EEGMobile: Enhancing Speed and Accuracy in EEG-Based Gaze Prediction with Advanced Mobile Architectures

Teng Liang[1][0009-0009-3074-2420] and Andrews Damoah[2][0009-0006-2840-6063]

[1] Palo Alto High School
tl40355@pausd.us
[2] University of Maryland - College Park
adamoah@terpmail.umd.edu

Abstract. Electroencephalography (EEG) analysis is an important domain in the realm of Brain-Computer Interface (BCI) research. To ensure BCI devices are capable of providing practical applications in the real world, brain signal processing techniques must be fast, accurate, and resource-conscious to deliver low-latency neural analytics. This study presents a model that leverages a pre-trained MobileViT alongside Knowledge Distillation (KD) for EEG regression tasks. Our results showcase that this model is capable of performing at a level comparable (only 3% lower) to the previous State-Of-The-Art (SOTA) on the EEGEyeNet Absolute Position Task while being 33% faster and 60% smaller. Our research presents a cost-effective model applicable to resource-constrained devices and contributes to expanding future research on lightweight, mobile-friendly models for EEG regression.

Keywords: Electroencephalography · Deep Learning · Brain-Computer Interfaces · Mobile Networks · Knowledge Distillation · Gaze Prediction · Human Computer Interaction

1. Introduction

Electroencephalography (EEG) signal analysis is a pivotal research subject contributing to the advancement of Brain-Computer Interfaces (BCI). Furthermore, the application of Machine Learning (ML) and Deep Learning (DL) algorithms has become a key component for EEG analysis, which has only grown steadily across the years (Sun and Mou, 2023). The EEGEyeNet dataset has become the centerpiece in this field fusing advanced EEG data compilation with cutting-edge ML and DL techniques. (Murungi et al., 2023a; Dou et al., 2022; Wolf et al., 2022; Rolff et al., 2022; Kastrati et al., 2023; Farago et al., 2022). Its popularity is underscored by its large collection of high-quality EEG and eye-tracking (ET) recordings alongside baseline ML and DL models to benchmark accuracy on a variety of ET tasks. The advent of EEGEyeNet has spawned further research into applications of new DL algorithms for EEG-ET tasks Yang and Modesitt (2023); Modesitt et al. (2024); Dou et al. (2022); Rolff et al. (2022). While much of this research is focused on improving predictive accuracy, for these methods to be practical in real-world scenarios, there is a need to develop more computationally efficient models.

As such, this study explores the efficacy of integrating a pre-trained MobileViT network into the existing EEG-based transformer model (EEGViT-TCNet) alongside incorporating knowledge distillation during training with a fine-tuned EEGViT-TCNet teacher model for EEG-based gaze estimation. This unique approach allows our model, dubbed EEGMobile, to leverage the MobileViT's effective design to increase computational efficiency during inferencing. The use of knowledge distillation also enhances task performance, allowing our model to exhibit vastly superior accuracy when compared to previous iterations where knowledge distillation was not used.

This research provides an extensive evaluation of EEGMobile and contrasts it with previous models focused on the same task, both in terms of speed and accuracy. We aim to highlight the strengths of the individual components that comprise our model as a whole and illustrate how these components unite to form a cost-effective model for EEG regression. Our research has the potential to expand the efficiency and scalability of practical applications of DL models for EEG analysis across a plethora of domains. Our code is publically available at: https://github.com/t0nyliang/EEGMobile.

1.1 Research Question

– Is a MobileViT-based model viable for faster, SOTA-comparable accuracy on the EEGEyeNet dataset?

By answering this question, we hope to contribute to research into the development of more memory-conscious models designed for resource-constrained devices, such as mobile phones. Our research contributions highlight the potential for expanding the accessibility of EEG-based eye-tracking technology to a wider audience, further interpolating neuroscience and Human-Computer Interaction (HCI) for medical applications.

