represents the number of source domain samples. Similarly, the GRU sample features in the target domain are denoted as H_T, where $H_T = \{h_1^t, h_2^t, h_3^t, ..., h_j^t\}$, and j represents the number of target domain samples. The function $\phi(\cdot)$ is a nonlinear mapping function that projects the features into a high-dimensional space to enhance distribution contrast.

$$\mathcal{L}_{\text{MMD}} = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s} \phi(h_i^s) - \frac{1}{n_t}\sum_{j=1}^{n_t} \phi(h_j^t) \right\|^2 \tag{5}$$

Here, n_s is the number of samples in the source domain, and n_t is the number of samples in the target domain. We use MMD to compute the feature distribution differences between the source and target domains. During backpropagation, the Transfer Loss influences the gradient updates of the GRU, enabling it to learn more stable features for the target domain and improve adaptability.

Compared to Kullback-Leibler (KL) divergence, which requires overlapping regions between two distributions to compute distribution differences, MMD directly measures the mean difference between source and target domain feature distributions. This method does not require any assumptions about the data distribution shape (e.g., Gaussian distribution, exponential distribution). Additionally, MMD is computationally simple and easy to optimize. Thus, using MMD to compute distribution differences and feeding them back to the GRU layer for optimization is a straightforward yet effective approach in our experiment.

Classification

The feature representations processed by the Feature Transformation module are fed into the classifier for binary classification as target or non-target. Subsequently, the row and column indices in the 6×6 matrix with the highest scores (i.e., those most frequently identified as targets) are retrieved to locate and output the target character. Leave-one-subject-out cross-validation is employed for both training and evaluation.

Experiment and Results

Dataset

The experiment used a public EEG dataset from Won et al. (2022), covering 55 healthy subjects (41 males, 14 females). EEG data were recorded via the 10-20 system at 512 Hz across 32 channels. The paradigm followed the classic speller setup shown in Figure 3. Each stimulus lasted 125 ms with a 62.5 ms interval, and each character needed 15 repeated stimulations. The experiment had calibration and test phases. In calibration, subjects spelled two words ("BRAIN" and "POWER") without feedback. In testing, they spelled four words ("SUBJECT", "NEURONS", "IMAGINE", and "QUALITY") with real-time feedback. All 55 subjects spelled the same words, undergoing 1800 calibration trials and 5040 test trials each.

Experiment Settings

The proposed model was implemented and trained using Python 3.10 and PyTorch on an RTX 4090. Given the 55-subject dataset's scale, leave-one-subject-out cross-validation was adopted. In each iteration, 54 subjects' data trained the model, and the remaining one's data tested it. The batch size was 64, and the initial learning rate was set to 0.0001. Gradient clipping capped the maximum gradient norm at 1.0, and the AdamW optimizer was used for efficient training.

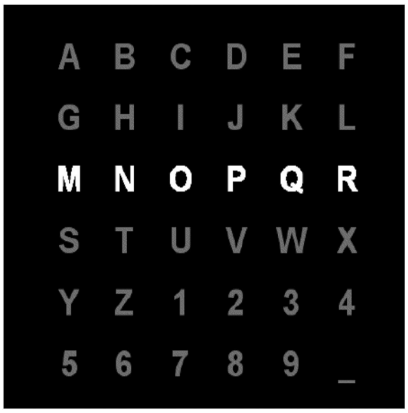

Figure 3: P300 Speller Paradigm. Illustration of the row/column flashing mechanism used in the P300 speller. Each character's stimulus is repeated 15 times over a complete sequence of 12 flashes. The training phase comprises two sub - phases, while the testing phase includes four sub - phases.

Experiment Results

The average letter recognition accuracy was used to assess the proposed model's decoding performance, reflecting correct recognitions among 28 tested letters. As a result, an average accuracy of **92.38%** in the 15th epoch is achieved by the proposed method (see TABLE 1). To assess the statistical significance of this improvement, a Wilcoxon signed-rank test was conducted. The results confirmed that our model substantially outperforms the SWLDA baseline ($p < 0.05$). 15 subjects achieved **100%** accuracy, 61% exceeded **90%**, yet four subjects stayed under **80%**. The proposed method was also compared with other representative ones evaluated on the same dataset.

