Fusing Pretrained ViTs with TCNet for Enhanced EEG Regression

Eric Modesitt[1], Haicheng Yin[2], Williams Huang Wang[3], and Brian Lu[4]

[1] University of Illinois at Urbana Champaign
[2] University of California, Santa Cruz
[3] Saint Francis Preparatory High School
[4] Palo Alto High School

Abstract. The task of Electroencephalogram (EEG) analysis is paramount to the development of Brain-Computer Interfaces (BCIs). However, to reach the goal of developing robust, useful BCIs depends heavily on the speed and the accuracy at which BCIs can understand neural dynamics. In response to that goal, this paper details the integration of pretrained Vision Transformers (ViTs) with Temporal Convolutional Networks (TCNet) to enhance the precision of EEG regression. The core of this approach lies in harnessing the sequential data processing strengths of ViTs along with the superior feature extraction capabilities of TCNet, to significantly improve EEG analysis accuracy. In addition, we analyze the importance of how to construct optimal patches for the attention mechanism to analyze, balancing both speed and accuracy tradeoffs. Our results showcase a substantial improvement in regression accuracy, as evidenced by the reduction of Root Mean Square Error (RMSE) from 55.4 to 51.8 on EEGEyeNet's Absolute Position Task, outperforming existing state-of-the-art models. Without sacrificing performance, we increase the speed of this model by an order of magnitude (up to 4.32x faster). This breakthrough not only sets a new benchmark in EEG regression analysis but also opens new avenues for future research in the integration of transformer architectures with specialized feature extraction methods for diverse EEG datasets.

Keywords: EEG Analysis · Temporal Convolutional Networks · Brain Computer Interfaces · Vision Transformers

1 Introduction

Analyzing Electroencephalogram (EEG) signals is fundamental to the progress of Brain-Computer Interfaces (BCIs), offering deep insights into the complex neural processes of the human brain. In the past decade, a wide range of machine learning and deep learning algorithms have been applied to EEG data, resulting in significant advancements in various applications. These applications encompass emotion recognition, motor imagery, mental workload evaluation, seizure detection, Alzheimer's disease classification, sleep stage scoring, and many others (Craik et al., 2019; Roy et al., 2019; Altaheri et al., 2023; Qu, 2022; Gao et al., 2021; Hossain et al., 2023; Yi and Qu, 2022; Key et al., 2024; Li et al., 2024; Koome Murungi et al., 2023; Dou et al., 2022; Zhou et al., 2022; Qu et al., 2020b,c,a, 2018, 2019; Murungi et al., 2023; Saeidi et al., 2021; Qu and Hickey, 2022; Rasheed et al., 2020; Wang and Qu, 2022; Dadebayev et al., 2022; Li et al., 2020; Aggarwal and Chugh, 2022). EEG regression, in particular, stands as a pivotal tool in both neuroscience and medical diagnostics, gaining prominence for its ability to decode complex neural dynamics. This technique plays a crucial role in a myriad of applications, ranging from pinpointing brain damage locations to monitoring cognitive activities and deciphering the neural basis of seizures (Subasi and Ercelebi, 2005; Sabbagh et al., 2020; Teplan, 2002). The essence of EEG regression lies in its capacity to transform raw EEG data into interpretable and meaningful information, thus providing an invaluable perspective into the brain's operations.

In the realm of machine learning, the advent of Transformer models has marked a revolutionary shift in EEG regression analysis. Initially celebrated for their breakthroughs in natural language processing, these models have been adeptly modified to cater to EEG data analysis, substantially elevating both the precision and efficiency of the analysis (Liu et al., 2022). A significant stride in this field is the adaptation of pre-trained Vision Transformers (ViTs) for EEG datasets, such as ImageNet Deng et al. (2009). The application of ViTs in EEG regression has demonstrated exceptional results, surpassing traditional methods across various benchmarks (Yang and Modesitt, 2023).

Concurrently, Temporal Convolutional Networks (TCNet) have emerged as a formidable force in the field of EEG signal processing. Exhibiting outstanding capabilities in feature extraction, TCNets excel in identifying intricate patterns and nuances in EEG data (Bai et al., 2018; Ingolfsson et al., 2020; Altaheri et al., 2022). Their robustness in capturing temporal dynamics and their efficacy in EEG signal handling render them an indispensable component in neural signal analysis.

This study delves into the synergistic integration of ViTs and TCNet, aiming to harness their combined strengths to substantially augment the accuracy and reliability of EEG regression. This innovative approach seeks to leverage the detailed feature extraction of TCNet and the contextual interpretation prowess of ViTs, hypothesizing a significant enhancement in EEG analysis.

Our research presents a comprehensive evaluation of this hybrid model, juxtaposing it against previous methodologies to underscore its superiority in EEG regression. We meticulously examine the performance of the ViT-TCNet combination, elucidating the contribution of each component to the overall effectiveness of the model. The implications of our findings extend beyond the confines of EEG analysis, potentially influencing a broad spectrum of data interpretation tasks in various scientific and AI-related fields.