Table 1. *Abbreviation Table*

Abbreviation	Definition
EEG	Electroencephalography
ET	Eye-Tracking
BCI	Brain-Computer Interfaces
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
ViT	Vision Transformer
CNN	Convolutional Neural Network
TCN	Temporal Convolutional Network
KD	Knowledge Distillation
SOTA	State Of The Art
RMSE	Root Mean Squared Error

2. Related Work

Table 2. *Root Mean Squared Error (RMSE) of EEGEyeNet Random guessing and baseline DL models, EEGViT, and EEGViT-TCN (Kastrati et al., 2021b; Yang and Modesitt, 2023)*

Model	**Absolute Position RMSE (mm)**
Naive Guessing	123.3 ± 0.0
CNN	70.4 ± 1.1
PyramidalCNN	73.9 ± 1.9
EEGNet	81.3 ± 1.0
InceptionTime	70.7 ± 0.8
Xception	78.7 ± 1.6
ViT-Base	58.1 ± 0.6
ViT-Base	61.5 ± 0.6
ViT-Base (Pre-trained)	58.1 ± 0.6
EEGViT	61.7 ± 0.6
EEGViT (Pre-trained)	55.4 ± 0.2
EEGViT-TCN	51.8 ± 0.2

2.1 EEG and Deep Learning

Electroencephalography (EEG) is extensively used in various research domains, including neural engineering, neuroscience, biomedical engineering, and brain-like computing, especially in the development of brain-computer interfaces (BCIs). The study of EEG signals is essential for the progress of BCIs, providing deep insights into the complex neural activities of the human brain.

In the past decade, a wide range of machine learning and deep learning algorithms have been applied to EEG data, leading to significant advancements in numerous applications. These applications include emotion recognition, motor imagery, mental workload assessment, seizure detection, Alzheimer's disease classification, and sleep stage scoring, among others (Craik et al., 2019; Kastrati et al., 2021a; 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; Murungi et al., 2023b; Dou et al., 2022; Zhou et al., 2022; Qu et al., 2020b,c,a, 2018, 2019; Saeidi et al., 2021; Qu and Hickey, 2022; Rasheed et al., 2020; Dadebayev et al., 2022; Wang and Qu, 2022;

Li et al., 2020; Aggarwal and Chugh, 2022). EEG and deep learning research have been in close proximity for decades, with advancements in both fields contributing to strides in our understanding of the brain. Deep learning algorithms are particularly popular in the context of EEG analysis due to their ability to extrapolate and generalize input information, making them ideal for decoding the complexities and noise within EEG signals into interpretable outputs.

The research presented in the EEGEyeNet study encapsulates this main point, presenting a large-scale EEG and eye-tracking dataset, designed specifically for ML and DL research Kastrati et al. (2021b). The study also presents a robust baseline for comparing the performance of new models on gaze estimation tasks. Due to its high-quality and accessible dataset, EEGEyeNet has been used by numerous studies experimenting with various DL techniques and architectures to improve performance. Their results, shown in Table 2, specifically highlight Convolutional Neural Networks (CNNs) as having a profound ability to interpret complex EEG data accurately, achieving a Root Mean Square Error (RMSE) value of 70.4. The contributions of combined EEG and deep learning heavily deepen our understanding of the brain and thus influence several related fields. EEG-based Artificial Intelligence (AI) systems can allow for the automated detection and monitoring of various neurological states and conditions, advancing fields such as neuroscience and BCI. In the case of EEGEyeNet, behavioral and neurological information can be deduced from eye tracking, further highlighting the importance of enhancing the capabilities of deep learning algorithms for EEG regression Cao (2020).

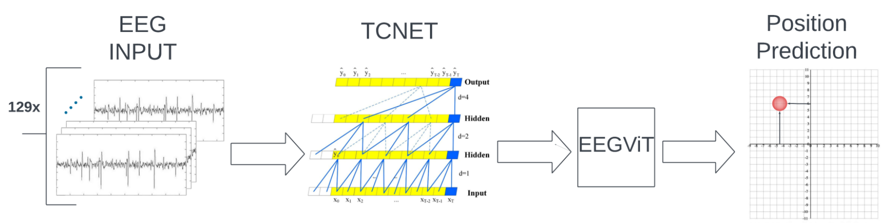

Fig. 1. *EEGViT-TCNet Model Diagram (Modesitt et al., 2024).*

2.2 Vision Transformers for EEG Regression

Vision Transformers are well regarded as being extremely adept at performing a wide variety of image tasks when pre-trained on large datasets Dosovitskiy et al. (2021). This is a result of the self-attention mechanism, which captures information across sequences of pixel patches, allowing the model to build a more robust representation of the entire image. While ViTs have been nominally utilized in image analysis, recent research has unmasked the applicability of ViTs for EEG analysis. The study presenting the EEGViT model highlighted how transformers' strong global and sequential data processing performs better at regression on time-series EEG data Yang and Modesitt (2023). Furthermore, the addition of prior feature extraction layers, such as a Temporal Convolutional Network (TCN) block shown in Figure 1, as discussed in EEGViT-TCNet, further extracts temporal features within sequences of EEG signals (Ingolfsson et al., 2020).

The strength of pre-trained ViTs for EEG regression lies in their ability to interpret sequence data as a whole and apply image-related priors to the EEG space. Moreover, evidence seems to suggest that decreased local connectivity may lead to more subject and noise invariance, broadening the variety of signal patterns that can be processed. Overall, ViTs prove to be an exceptional tool in the field of EEG signal analysis, overcoming the many limitations of other DL algorithms specific to EEG (Table 2) Yang and Modesitt (2023).

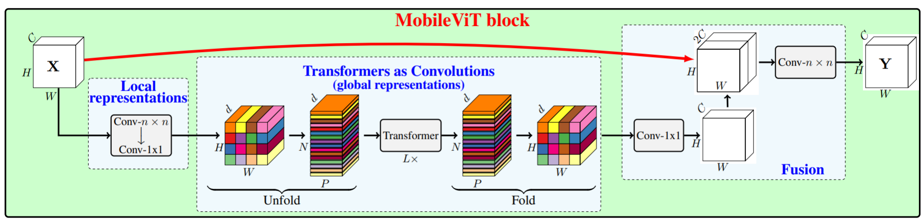

Fig. 2. *Model diagram for the MobileViT block, primary innovation of the MobileViT (Mehta and Rastegari, 2022a).*

2.3 Lightweight and Mobile Networks

The use of lightweight models is the standard for vision tasks on mobile or resource-constrained devices. Typically, these models are adapted from CNNs, which, due to their spatial inductive bias, yield better performance than ViTs at lower parameter sizes. Lightweight CNNs can also be more effortlessly adapted for downstream tasks and are generally easier to optimize than ViTs, which require extensive data augmentation and regularization Xiao et al. (2021). However, to learn global representations, ViTs need to be integrated in some manner. MobileViT introduces a hybrid, lightweight model that treats transformers as convolutions Mehta and Rastegari (2022a). They achieve this by modifying standard convolution operations to encode global features via a transformer, seen in Figure 2, permitting both local and global processing, resulting in better accuracy than other lightweight networks. Further optimizations were made to MobileViT, in the subsequently named MobileViTV2. This primarily involved altering the self-attention mechanism from using the multi-headed standard to a separable approach. Using this method, rather than attention scores being computed with respect to each patch, they are computed with respect to a single latent token, reducing the time complexity from quadratic to linear with respect to the number of patches. Additionally, the removal of the skip connection within the MobileViT block provided a minor improvement to the model's task performance.

The advancements of lightweight networks have implications directly related to building more deployable AI systems. Faster and more accurate models improve vision-related tasks on mobile devices, such as image recognition and augmented reality functionality. This extends outward from solely mobile phones and can apply to other resource-constrained devices, such as drones and wearable technology. When it comes to EEG analysis, both CNNs and Transformers are proven to be powerful tools for enhancing performance, and the adaptation of mobile networks can make strides in the practical application of DL in EEG signal processing Kastrati et al. (2021b); Yang and Modesitt (2023); Modesitt et al. (2024).