1. **SWLDA** (Won et al., 2022): A stepwise weighted linear discriminant analysis (SWLDA) method for feature selection and classification.
2. **CNN_ZT** (Lee et al., 2020): A zero-training CNN model. This model aims to classify EEG data without individual calibration, meaning it can be directly applied to new subjects without requiring additional training data. It is pre-trained on cross-subject EEG data and tested directly on new subjects.
3. **SWLDA_ZT** (Lee et al., 2020): A zero-training version of the stepwise linear discriminant analysis (SWLDA) method.

Table 1: Comparison of recognition accuracy with traditional methods. The sign (≈) denotes estimated values based on implicit reporting in original articles.

Model Name	Min	STD	Mean Acc. (15th)
SWLDA	≈ 40%	≈ 13.00	92.00%
CNN_ZT	42.86%	13.65	89.22%
SWLDA_ZT	17.86%	23.57	69.80%
Our Model	**64.29%**	**11.24**	**92.38%**

By comparing with the aforementioned methods, our proposed model achieves a mean accuracy **22.58%** higher than that of SWLDA_ZT (Lee et al., 2020). The experimental results demonstrate that the SWLDA model without calibration (SWLDA_ZT) yields notably lower performance in cross-subject scenarios. Compared to SWLDA_ZT, our model can capture and model the nonlinear relationships between features. Furthermore, our proposed model outperforms the CNN_ZT model introduced by Lee et al. (2020), achieving an improvement of **3.16%** in average character recognition accuracy. This notable performance gain highlights the efficacy of our model architecture in capturing discriminative features for the task. This is because RNNs inherently capture temporal dependencies between features, which is particularly beneficial for EEG data, a type of time-series data. As a result, RNNs can extract crucial EEG features more effectively than CNNs. Moreover, we utilize Maximum Mean Discrepancy (MMD) to compute the distribution discrepancy between the source and target domains, thereby optimizing and adjusting the model's feature extraction capability. Meanwhile, compared to the SWLDA model proposed by Won et al. (2022), the proposed model achieves a higher accuracy, demonstrating stronger feature modeling capability and robustness.To better illustrate the experimental results, the accuracy of the proposed model at the 15th epoch for subjects s01 to s55 is shown in Figure 4

The experimental results above indicate that the proposed method achieves improved P300 character recognition performance compared to traditional approaches. This highlights the effectiveness of combining domain adaptation with recurrent neural networks in reducing individual variability and enhancing character recognition accuracy, especially in critical scenarios such as assistive communication systems for people with motor disabilities.

Ablation Study

After obtaining the experimental results, we conducted an ablation study to evaluate the effectiveness of each module in our proposed model. Specifically, we performed ablation experiments by removing the TL (Transfer Loss) layer and the

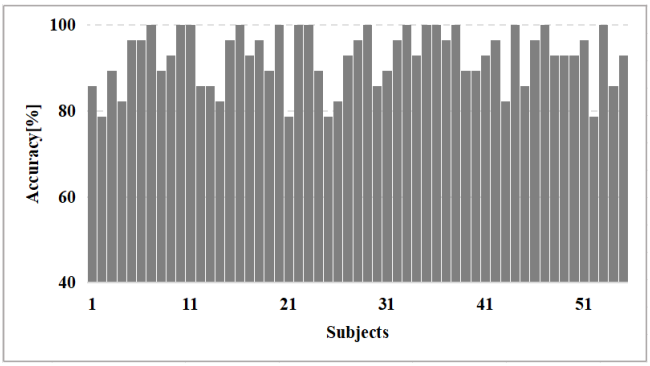

Figure 4: Character recognition accuracy of each subject (S01 to S55) at the 15th epoch.

TWN (Time-step Weighted Calculation and Normalization) layer separately. Finally, we removed both modules simultaneously, leaving only the GRU layer and the feature transformation layer. The experimental results demonstrate that both the TL and TWN modules significantly enhance model performance. To ensure the rigor and reliability of our experiments, we maintained a single-variable change in each ablation setting. Each ablation experiment was conducted over 15 epochs, and the average character recognition accuracy was used as the evaluation metric. The full model proposed in this study is used as the baseline for all variants. The ablation variants are described as follows:

- **w/o TL**: This variant removes the Transfer Loss submodule.
- **w/o TWN**: This variant removes the Time-step Weighted Calculation and Normalization sub-module.
- **w/o TL AND TWN**: This variant removes both the Transfer Loss and Time-step Weighted Calculation and Normalization sub-modules.

The average character recognition accuracies after 15 training epochs are summarized as follows: w/o TL: 89.29%, w/o TWN: 85.71%, and w/o TL AND TWN: 78.57%. To visually and intuitively demonstrate the performance contributions of each module, these results are illustrated in Figure 5. From the figure, it is evident that removing the TL module results in a 3.09% decrease in average accuracy. This can be attributed to the elimination of the MMD-based optimization process, which is crucial for fine-tuning the feature extraction layer. Without it, the model's generalization capability is significantly reduced, especially in cross-subject scenarios. The w/o TWN model exhibits a more substantial decrease of 6.67% in average accuracy, highlighting the importance of the time-step weighting and normalization (TWN) module. Given that target stimuli are relatively rare compared to non-target stimuli, the weighted feature calculation is essential for emphasizing important target features, thereby enhancing recognition accuracy and improving the feature extraction process.

In the w/o TL AND TWN setting, the accuracy drops drastically by 13.81%, indicating severe performance degradation when both modules are removed. Without these modules, the model essentially degenerates into a plain RNN framework, lacking mechanisms for emphasizing important features and adjusting the feature representation during training, which leads to suboptimal performance. Through these ablation studies, we demonstrate the critical roles of both the TL and TWN modules in our proposed model. These findings provide strong empirical evidence supporting the effectiveness of our architecture, especially for cross-subject EEG classification tasks. Moreover, they offer valuable insights and confidence for future applications in online BCI experiments.

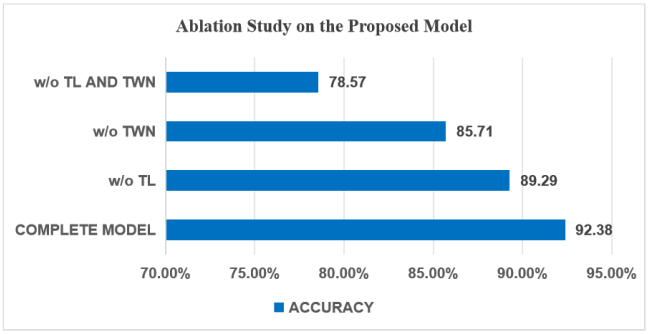

Figure 5: Ablation Study Results. Character recognition accuracy after 15 epochs.

Limitations and Future Work

Despite promising results, this study is limited by its reliance on a single public dataset and offline analysis. Future research will focus on cross-dataset validation and real-time implementation. Furthermore, we plan to benchmark our framework against recent methods like MMDA and Riemannian-based techniques to further verify its robustness and competitive edge in diverse BCI environments

Conclusion

This paper proposes a method that combines adaptive transfer learning with recurrent neural networks (RNNs) to achieve cross-subject P300 detection classification. Experimental results show that the method combining adaptive transfer learning with recurrent architecture effectively solves the problem of strong topic dependencies. In future work, we will further refine our method and integrate it into BCI systems, aiming for potential applications in the auxiliary assessment of disorders of consciousness (DoC) patients.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant 62576142 and Guangdong Talent Plan under grant 2024TQ08X162.

References

Aljlaly, A. M., Elshami, I. F., & Almijbari, A. (2024). Long short-term memory employed for classifying P300-based BCI. *Proceedings of the 2024 IEEE 4th International Maghreb Meeting on Science, Technology, Automation, Control, and Computer Engineering (MI-STA)*, 666–671.

Brouwer, A.-M., & Van Erp, J. B. (2010). A tactile P300 brain-computer interface. *Frontiers in Neuroscience*, *4*, 1440.

Cecotti, H., & Graser, A. (2011). Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(3), 433–445.

Chen, X., & Huang, Z. (2022). A P300 BCI calibration-free algorithm based on intersubject transfer and reinforcement learning. *Proceedings of the 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–6.

Eldawlatly, S. (2024). On the role of generative artificial intelligence in the development of brain-computer interfaces. *BMC Biomedical Engineering*, *6*(1).

Furdea, A., Halder, S., Krusienski, D., Bross, D., Nijboer, F., Birbaumer, N., & Kübler, A. (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, *46*, 617–625.

Gao, W., Yu, T., & Yu, J. G. (2021). Learning invariant patterns based on a convolutional neural network and big electroencephalography data for subject-independent P300 brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 1047–1057.

Gao, Y., Liu, Y., She, Q., & Zhang, J. (2023). Domain adaptive algorithm based on multi-manifold embedded distributed alignment for brain-computer interfaces. *IEEE Journal of Biomedical and Health Informatics*, *27*, 296–307.

Hwaidi, J. F., & Chen, T. M. (2022). Classification of motor imagery EEG signals based on deep autoencoder and convolutional neural network approach. *IEEE Access*, *10*, 48071–48081.

Jin, J., Zhang, H., Daly, I., Wang, X., & Cichocki, A. (2017). An improved P300 pattern in BCI to catch user's attention. *Journal of Neural Engineering*, *14*(3), 036001.

Lee, J., Won, K., Kwon, M., Jun, S. C., & Ahn, M. (2020). CNN with large data achieves true zero-training in online P300 brain-computer interface. *IEEE Access*, *8*, 74385–74400.

Li, F., Xia, Y., Wang, F., Zhang, D., Li, X., & He, F. (2020). Transfer learning algorithm of P300-EEG signal based on XDAWN spatial filter and riemannian geometry classifier. *Applied Sciences*, *10*(5), 1804.

Liang, Z., Zheng, Z., Chen, W., Pei, Z., Wang, J., & Chen, J. (2024). A novel deep transfer learning framework integrating general and domain-specific features for EEG-based brain-computer interface. *Biomedical Signal Processing and Control*, *95*(Pt. B), 106311.

Liu, M., Wu, W., Gu, Z., et al. (2018). Deep learning based on batch normalization for P300 signal detection. *Neurocomputing*, *275*, 288–297.

Ma, T., Falkenburg, B., & Speier, W. (2024). Incorporating flash adjacency into the classifier for a language model-based P300 speller. *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, 1–8.

Maddula, R., Stivers, J., Mousavi, M., Ravindran, S., & de Sa, V. (2017). Deep recurrent convolutional neural networks for classifying P300 BCI signals. *Proceedings of the 7th Graz Brain-Computer Interface Conference (GBCIC)*, *201*, 18–22.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, *51*(6), 1034–1043.

Wang, H., Han, H., Gan, J. Q., & Wang, H. (2024). Lightweight source-free domain adaptation based on adaptive euclidean alignment for brain-computer interfaces. *IEEE Journal of Biomedical and Health Informatics*.

Won, K., Kwon, M., Ahn, M., & Jun, S. C. (2022). EEG dataset for RSVP and P300 speller brain-computer interfaces. *Scientific Data*, *9*(1), 388.

Xia, K., Deng, L., Duch, W., & Wu, D. (2022). Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, *69*, 3365–3376.

Xu, D.-q., Sun, Y.-j., & Li, M.-a. (2023). An adaptive cross-class transfer learning framework with two-level alignment. *Biomedical Signal Processing and Control*, *86*, 105155.

Zanini, P., Congedo, M., Jutten, C., Said, S., & Berthoumieu, Y. (2017). Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, *65*, 1107–1116.

Zhang, G., Davoodnia, V., Sepas-Moghaddam, A., Zhang, Y., & Etemad, A. (2020). Classification of hand movements from EEG using a deep attention-based LSTM network. *IEEE Sensors Journal*, *20*(6), 3113–3122.