In addition to the aforementioned aspects, a notable facet of this study is the emphasis on the processing speed of the integrated ViT-TCNet model. Speed is a critical parameter in EEG analysis, especially for real-time applications in Brain-Computer Interfaces (BCIs) where rapid response times are essential By optimizing the architecture and employing advanced techniques in model training and inference, we have successfully accelerated the processing speed of the EEG analysis. This advancement is particularly significant in scenarios where real-time data processing is crucial, such as in neurofeedback systems or in clinical settings where prompt decision-making is imperative. The increase in processing speed, achieved without compromising the model's performance, marks a substantial leap forward in making EEG-based BCIs more viable and user-friendly.

In the ensuing sections, we will outline the methodology utilized in our study, present our empirical findings, and discuss the broader implications and future research directions stemming from our work. This research not only enriches the existing literature in EEG regression but also sets the stage for future explorations into the amalgamation of advanced machine learning architectures for refined neural data analysis.

In summary, the contributions of this work can be articulated in three primary areas:

1. **Innovative Combination of ViTs and TCNet for Advanced EEG Regression**: This research marks a significant advancement in EEG regression analysis through the novel integration of pretrained Vision Transformers (ViTs) with Temporal Convolutional Networks (TCNet). This fusion harnesses ViTs' exceptional capability in processing sequential data and TCNet's robust feature extraction techniques, culminating in a notable improvement in EEG regression accuracy.

2. **Enhancement of Model Processing Speed and Efficiency**: A key contribution of this study is the substantial improvement in the processing speed of the EEG analysis model. Recognizing the importance of swift data processing in real-time applications such as Brain-Computer Interfaces, the research introduces optimizations that significantly accelerate the model's performance.

3. **Ablation Studies and Future Research Directions**: The research undertakes comprehensive ablation studies to understand the individual and combined contributions of ViTs and TCNet to the model's performance. These studies offer valuable insights into the mechanics of the model, paving the way for further optimizations.

2 Related Work

2.1 Deep Learning in EEG

The evolution of EEG signal processing has been significantly influenced by the emergence of deep learning techniques. Traditional machine learning methods, while effective, often fall short in capturing the high-dimensional and complex nature of EEG data. The introduction of deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), revolutionized this field. These models brought enhanced capabilities in handling large datasets, extracting relevant features, and recognizing intricate patterns in EEG signals (Subasi and Ercelebi, 2005; Sabbagh et al., 2020; Teplan, 2002). This shift not only improved the accuracy of EEG analyses but also expanded the potential applications in neurological research and clinical diagnostics.

Deep learning's impact on EEG signal processing is profound, offering new perspectives in understanding brain activity. The ability of these models to learn from data autonomously, without the need for extensive feature engineering, has opened avenues for more nuanced and detailed analyses of neural signals. This advancement is crucial in fields where EEG data plays a pivotal role, such as in the study of cognitive processes, sleep patterns, and brain-computer interfaces. The integration of advanced deep learning architectures in EEG analysis heralds a new era of innovation and discovery in neuroscience.

2.2 ViTs in Non-Image Data Analysis

Vision Transformers (ViTs), originally designed for image recognition, have demonstrated remarkable versatility by extending their application to various other domains, including EEG data analysis (Dosovitskiy et al., 2020; Wu et al., 2020; Han et al., 2022). The cornerstone of their success, the self-attention mechanism, enables ViTs to efficiently manage sequential data, a feature crucial in interpreting EEG signals Vaswani et al. (2017). This characteristic of ViTs facilitates an understanding of the complex, temporal relationships inherent in EEG data, making them an ideal choice for this type of analysis.

The adaptation of ViTs to non-image data, such as EEG signals, signifies a major shift in the approach to data analysis across disciplines. It underscores the potential of transformer models to handle diverse types of data beyond their initial scope. This cross-domain applicability of ViTs not only enriches the toolkit available for EEG analysis but also inspires innovative approaches to data interpretation. The flexibility and effectiveness of ViTs in handling sequential data pave the way for their broader adoption in various scientific and analytical fields.

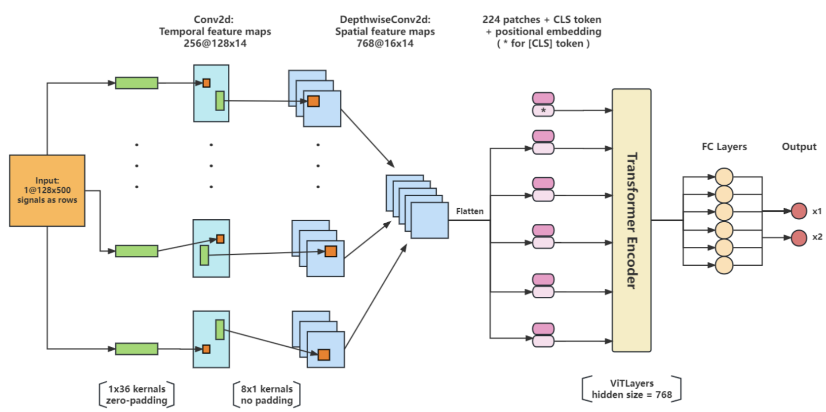

Fig. 1. *EEGViT architecture, SOTA on EEGEYENET* (Yang and Modesitt, 2023).