2.4 Knowledge Distillation for Lightweight Models

While lightweight models perform reasonably well across various image task benchmarks, they still fall short in terms of accuracy when compared to their larger counterparts. Due to the necessity of a resource-conscious design, architectural modifications are more difficult to make, limiting their overall performance Gao and Zhou (2023). One proven way to improve the performance of lightweight models is through model compression, with knowledge distillation being one well-known technique to achieve this. The main methodology behind this procedure is to have a larger, more complex network (teacher) with a smaller, simpler model (student) and train the student model using a ”distillation loss” calculated from the combined losses of the student and teacher models Hinton et al. (2015). This causes the student model to mimic the teacher model's behavior, effectively ”compressing” the teacher into a smaller network. This is especially good for lightweight networks, as they are able to enjoy enhanced predictive accuracy alongside their computational efficiency.

Knowledge distillation is an incredibly popular and useful technique in designing deployable DL models for resource-constrained devices, as this training procedure can result in better accuracy without affecting the model's size or speed Cui et al. (2024). This makes it an attractive method to aid in developing enhanced lightweight networks for EEG analysis. These networks, with higher accuracy and efficiency, have the potential to increase practical applications in medical contexts, where speed and accuracy are vital.

3. Methods

In our study, we concentrated on the Absolute Position task within the EEGEyeNet dataset, opting for the MobileViT architecture due to its proven superior performance relative to other similarly sized models. We incorporated a meticulously fine-tuned EEGViT-TCNet as the teacher model in our knowledge distillation process, aimed at enhancing the MobileViT student model's accuracy. The accuracy of our proposed model was rigorously evaluated using the Root Mean Square Error (RMSE) on the test set, in addition to measuring the computational speed during inference, and the model's parameter count.

3.1 Dataset

This study employed the EEGEyeNet dataset for the training and validation of our model. (Wang and Wang, 2022; Fuhl et al., 2023; Xiang and Abdelmonsef, 2022; Modesitt et al., 2023; Mishra et al., 2023) This dataset encompasses EEG and ET recordings from 356 adults, 190 of whom were female and 166 were male, ranging in age from 18 to 80 years old. Researchers obtained written consent from all participants prior to data collection and compensated participants monetarily. The EEG data was recorded on the EEG Geodesic Hydrocel system with 128 channels and a 500 Hz sampling rate. The impedance of each electrode was analyzed between recording sessions and kept at a maximum of 40 KOhm. Eye positions were concurrently recorded with an EyeLink 1000 Plus at the same sampling rate. The eye tracking was calibrated using a 9-point grid before each recording and validated to ensure an average error of less than 1° for the measurement of all points. Participants were seated 68 cm from a 24-inch monitor with a resolution of 800x600 pixels, with their heads stabilized in a chin rest position.

To address artifacts present in the EEG recordings as a result of environmental and psychological noise, the necessary preprocessing steps were taken. This process involved the detection and correction of bad electrodes, along with running a 40 Hz high-pass filter and a 0.5 Hz low-pass filter on the data. The EEG data was then synchronized alongside the eye-tracking data to ensure time-locked analyses at the onset of relevant events with errors not exceeding 2 ms.

In the Large Grid Paradigm, participants were asked to focus on a sequence of 25 dots located in different positions on the screen. Dots were each presented for roughly 1.5 to 1.8 seconds, and their positions were selected to ensure maximal coverage of the screen area. This procedure was separated into five experimental blocks, each displaying a series of 27 dots, with the center dot appearing three times in a pseudo-randomized order to reduce the predictability of subsequent dots. This entire experiment was then repeated six times.

The Absolute Position benchmark data was performed using the Large Grid Paradigm. This task involves determining the subject's gaze position as an XY coordinate pair. Each sample of one second describes a single fixation from a participant. The benchmark contains 21464 samples from 27 participants. This task was specifically important for our research as it provided a diverse array of gaze positions combined with a high sample count, allowing for a more comprehensive analysis of EEG-ET patterns, integral to determining the XY coordinate positions (Kastrati et al., 2021b).

3.2 Model Architecture

The architecture of our model can be seen in Figure 3. Our model design is adapted from the EEGViT-TCNet architecture and includes three main components that construct the entire network architecture.

Temporal Convolutional Network Block: The TCN block is initialized with a 129-dimension input layer, specifically attuned to the number of EEG channels with an extra dimension for encoding grounding information. Three additional layers with channel sizes 64, 128, and 256 are utilized to build a comprehensive representation of the temporal dependencies within the recorded EEG signals. A uniform kernel of

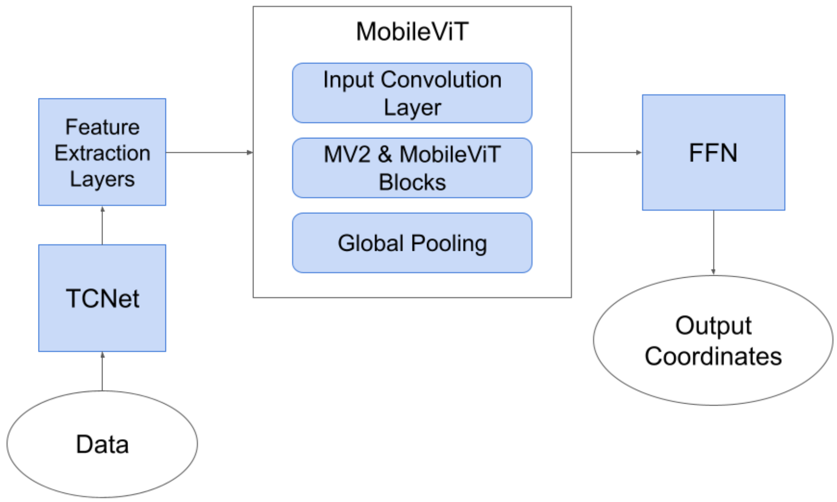

Fig. 3. *EEGMobile Model Flow Chart: MV2 refers to MobileNetV2 and FFN refers to Feed Forward Network.*

size 3 coupled with ReLU activation is applied to all layers with a dropout rate of 0.75 to counteract overfitting. Finally, causality and weight normalization are applied to improve stability Bai et al. (2018).

Feature Extraction Layers: This block consists of two convolutional layers designed to initiate feature extraction on the data output by TCN in preparation to be fed to the MobileViT layer. The first layer consists of a convolutional kernel of size (1, 36) scaling the input up to 256 channels, initiating spatial feature extraction from the input data. The second layer consists of a kernel of size (256, 1), further scaling the data to 768 channels and compressing spatial information in preparation for the transformer component Yang and Modesitt (2023).

MobileViT Network: This final block consists of a pre-trained MobileViTV2 network configured from the Hugging Face "apple/mobilevitv2-1.0-imagenet1k-256" model. The model is initialized with an input dimensionality of 768, perfectly aligning with the previous feature extraction component, along with a kernel size of (3, 3) for the internal convolutional layers and a patch size of (1, 1) for the internal transformers. Lastly, a classifier head is comprised of a classifier layer followed by a dropout of 0.1, finally connecting to a linear layer that outputs the final XY coordinate gaze prediction Mehta and Rastegari (2022a,b); Yang and Modesitt (2023).

3.3 Training and Evaluation:

Data from the Absolute Position tasks was split 70% for training, 15% for validation, and 15% for testing. To maintain data integrity, we ensured data points with eye positions outside of the 800-by-600 pixel range were excluded. Our model was then trained for a total of 15 epochs, with each training iteration consisting of 64 sample batches

Kastrati et al. (2021b). Our model utilized the Adam optimizer with a learning rate of 1e-3 and weight decay of 0.3, and a learning rate scheduler with a step size of 6 was utilized to prevent overfitting. Additionally, we employed knowledge distillation using a fine-tuned EEGViT-TCN teacher model to train our EEGMobile student model, due to its superior performance over other models on the Absolute Position Task. The loss was calculated as a function of a lambda-controlled, weighted sum between the Mean Square Error (MSE) between the student predictions and observed values, the ”true loss”, and the Kullback-Leibler divergence loss of the soft targets of both models, the ”distillation loss”. We set the temperature parameter, controlling the strength of the distillation, to 20 and lambda to 0.9, giving a higher weight to the distillation loss. Hinton et al. (2015).

Model accuracy was evaluated as the Euclidean distance between the model's predictions and observed values in millimeters, represented as the RMSE Kastrati et al. (2021b). We also evaluated the model's speed through an inference test, where the model was set to run inferencing across the entire Absolute Position task dataset ten times, after which the runtime in minutes was computed. Finally, we report parameter counts for the compared models to examine their relative sizes. All training and testing were conducted on the same machine using a single P5000 GPU.

4. Results

Results from our exhaustive evaluation illustrate EEGMobile's ability to deliver SOTA-comparable accuracy at a faster inferencing speed and smaller size, outperforming the other transformer-based models.

Table 3. *RMSE indicates validation error in millimeters. Runtime measures the time to complete inferencing on the entire dataset 10 times. Parameter count is measured in millions. All values represent the mean and standard deviation of 5 runs. All tests were run on the same hardware.*

Model	RMSE (mm)	Runtime (mins)	Parameter Count
CNN Kastrati et al. (2021b)	70.4 ± 1.1	2.1 ± 0.2	0.6 M
EEGViT Yang and Modesitt (2023)	55.4 ± 0.2	78.8 ± 0.2	86.0 M
EEGViT-TCN Modesitt et al. (2024)	51.8 ± 0.2	12.1 ± 1.6	137.2 M
EEGMobile (Ours)	**53.6 ± 0.6**	**8.1 ± 0.9**	**55.9 M**

Table 3 shows that our EEGMobile architecture attains an impressively low RMSE of 53.6 mm. This eclipses all EEGEyeNet baseline models, the lowest of which is the CNN with an RMSE of 70.4 (Kastrati et al., 2021b). Likewise, our model outperforms the highest-performing EEGViT model with an RMSE of 55.4 (Yang and Modesitt, 2023). EEGMobile also attains a similar accuracy to EEGViT-TCNet, just 3% shy of its 51.8 RMSE (Modesitt et al., 2024). This notable performance accuracy underscores the robustness of EEGMobile and its ability to capture and generalize key information within the training data.

In addition to evaluating the accuracy, we also assessed EEGMobile's computational efficiency through an inference speed test and analyzed the model's relative size. Compared to the other EEG-based ViT architectures, our model boasts a 33% increase in inferencing speed when compared to EEGViT-TCNet and is over 9.7× faster than the original EEGViT with a runtime of about 8.1 minutes, compared to 78.8 and 12.1 minutes for EEGViT and EEGViT-TCNet, respectively. However, EEGMobile is still slower than the baseline CNN model by about 3.9×. Furthermore, we find that while EEGMobile has a smaller size compared to the other ViT architectures, with 55.9 million parameters, followed by EEGViT with 86 million and EEGViT-TCN with 137.2 million, It is still much larger than the CNN model with only 609 thousand parameters. These sizes indicate that while EEGMobile is still about 91.7× larger than the CNN, it is also about 35% smaller than EEGViT and 60% smaller than EEGViT-TCN. This noteworthy improvement to both speed and size highlights EEGMobile's efficiency over the other transformer-based models and is especially paramount for applications involving real-time predictions on memory-restricted devices.

5. Discussion

Our meticulous performance evaluation legitimizes EEGMobile as a robust model capable of executing EEG-ET tasks efficiently when compared to other transformer-based models, reinforcing the viability of combining lightweight networks with distillation techniques to develop cost-effective models. Our RMSE results suggest that EEGMobile is capable of providing predictive capabilities comparable to the other transformer models on EEG gaze estimation tasks and can gain a robust interpretation of the EEG data despite its small size. Similarly, without the use of KD, EEGMobile has a much higher RMSE of about 76.8. This alludes to the idea that KD aids in guiding the model to learn better representations, enhancing its accuracy. This is further highlighted by the high temperature and lambda parameters, which greatly increased the strength and effect of distillation. We also note that EEGMobile's memory and time efficiency are still very much below those of CNN. However, when examining the components of all three ViT-based models, a significant amount of the total parameter counts stem from the feature extraction layers, which contain over 50 million parameters. As such, there is a strong possibility that integrating feature extraction methods with significantly fewer parameters may further close the size gap between EEGMobile and CNN in terms of speed and size. Our analysis of EEGMobile's size, speed, and task performance illustrates its potential as an efficient model for practical EEG signal analysis, especially on weaker devices, across various domains.

This improved efficiency is material to real-time EEG and eye-tracking applications, where speed and accuracy are important EL Menshawy et al. (2015). Our findings are also particularly relevant to the HCI community, such as in the integration of EEG recording devices into virtual reality and augmented reality technologies to improve interactivity Xiang and Abdelmonsef (2022); Rolff et al. (2022); Xu et al. (2023). Additionally, due to the model's low parameter count, there are also memory efficiency benefits when it is utilized by resource-constrained BCI devices, allowing the model to be run on-device, making medical diagnostic tools more accessible to the average person Jebelli et al. (2019).

Due to time constraints, our experiment only tested the MobileViT and MobileViTV2 architectures, ultimately selecting the V2 in our final results. However, a MobileViTV3, which makes a number of improvements to further enhance the model's performance, has also been developed Wadekar and Chaurasia (2022). Not only does utilizing this pre-trained model have the potential to improve our results, but it would also be more representative of the capabilities of MobileViT on the Absolute Position Task. Similarly, testing our model on other EEG datasets would allow us to better evaluate its robustness for EEG analysis as a whole.

In future studies, exploring alternative lightweight model architectures has the potential to yield better accuracy and speed. Experimenting with alternative distillation techniques, such as feature distillation, may also lead to improved performance. Additionally, since there is no theoretical constraint on the teacher model used in the distillation paradigm, research could focus on developing a large, high-performance model to train a smaller, more practical model using the same training procedure as EEGMobile. One notable concern with the current methodology is the independent training of both the teacher and student models, which can be inefficient, especially with very large and slow teacher models. Research into parallelizing this training process may increase the scalability of these models. Furthermore, exploring various deep learning techniques on different datasets for comparative studies (An et al. (2023a,b); Jiang et al. (2023); Lu et al. (2023); Chen et al. (2024); Ma (2022); Ma et al. (2024); Gui 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.

6. Conclusion

This study highlights the efficacy of lightweight models for EEG-ET tasks. By integrating the MobileViT architecture into a Hybrid Transformer model and utilizing knowledge distillation techniques, we present a model with enhanced speed and size with a minimal cost to accuracy compared to the SOTA. Our findings further validate the potential of lightweight networks on EEG regression tasks, but further expand the accessibility of DL-based EEG analysis tools for real-world applications, especially on resource-constrained devices, with the potential to advance the unification of neuroscience and HCI.

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, 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.

Zehong Cao. A review of artificial intelligence for eeg-based braincomputer interfaces and applications. *Brain Science Advances*, 6(3):162–170, 2020. `https://doi.org/10.26599/BSA.2020.9050017`. URL `https://doi.org/10.26599/BSA.2020.9050017`.

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.

Yiming Cui, Jiajia Guo, Zheng Cao, Huaze Tang, Chao-Kai Wen, Shi Jin, