Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/154/2025.TJ23002

Enhancing Brain-Computer Interface Performance and Security through Advanced Artificial Intelligence Techniques

Weijie Liu

Guangdong University of Technology, Guangdong, China
3123009086@mail2.gdut.edu.cn

Abstract: The brain-computer interface has become a rapidly developing field, but it has also brought many problems with its development. The main issues are the sparse amount of brain-computer interface data, the inaccurate decoding and classification of data, and the data security of the brain-computer interface. With the development of artificial intelligence, artificial intelligence also provides solutions to many problems. This study mainly uses artificial intelligence algorithms to solve these problems. This paper reviews the integration of artificial intelligence techniques—specifically transfer learning, generative adversarial networks (GANs), Transformer models, and federated learning—to address critical challenges in brain-computer interfaces (BCIs), including data scarcity, classification accuracy, and data security. The hybrid model has many outstanding performances in solving the brain-computer interface problem, and this paper mainly mentions the joint extraction of spatiotemporal features of the CNN-Transformer to make up for the shortcomings of a single model and improve the overall performance. The GAN-TL hybrid model can effectively reduce the influence of individual differences on the model. This paper illustrates the advantages of the hybrid model, which is also the main direction of future research. It highlights how hybrid AI models significantly enhance BCI performance while outlining current limitations and future research directions to ensure robust, efficient, and secure BCI applications.

Keywords: brain-computer interfaces, Convolutional neural networks, Generate adversarial networks, Federated Learning, Transformer

1. Introduction

With the development of science and technology, brain-computer interfaces are also making continuous progress, but they face many problems. The rise of artificial intelligence has also brought solutions and more possibilities. Over the past decade, the continuous advancement of brain-computer interfaces has also brought new possibilities to the development of many industries. Brain-computer interfaces were first used in the medical industry [1]. Brain-computer interfaces (BCIs) enable direct communication between the human brain and external devices, with profound implications in medical rehabilitation, assistive technologies, and advanced human-machine interaction. Despite these advances, challenges such as limited training data, difficulty in accurately classifying EEG signals, and concerns over user privacy persist. The rapid advancement of AI technologies provides robust solutions, making their integration crucial for next-generation BCIs.

There are many noises and artifacts in EEG signals, and they have low spatial resolution, high temporal resolution, low signal-to-noise ratio, and large individual differences. In addition, EEG signals indirectly reflect an individual's private information. So, these characteristics lead to some problems: there are different differences between subjects, and it is difficult to build a universal model that can be adapted to each individual and different topic. EEG data is difficult to decode, and EEG signals are difficult to process and classify. There is a threat of leakage of brain-computer interfaces. These problems can be solved by artificial intelligence algorithms, mainly through machine learning and deep learning algorithms. However, these algorithms also bring new problems as well as solving issues: first, the fusion model can play its advantages, but its disadvantages may also be superimposed, for example, the interaction of GAN-TL models may exacerbate the collapse of the model. Secondly, the classification problem has relatively high requirements for data preprocessing. Finally, federated learning can run into the problem of back-attacks.

This paper mainly expounds on how artificial intelligence algorithms solve the problem of brain-computer interface and puts forward the possibilities brought by fusion models. New issues were identified during the review process and suggestions were provided.

2. Addressing key challenges in BCIs through AI techniques

2.1. Addressing data scarcity with transfer learning and generative adversarial networks

With the development of brain-computer interfaces, there are currently three main modes of brain-computer interfaces, the most common being non-invasive BCI, which mainly processes and analyzes EEG data. However, it is difficult to construct a generic model because of the sensitivity of this modality to noise and response, and because there are differences between subjects, and it is very time-consuming and laborious to collect new data for each user or topic [2]. At this point, we can take advantage of the transfer learning algorithm, which mainly moves the knowledge learned in one task or domain (source domain) to another related, but different user or domain (target domain) to improve the performance of the model on the target domain. Its application is to migrate the user data that has been trained under a specific topic to another topic through a migration algorithm, the original topic is the source domain, and the new topic is the target domain.

In addition, the hybrid model of transfer learning can solve the problem of community sparseness. For example, federated transfer learning, which combines transfer learning and federated learning, can solve the problem of insufficient data volume [3]. However, if only transfer learning is used to adapt brain-computer interfaces to different topics, it is not enough, because the source domain will also have the problem of data sparsity. And because of the high cost of EEG data collection, if the data is too sparse, there will also be the problem of overfitting. Therefore, we also need another algorithm, which is to generate an adversarial network, that augments the training set by generating pseudo-data similar to the real signal distribution, and the GAN has good robustness. In addition, GAN has demonstrated its applicability in many places, such as GAN generating SSVEP signals, improving the authenticity and depth of generated data, and improving the accuracy of learning classification [4]. The purpose of transfer learning is to transfer the knowledge of the source domain to the target domain to improve the generalization ability of the model, while the purpose of GAN is to generate synthetic data similar to the distribution of real data.

The purpose of their hybrid model is that the shortcomings of TL are compensated by GAN, and the transfer learning relies on the source domain data, and if the source domain data is insufficient, the transfer effect is poor. GANs can generate synthetic data to augment source or target domain datasets. TL is sensitive to domain differences, and the data generated by GAN can be reduced by style transfer. The shortcomings of GAN are made up by TL, and the data generated by GAN may collapse in mode, and transfer learning can extract high-level features through pre-trained models to constrain the semantic rationality of the generated data. Although the GAN-TL model can collaboratively solve the problem of data scarcity and domain adaptation, there are still the following pitfalls: first, the training instability, due to the inherent defects of GAN, the adversarial training of generators and discriminators is easy to be unbalanced, resulting in gradient vanishing or pattern collapse. The parameter initialization of transfer learning may exacerbate the shock of GAN training, and the parameters need to be finely tuned. Secondly, if the difference between the source and destination domains is too large, the negative migration effect of TL will reduce the availability of the generated data. If the data generated by GAN deviates from the true distribution, it would further mislead the transfer learning model.

2.2. Improving EEG classification using hybrid transformer-CNN models

Due to the low spatial resolution, high temporal resolution low signal-to-noise ratio, and large individual differences in EEG signals [5]. These characteristics pose great challenges to the signal processing and accurate classification of motion images and electroencephalogram (EEG) of brain-computer interface systems, and the development of artificial intelligence has also brought corresponding solutions. When it comes to classification, traditional methods such as SVM have good performance in classification tasks, but SVM has certain limitations on the performance of data classification. EEG signals often contain complex nonlinear dynamic features, and SVMs use linear functions by default, so they need to be mapped to high-dimensional spaces with dependencies and skills.

CNN and Transform can automatically extract features through hierarchical nonlinear transformation, without design and function, and SVN makes it difficult to process high-dimensional spatiotemporal data, SVN needs to rely on manual extraction of CSP features, while CNN can learn spatiotemporal features directly from the original signal, which shows that traditional SVM is not well qualified for EEG data classification tasks. In addition, the attention mechanism of RNNs has been used for EEG classification tasks [6], Transformers has been used for classification tasks, and CNNs have been used for decoding tasks of brain-computer interfaces, such as the attention-based temporal convolutional network ATCNet to decode MI-EEG brain signals [7]. First of all, the RNN algorithm is used to classify, the more common is the CNN-RNN hybrid model, which first uses CNN to extract spatial features, and then uses it to capture time dynamics, but the limitations of RNN itself make it difficult to process long EEG signals, such as ten-second EEG signal sampling, and it will have gradient problems, the most important thing is that he cannot perform parallel operations, which shows that the CNN-RNN model is more suitable for short EEG signal classification. The second is the Transformer model for classification. JING et al. constructed five types of models, among which the f-ctrans model performed well in short-term multi-classification-based data [8].

However, this shows that the transform model can be well classified in the classification of short sequence-based data. Moreover, the transform model has good classification performance in moving images and can support parallel processing and training efficiency through the attention mechanism. For example, the CNN-Transformer model, called the EEG Conformer model, can be decoded and classified to learn global time dependencies [9]. CNN extracts local spatial features based on convolutional kernels, but it is difficult to model long-time series dependence. RNNs deal with timing dependence through a cyclic structure, but it is difficult to calculate in parallel, there are still gradient problems in long sequences, and they are insensitive to sudden signal changes. Based on the self-attention mechanism to capture global dependencies, Transformer has strong parallelization ability, high training efficiency, excellent long sequence modeling ability, and good interpretability. However, a large amount of data is required, small samples are easy to overfit, and computational resources are consumed. Their hybrid model is mainly CNN-RNN and CNN-Transformer. CNNs extract EEG spatial patterns, while RNNs and Transformers capture temporal dynamics. Their hybrid model makes up for the shortcomings of a single model: the locality of CNN and the global nature of Transformers can improve the classification robustness. The sequence sensitivity of RNNs and the noise reduction ability of CNNs can improve EEG decoding with a low signal-to-noise ratio. However, they also have certain limitations: first, they are computationally complex, and the number of parameters increases due to the stacking of multiple modules. Secondly, the training is difficult, and the convergence speed of different modules is different, for example, CNN is faster than Transformer, so it is necessary to fine-tune the parameters. Finally, in the case of small data with overfitting risk, the hybrid model may be overly dependent on a certain module. The EEG-CNN-Transformer model shows good performance in various classification tasks of EEG.

2.3. Enhancing data security in BCIs via federated learning and GANs

Brain-computer interfaces generally require EEG data from multiple subjects, and these data are usually the privacy data of others, EEG data can reflect the physical and mental health status, cognition and emotion, identity and biometric privacy information, and EEG-based BCI is prone to the threat of privacy leakage. Then among the machine learning algorithms, there is an algorithm that can protect the security of data very well, even if it is federated learning. Federated learning is used to protect BCI's private data, mainly by making the data of individual clients inaccessible to the central server. The central server maintains a global model and sends it to the local client for updates. Each client updates the global model parameters based on its own EEG data and aggregates them with the sending server [10].

However, this also introduces a new problem, as an attacker can still reverse engineer the raw data. In response to this problem, two solutions are proposed. For example, the sandwich framework formed by the combination of transfer learning and federated learning can protect the four privacy attributes, the privacy parameters of the data body, the encryption of privacy in the process, and the inference level [11]. In addition, there are some limitations to people's learning, the first is the problem of data poisoning, maliciously uploading tampered data on the client, and we can use robust learning algorithms to solve this problem. For example, based on the geometric median to resist poisoning attacks, in addition to federated learning, we can also use generative adversarial networks to generate synthetic EEG data instead of real data, which can not only avoid leaking real user data but also generate adversarial samples to confuse the attack model, to prevent reverse attacks.

3. Current challenges and future directions in BCI development

3.1. Optimizing hybrid AI models

At present, several deficiencies remain in the integration of AI models with BCIs. The GAN-TL model, which combines generative adversarial networks (GANs) and transfer learning, often experiences model collapse due to inherent training instabilities of GANs and additional parameter bias introduced by transfer learning. Negative transfer effects can also occur, reducing the effectiveness of transfer learning. Additionally, the EEG-CNN-Transformer hybrid model requires robust noise reduction preprocessing to achieve optimal performance.

Several strategies can address these challenges. First, stabilizing the training mechanism is essential; this can be achieved by replacing traditional GANs with Wasserstein GANs or diffusion models to mitigate mode collapse. Second, implementing dynamic transfer strategies that adaptively adjust the weights of source and target domains via meta-learning can minimize negative transfer effects. Lastly, improving the CNN-Transformer model through depth-wise separable convolutions can significantly reduce computational costs. Combining CNN-Transformer models with GAN-generated synthetic data can further address the scarcity of large EEG datasets required for training.

3.2. Improving BCI hardware and comfort

Despite advancements in AI models, significant challenges persist in brain-computer interface hardware design. Current BCI devices are often bulky and rigid, limiting their practical daily use. Although innovative solutions such as the in-ear SpiralE device offer improvements, they still face limitations regarding comfort and long-term usability [12]. Given that BCIs are intended for human users, user comfort is paramount. Flexible and biocompatible electronic components, which are currently becoming more widely available, will be critical for ensuring comfortable and prolonged use [13].

3.3. Expanding practical applications of BCIs

3.3.1. Smart gaming and social interaction

Personal information can be securely stored within brain-computer interfaces, enabling seamless player authentication across multiple gaming platforms. Real-time monitoring of brain activity can provide immediate visual or auditory feedback, enhancing the gaming experience. Furthermore, BCIs allow users to socialize and interact intuitively with other players in immersive virtual environments, increasing realism and user engagement.

3.3.2. Brain-computer-controlled robots

Brain-computer interfaces enable the control of robots and prosthetic devices directly through conscious thought, offering profound benefits for individuals with disabilities. Unlike traditional remote-control methods, consciousness-driven interfaces allow robots to execute actions more closely aligned with human intentions. Additionally, consciousness-controlled robots can safely undertake dangerous tasks, such as deep-sea exploration, high-altitude work, or test-driving new vehicles, reducing human exposure to hazardous conditions.

3.3.3. Treatment of neurological diseases

BCIs hold significant therapeutic potential for the treatment and management of neurological disorders. By decoding brain signals and facilitating interaction with external devices, BCIs enable applications such as neurofeedback training, neural bypass technologies, and closed-loop neural regulation. These techniques offer promising approaches to restoring lost functions and managing neurological conditions.

4. Conclusion

This study systematically reviews the application of artificial intelligence algorithms to address major challenges in brain-computer interfaces. Specifically, it discusses the effectiveness of the GAN-TL model in addressing data scarcity, the CNN-Transformer model in enhancing EEG data classification accuracy, and federated learning in ensuring data security. These hybrid approaches show significant promise; however, limitations remain. This paper outlines targeted solutions to these challenges. Future research should focus on translating these theoretical solutions into practical implementations, continuously refining hybrid AI models, and enhancing their applicability in real-world scenarios.

References

[1] G. Pfurtscheller, G. R. Müller-Putz, R. Scherer and C. Neuper, "Rehabilitation with brain–computer interface systems", Computer, vol. 41, no. 10, pp. 58-65, 2008.

[2] D. Wu, Y. Xu and B. -L. Lu, "Transfer Learning for EEG-Based Brain–Computer Interfaces: A Review of Progress Made Since 2016," in IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 1, pp. 4-19, March 2022, doi: 10.1109/TCDS.2020.3007453.

[3] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu and C. Guan, "Federated Transfer Learning for EEG Signal Classification," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 3040-3045, doi: 10.1109/EMBC44109.2020.9175344.

[4] J. Wang et al., "Using Determinant Point Process in Generative Adversarial Networks for SSVEP Signals Synthesis," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023, pp. 1-4, doi: 10.1109/EMBC40787.2023.10340247.

[5] A. Craik, Y. He and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review", J. Neural Eng., vol. 16, Jun. 2019.

[6] G.Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network", IEEE Sensors J., vol. 20, no. 6, pp. 3113-3122, Mar. 2020.

[7] H. Altaheri, G. Muhammad and M. Alsulaiman, "Physics-Informed Attention Temporal Convolutional Network for EEG-Based Motor Imagery Classification," in IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 2249-2258, Feb. 2023, doi: 10.1109/TII.2022.3197419.

[8] J. Xie et al., "A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 2126-2136, 2022, doi: 10.1109/TNSRE.2022.3194600.

[9] Y. Song, Q. Zheng, B. Liu and X. Gao, "EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 710-719, 2023, doi: 10.1109/TNSRE.2022.3230250.

[10] T. Jia, L. Meng, S. Li, J. Liu and D. Wu, "Federated Motor Imagery Classification for Privacy-Preserving Brain-Computer Interfaces," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 3442-3451, 2024, doi: 10.1109/TNSRE.2024.3457504.

[11] The 'Sandwich' meta-framework for architecture agnostic deep privacy-preserving transfer learning for non-invasive brainwave decoding Xiaoxi Wei, Jyotindra Narayan and A Aldo Faisal"The 'Sandwich' meta-framework for architecture agnostic deep privacy-preserving transfer learning for non-invasive brainwave decoding"Published 23 January 2025 • © 2025 The Author(s). Published by IOP Publishing Ltd Journal of Neural Engineering, Volume 22, Number 1Citation Xiaoxi Wei et al 2025 J. Neural Eng. 22 016014DOI 10.1088/1741-2552/ad9957

[12] Wang, Z., Shi, N., Zhang, Y. et al. Conformal in-ear bioelectronics for visual and auditory brain-computer interfaces. Nat Commun 14, 4213 (2023). https://doi.org/10.1038/s41467-023-39814-6

[13] Y. Zhang, T. Zhang, Z. Huang, J. Yang, A New Class of Electronic Devices Based on Flexible Porous Substrates. Adv. Sci. 2022, 9, 2105084. https://doi.org/10.1002/advs.202105084