2.3 Temporal Convolutional Networks (TCNet)

:

Temporal Convolutional Networks (TCNet) have gained significant attention for their ability to process time-series data, particularly in EEG signal analysis. The architecture of TCNet, with its focus on capturing temporal dependencies through convolutional layers, makes it exceptionally suited for extracting detailed features from EEG data (Farha and Gall, 2019; Hewage et al., 2020). The efficacy of TCNet in identifying subtle patterns and temporal features in complex datasets has established it as a leading tool in the field of neural signal processing.

The role of TCNet in EEG data interpretation extends beyond mere feature extraction. It involves a deeper understanding of the temporal dynamics and inherent structures within the EEG signals. This understanding is vital in applications where precise timing and sequence of neural events are critical, such as in epilepsy research or brain-computer interface development. The combination of TCNet with other advanced models like ViTs presents a promising avenue for enhancing EEG analysis, potentially leading to more accurate and insightful interpretations of neural data.

3 Methods

Our study employs an innovative approach by integrating pre-trained Vision Transformers (ViTs) with Temporal Convolutional Networks (TCNet) to enhance EEG regression analysis. This section outlines the dataset utilized, the specifics of the proposed model, and the methodology for evaluating its effectiveness.

3.1 EEGEyeNet Dataset

The data presented here is derived from the EEGEyeNet dataset Kastrati et al. (2021). The EEGEyeNet dataset encompasses recordings from 356 healthy adults, including 190 females and 166 males, aged 18 to 80 years. All individuals in this study provided written informed consent, compliant with the Declaration of Helsinki, and were compensated monetarily.

The EEG recordings in the EEGEyeNet dataset were obtained using a high-density 128-channel EEG Geodesic Hydrocel system, operating at a sampling rate of 500 Hz with a central recording reference. Eye positions were concurrently recorded using an EyeLink 1000 Plus system at the same sampling rate. This setup maintained electrode impedances below 40 kOhm and ensured accurate eye tracker calibration. Participants were positioned 68 cm from a 24-inch monitor, with their head stabilized using a chin rest.

EEG data, as recorded in the EEGEyeNet dataset, are prone to various artifacts, including environmental noise and physiological interferences such as eye movements and blinks. To address this, the dataset underwent rigorous preprocessing with two levels: minimal and maximal. The minimal preprocessing involved identifying and interpolating faulty electrodes, along with applying a high-pass filter at 40 Hz and a low-pass filter at 0.5 Hz. The maximal preprocessing, aimed at neuroscientific analyses, further incorporated Independent Component Analysis (ICA) and IClabel for artifact component removal.

The EEGEyeNet dataset also includes synchronized EEG and eye-tracking data, facilitating time-locked analyses relative to event onsets. This synchronization was stringently verified to ensure a maximum error margin of 2 ms.

The Absolute Position Task, a key component of the EEGEyeNet dataset, involved participants fixating on sequentially displayed dots at various screen positions. Each dot appeared for 1.5 to 1.8 seconds, located at one of 25 distinct screen positions. The central dot was presented thrice, resulting in 27 trials per block. This setup, covering the entire screen area, captured a broad range of gaze positions. Adapted from Son et al. (2020) for fMRI studies, modifications were made for EEG compatibility, including stimulus duration and repetition adjustments. The dot presentation followed a pseudo-randomized sequence across five experimental blocks, repeated six times, totaling 810 stimuli per participant.

The Absolute Position task is particularly relevant for our research as it provides a comprehensive dataset for analyzing eye movement patterns and gaze positions. The variety in dot positions and the high number of trials allow for a complete assessment of the participants' gaze behavior, which is crucial for our objective of determining the exact XY-coordinates of a participant's gaze using EEG data.

3.2 EEGViT-TCNet Model Architecture

Intending to advance EEG signal analysis, we developed the EEGViT-TCNet model, a novel architecture that combines Temporal Convolutional Networks (TCNet) with a pre-trained Vision Transformer (ViT). This model was meticulously designed to decipher the temporal dynamics and spatial characteristics embedded within EEG signals.

Temporal Convolutional Network (TCNet) Component: The EEGViT-TCNet model begins with the TCNet component, tailored to embrace the complexities of EEG data. This component is characterized by:

- **Input Layer:** Accepting EEG signals, the TCNet is prepared to handle an input dimensionality of 129, corresponding to the number of recorded EEG channels +1 for including grounding information (as done in the original EEGEyeNet paper).
- **Sequential TCNet Layers:** The architecture encompasses three layers, with the number of channel dimensions expanding progressively to 64, 128, and 256. This hierarchy is instrumental in capturing a comprehensive spectrum of temporal dependencies inherent in the EEG signals.
  - *Kernel Size:* A kernel size of 3 is uniformly applied across the TCNet layers.
  - *Dropout:* To counteract the potential for overfitting, a dropout rate of 0.75 is employed.

- *Causality and Normalization:* The layers incorporate weight normalization alongside the ReLU activation function to enhance the model's stability and performance.

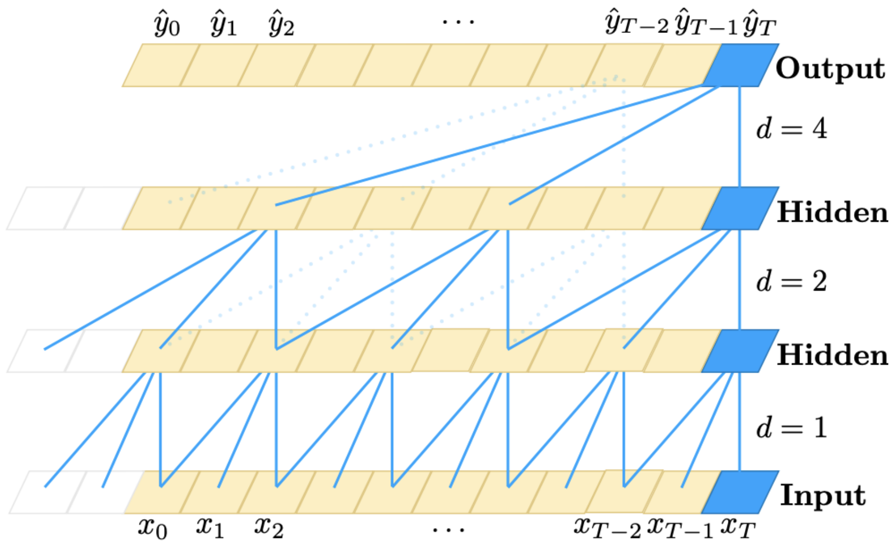

Fig. 2. *An outline of the TCNet functionality* (Bai et al., 2018).

Convolutional and Batch Normalization Layers: The pathway pathway we designed from the TCNet layers to the ViT involves:

– **Convolutional Layers:** Two convolutional layers connect the TCNet and ViT. The first layer, equipped with 256 filters and a kernel size of (1, 36), is succeeded by batch normalization and ReLU activation. This configuration initiates the spatial feature extraction. Subsequently, the second layer amplifies the channel dimension to 768, aligning with the ViT's input specifications. In addition, this layer employs a kernel size of (256, 1) to effectuate a spatial compression conducive to the subsequent transformer analysis.

Vision Transformer (ViT) Component: The culmination of the EEGViT-TCNet model's preprocessing lies in the ViT component:

– **EEG Data Adaptation:** Leveraging the "google/vit-base-patch16-224" model from Huggingface, the configuration is modified to fit the unique format of EEG data.

- **Patch Embeddings Projection:** This layer is reimagined as a 1D convolutional layer, directly accommodating the output from preceding stages BS ensuring the input's integration into the transformer architecture.
- **Classifier Head:** The model ends with a randomly initialized classifier layer, transitioning through a linear layer, followed by a dropout layer (p=0.1), culminating in a linear layer that predicts the gaze XY coordinates.

3.3 Training and Evaluation Procedure

For training, we employed a supervised learning approach. The model was trained on a split of the EEG dataset, with 70% used for training and 30% for validation. During training, we employed a mean squared error loss function, optimized using the Adam optimizer with a learning rate of 1e-4. To prevent overfitting, we implemented early stopping based on the validation loss, with a patience of 10 epochs.

The primary metric for evaluating our model's performance was the Root Mean Square Error (RMSE). This metric provides a clear indication of the model's accuracy in predicting the gaze coordinates. A lower RMSE value indicates a closer approximation to the actual gaze positions. To ensure the robustness of our findings, we conducted five independent runs for each model configuration and reported the mean and standard deviation of these runs.

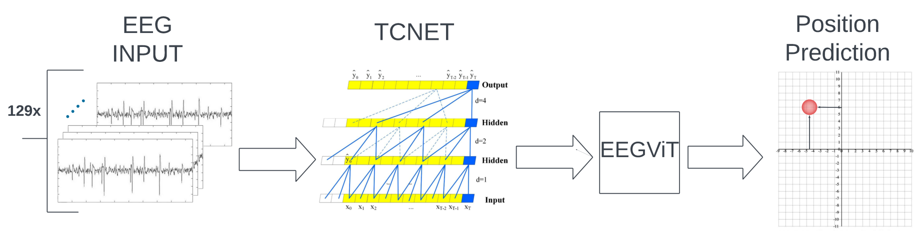

Fig. 3. *An outline of our addition to EEGViT, demonstrating our distinct feature extraction methodology.*

4 Results

Through rigorous testing and comparison, our model has demonstrated its capability to predict gaze positions with state-of-the-art precision, outperforming a spectrum of both conventional and advanced methodologies.

4.1 Performance Benchmarking

Our EEGViT-TCNet model achieved a Root Mean Square Error (RMSE) of 51.8mm, marking a significant advancement over existing models. This performance showcases a 6.5% enhancement in precision over standalone ViT models

Yang and Modesitt (2023). The stark contrast is further accentuated when juxtaposed with traditional approaches such as Linear Regression and Random Forest, where the RMSE figures exceed 115mm.

Our model's evaluation, focusing on a single data partition divided by subject as done in Kastrati et al. (2021), demonstrates its ability to generalize effectively. This approach, where the testing data consisted of completely unseen, new groups of data, highlights the model's robustness and adaptability. Despite being tested on different subsets of subjects than what the model was trained on, it maintained consistent RMSE metrics, underscoring its sophisticated architecture's capability to handle the complexities of EEG data. This consistency in performance across various subject-based segments is crucial for real-world applications, affirming the model's potential for effective generalization.

Model	**Absolute Position RMSE (mm)**
Naive Guessing	123.3 ± 0.0
KNN	119.7 ± 0
RBF SVR	123 ± 0
Linear Regression	118.3 ± 0
Random Forest	116.7 ± 0.1
CNN	70.4 ± 1.1
EEGViT (Pre-trained)	55.4 ± 0.2
EEGViT-TCNet	**51.8 ± 0.6**

Table 1. Comparative analysis of Root Mean Squared Error (RMSE) across various models, highlighting the superior performance of the EEGViT-TCNet model. The values represent the mean ± standard deviation over five independent runs, illustrating the model's consistency and accuracy in the Absolute Position Task.

4.2 Ablation Studies

To assess the individual contributions of various components within our EEGViT-TCNet model, we conducted a series of ablation studies. These studies aimed to isolate the effects of specific elements of the model, such as the convolutional layers, dropout rates in the TCNet, and the use of a pretrained Vision Transformer (ViT). Each variation of the model was evaluated using the same dataset and metrics, allowing us to directly compare their performance.

Impact of Convolutional Layers We first examined the impact of the additional convolutional layers that bridge the gap between the TCNet and the

ViT on the model's performance. In particular, we analyze the results after removing all possible combinations of 1 convolutional layer (spatial, temporal, and pointwise convolution). By removing the pointwise layer, we observed a slight increase in the Root Mean Square Error (RMSE) from 51.8 ± 0.6mm to 52.5 ± 0.8mm. This suggests that the pointwise convolutional layer plays a modest yet significant role in feature extraction and spatial representation, contributing to the model's overall accuracy.

Influence of Dropout Rates in TCNet The role of dropout rates in TCNet was another focus of our study. By varying the dropout rates, we investigated their effect on the model's capability to generalize and prevent overfitting. The original model with a 0.75 dropout rate achieved an RMSE of 51.8 ± 0.6mm. Reducing the dropout rate to 0 increased the RMSE to 54.1 ± 0.6mm, indicating a higher propensity for overfitting. Conversely, lower dropout rates of 0.25 and 0.5 yielded RMSEs of 52.5 ± 0.4mm and 52.1 ± 0.4mm, respectively. These findings illustrate a nuanced balance between dropout rate and model performance, with moderate dropout rates contributing positively to the model's accuracy and generalizability.

Contribution of Pretrained ViT Finally, we evaluated the contribution of using a pretrained ViT in our model. By replacing the pretrained ViT with a non-pretrained counterpart, the RMSE increased to 53.2 ± 0.5mm. This increase underscores the significance of pretraining in enhancing the model's feature recognition capabilities, particularly in the context of EEG data analysis.

Model Variation	**Absolute Position RMSE (mm)**
EEGViT-TCNet (Ours)	51.8 ± 0.6
No 2nd Conv Layer	52.5 ± 0.8
0% Dropout	54.1 ± 0.6
25% Dropout	52.5 ± 0.4
50% Dropout	52.1 ± 0.4
No Pretrained ViT	53.2 ± 0.5

Table 2. Ablation study results comparing RMSE across various EEGViT-TCNet model configurations. The values represent the mean ± standard deviation over five runs.

These ablation studies reveal the delicate interplay of different architectural components in optimizing the EEGViT-TCNet model for EEG regression analysis. The presence of the second convolutional layer, the calibration of dropout rates in TCNet, and the incorporation of a pretrained ViT each contribute uniquely to the model's performance. Our findings highlight the importance of these components in achieving high precision in EEG regression tasks, providing valuable insights for future enhancements and applications of the model.

5 Discussion

The integration of Vision Transformers (ViTs) with Temporal Convolutional Networks (TCNet) within the EEGViT-TCNet model represents a substantial leap forward in EEG signal analysis, particularly for applications within brain-computer interfaces (BCIs) and the broader realm of neural signal processing. This novel approach has not only showcased a marked improvement in regression accuracy but also set new precedents in processing speed and efficiency. The results obtained from this study reflect the significant potential of leveraging the strengths of both ViTs and TCNet, underscoring the profound impact that such hybrid models can have on understanding and interpreting complex neural dynamics.

The success of the EEGViT-TCNet model in reducing the Root Mean Square Error (RMSE) to unprecedented levels emphasizes the model's capability to provide a more accurate interpretation of EEG data. This breakthrough is particularly relevant in the development of BCIs, where the precision of signal interpretation directly correlates to the effectiveness and user-friendliness of the interface. In clinical settings, the enhanced accuracy and speed of EEG analysis facilitated by the EEGViT-TCNet model could lead to more timely and accurate diagnoses of neurological conditions, potentially transforming patient care.

Throughout the research, adapting ViT to the unique nature of EEG data highlighted the complexities inherent in neural signal processing. The preprocessing of EEG signals, essential for maintaining the integrity of temporal features, posed significant challenges. This process is critical in ensuring the model's adaptability and generalizability across different subjects and experimental conditions, a vital aspect for the practical application of such technologies.

Looking forward, the field beckons for further exploration into the scalability of hybrid models like EEGViT-TCNet, particularly in handling larger datasets and assessing performance in diverse real-world scenarios. A key area of interest lies in enhancing the interpretability of these deep learning models. Improved interpretability is crucial for clinical acceptance and can lead to advancements in personalized medicine, where EEG analysis can be tailored to individual patients for monitoring or therapeutic purposes.

Moreover, the exploration into other hybrid architectures and their efficacy across various domains of neural data presents an exciting avenue for research. The integration of multimodal data sources, alongside the application of transfer learning techniques, could further refine the accuracy and applicability of EEG signal analysis methods. Such advancements could pave the way for the development of more sophisticated BCIs, offering improved interaction mechanisms between humans and machines.

6 Conclusion

This study represents a significant advancement in EEG regression analysis, underscoring the indispensable role of meticulous feature extraction in the efficacy of sophisticated computational models like Vision Transformers (ViT). The integration of Temporal Convolutional Networks (TCNet) with pretrained ViTs has unveiled the vast potential of harmonizing specialized feature extraction techniques with advanced deep learning frameworks. This synergy not only elevates the accuracy of EEG analysis but also establishes a new benchmark in the field, showcasing the profound benefits of refined feature representation.

Our findings highlight the criticality of nuanced feature extraction in interpreting complex EEG data, with the EEGViT-TCNet model demonstrating notable performance enhancements. This indicates that features, often overlooked by traditional models, can be captured and leveraged for more accurate regression analysis, suggesting a broad array of applications, from clinical diagnostics to enhanced brain-computer interfaces.

As we chart the course for future research, the horizon of EEG analysis and deep learning promises continued expansion and innovation. The development of increasingly sophisticated models capable of navigating the complexities inherent in EEG data is anticipated. A pivotal challenge will be enhancing the interpretability of these models, ensuring they not only perform optimally but also offer actionable insights for practitioners. Moreover, integrating these advanced models into real-world applications will be crucial, extending the benefits of this research to society at large. Additionally, exploring various deep learning techniques on different datasets for comparative studies (An et al. (2023a,b); Jiang et al. (2023); Gui et al. (2024); Lu et al. (2023); Chen et al. (2024); Ma (2022); Ma et al. (2024); Tan et al. (2023, 2021); Qiu et al. (2023); Zhao et al. (2024); Zhang et al. (2022, 2023)) could provide valuable insights and further enhance the field.

In sum, the fusion of TCNet and pretrained ViTs within the EEG regression domain exemplifies the transformative power of targeted feature extraction and advanced data processing. This study not only redefines the standards for EEG analysis but also lights the way for future endeavors in the realms of deep learning and neural data interpretation. As we delve deeper into the complexities of the human brain, the significance of innovative computational models grows ever more evident, harboring the potential for groundbreaking discoveries in neuroscience and artificial intelligence.

Bibliography

Swati Aggarwal and Nupur Chugh. Review of machine learning techniques for eeg based brain computer interface. *Archives of Computational Methods in Engineering*, 29(5):3001–3020, 2022.

Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention temporal convolutional network for eeg-based motor imagery classification. *IEEE Transactions on Industrial Informatics*, 19(2):2249–2258, 2022.

Hamdi Altaheri, Ghulam Muhammad, Mansour Alsulaiman, Syed Umar Amin, Ghadir Ali Altuwaijri, Wadood Abdul, Mohamed A Bencherif, and Mohammed Faisal. Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review. *Neural Computing and Applications*, 35(20):14681–14722, 2023.

Sizhe An, Ganapati Bhat, Suat Gumussoy, and Umit Ogras. Transfer learning for human activity recognition using representational analysis of neural networks. *ACM Transactions on Computing for Healthcare*, 4(1):1–21, 2023a.

Sizhe An, Yigit Tuncel, Toygun Basaklar, and Umit Y Ogras. A survey of embedded machine learning for smart and sustainable healthcare applications. In *Embedded Machine Learning for Cyber-Physical, IoT, and Edge Computing: Use Cases and Emerging Challenges*, pages 127–150. Springer, 2023b.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Jintai Chen, Yaojun Hu, Yue Wang, Yingzhou Lu, Xu Cao, Miao Lin, Hongxia Xu, Jian Wu, Cao Xiao, Jimeng Sun, et al. Trialbench: Multi-modal artificial intelligence-ready clinical trial datasets. *arXiv preprint arXiv:2407.00631*, 2024.

Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.

Didar Dadebayev, Wei Wei Goh, and Ee Xion Tan. Eeg-based emotion recognition: Review of commercial eeg devices and machine learning techniques. *Journal of King Saud University-Computer and Information Sciences*, 34(7):4385–4401, 2022.

Jia Deng et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009.

Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Guangyao Dou, Zheng Zhou, and Xiaodong Qu. Time majority voting, a pc-based eeg classifier for non-expert users. In *International Conference on Human-Computer Interaction*, pages 415–428. Springer, 2022.

Yazan Abu Farha and Jurgen Gall. Ms-tcn: Multi-stage temporal convolutional network for action segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.

Zhongke Gao, Weidong Dang, Xinmin Wang, Xiaolin Hong, Linhua Hou, Kai Ma, and Matjaž Perc. Complex networks and deep learning for eeg signal analysis. *Cognitive Neurodynamics*, 15(3):369–388, 2021.

Shengxi Gui, Shuang Song, Rongjun Qin, and Yang Tang. Remote sensing object detection in the deep learning era—a review. *Remote Sensing*, 16(2):327, 2024.

Kai Han et al. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):87–110, 2022.

Pradeep Hewage et al. Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing*, 24:16453–16482, 2020.

Khondoker Murad Hossain, Md Ariful Islam, Shahera Hossain, Anton Nijholt, and Md Atiqur Rahman Ahad. Status of deep learning for eeg-based brain–computer interface applications. *Frontiers in computational neuroscience*, 16: 1006763, 2023.

Thorir Mar Ingolfsson et al. Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020.

Chao Jiang, Bo Hui, Bohan Liu, and Da Yan. Successfully applying lottery ticket hypothesis to diffusion model. *arXiv preprint arXiv:2310.18823*, 2023.

Ard Kastrati et al. Eegeyenet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. *arXiv preprint arXiv:2111.05100*, 2021.

Matthew L Key, Tural Mehtiyev, and Xiaodong Qu. Advancing eeg-based gaze prediction using depthwise separable convolution and enhanced pre-processing. In *International Conference on Human-Computer Interaction*, pages 3–17. Springer, 2024.

Nathan Koome Murungi, Michael Vinh Pham, Xufeng Dai, and Xiaodong Qu. Trends in machine learning and electroencephalogram (eeg): A review for undergraduate researchers. *arXiv e-prints*, pages arXiv–2307, 2023.

Gen Li, Chang Ha Lee, Jason J Jung, Young Chul Youn, and David Camacho. Deep learning for eeg data analytics: A survey. *Concurrency and Computation: Practice and Experience*, 32(18):e5199, 2020.

Weigeng Li, Neng Zhou, and Xiaodong Qu. Enhancing eye-tracking performance through multi-task learning transformer. In *International Conference on Human-Computer Interaction*, pages 31–46. Springer, 2024.

Jiyao Liu et al. Spatial-temporal transformers for eeg emotion recognition. In *Proceedings of the 6th International Conference on Advances in Artificial Intelligence*, 2022.

Yingzhou Lu, Minjie Shen, Huazheng Wang, Xiao Wang, Capucine van Rechem, and Wenqi Wei. Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062*, 2023.

Xiaobo Ma. *Traffic performance evaluation using statistical and machine learning methods*. PhD thesis, The University of Arizona, 2022.

Xiaobo Ma, Abolfazl Karimpour, and Yao-Jan Wu. Data-driven transfer learning framework for estimating on-ramp and off-ramp traffic flows. *Journal of Intelligent Transportation Systems*, pages 1–14, 2024.

Nathan Koome Murungi, Michael Vinh Pham, Xufeng Caesar Dai, and Xiaodong Qu. Empowering computer science students in electroencephalography (eeg) analysis: A review of machine learning algorithms for eeg datasets. *SIGKDD*, 2023.

Yansheng Qiu, Ziyuan Zhao, Hongdou Yao, Delin Chen, and Zheng Wang. Modal-aware visual prompting for incomplete multi-modal brain tumor segmentation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3228–3239, 2023.

Xiaodong Qu. *Time Continuity Voting for Electroencephalography (EEG) Classification*. PhD thesis, Brandeis University, 2022.

Xiaodong Qu and Timothy J Hickey. Eeg4home: A human-in-the-loop machine learning model for eeg-based bci. In *Augmented Cognition: 16th International Conference, AC 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*, pages 162–172. Springer, 2022.

Xiaodong Qu, Yixin Sun, Robert Sekuler, and Timothy Hickey. Eeg markers of stem learning. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE, 2018.

Xiaodong Qu, Mercedes Hall, Yile Sun, Robert Sekuler, and Timothy J Hickey. A personalized reading coach using wearable eeg sensors. *CSEDU*, 2019.

Xiaodong Qu, Peiyan Liu, Zhaonan Li, and Timothy Hickey. Multi-class time continuity voting for eeg classification. In *Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2*, pages 24–33. Springer, 2020a.

Xiaodong Qu, Saran Liukasemsarn, Jingxuan Tu, Amy Higgins, Timothy J Hickey, and Mei-Hua Hall. Identifying clinically and functionally distinct groups among healthy controls and first episode psychosis patients by clustering on eeg patterns. *Frontiers in psychiatry*, 11:541659, 2020b.

Xiaodong Qu, Qingtian Mei, Peiyan Liu, and Timothy Hickey. Using eeg to distinguish between writing and typing for the same cognitive task. In *Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2*, pages 66–74. Springer, 2020c.

Khansa Rasheed, Adnan Qayyum, Junaid Qadir, Shobi Sivathamboo, Patrick Kwan, Levin Kuhlmann, Terence O'Brien, and Adeel Razi. Machine learning for predicting epileptic seizures using eeg signals: A review. *IEEE reviews in biomedical engineering*, 14:139–155, 2020.

Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5): 051001, 2019.

David Sabbagh et al. Predictive regression modeling with meg/eeg: from source power to signals and cognitive states. *NeuroImage*, 222:116893, 2020.

Maham Saeidi, Waldemar Karwowski, Farzad V Farahani, Krzysztof Fiok, Redha Taiar, Peter A Hancock, and Awad Al-Juaid. Neural decoding of eeg signals with machine learning: a systematic review. *Brain Sciences*, 11(11):1525, 2021.

Jake Son, Lei Ai, Ryan Lim, Ting Xu, Stanley Colcombe, Alexandre Rosa Franco, Jessica Cloud, Stephen LaConte, Jonathan Lisinski, Arno Klein, et al. Evaluating fmri-based estimation of eye gaze during naturalistic viewing. *Cerebral Cortex*, 30(3):1171–1184, 2020.

Abdulhamit Subasi and Ergun Ercelebi. Classification of eeg signals using neural network and logistic regression. *Computer methods and programs in biomedicine*, 78(2):87–99, 2005.

Jieyuan Tan, Xiang Shen, Xiang Zhang, and Yiwen Wang. Multivariate encoding analysis of medial prefrontal cortex cortical activity during task learning. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 6699–6702. IEEE, 2021.

Jieyuan Tan, Xiang Zhang, Shenghui Wu, Zhiwei Song, Shuhang Chen, Yifan Huang, and Yiwen Wang. Audio-induced medial prefrontal cortical dynamics enhances coadaptive learning in brain–machine interfaces. *Journal of Neural Engineering*, 20(5):056035, 2023.

Michal Teplan. Fundamentals of eeg measurement. In *Measurement Science Review*, volume 2, 2002.

Ashish Vaswani et al. Attention is all you need. In *Advances in neural information processing systems*, volume 30, 2017.

Ruyang Wang and Xiaodong Qu. Eeg daydreaming, a machine learning approach to detect daydreaming activities. In *International Conference on Human-Computer Interaction*, pages 202–212. Springer, 2022.

Bichen Wu et al. Visual transformers: Token-based image representation and processing for computer vision. *arXiv preprint arXiv:2006.03677*, 2020.

Ruiqi Yang and Eric Modesitt. Vit2eeg: Leveraging hybrid pretrained vision transformers for eeg data. *arXiv preprint arXiv:2308.00454*, 2023.

Long Yi and Xiaodong Qu. Attention-based cnn capturing eeg recording's average voltage and local change. In *Artificial Intelligence in HCI: 3rd International Conference, AI-HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*, pages 448–459. Springer, 2022.

Zhengming Zhang, Renran Tian, Rini Sherony, Joshua Domeyer, and Zhengming Ding. Attention-based interrelation modeling for explainable automated driving. *IEEE Transactions on Intelligent Vehicles*, 8(2):1564–1573, 2022.

Zhengming Zhang, Renran Tian, and Zhengming Ding. Trep: Transformer-based evidential prediction for pedestrian intention with uncertainty. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3534–3542, 2023.

Shenghao Zhao, Xulei Yang, Zeng Zeng, Peisheng Qian, Ziyuan Zhao, Lingyun Dai, Nayana Prabhu, Pär Nordlund, and Wai Leong Tam. Deep learning based cetsa feature prediction cross multiple cell lines with latent space representation. *Scientific Reports*, 14(1):1878, 2024.

Zheng Zhou, Guangyao Dou, and Xiaodong Qu. Brainactivity1: A framework of eeg data collection and machine learning analysis for college students. In *International Conference on Human-Computer Interaction*, pages 119–127. Springer, 2022.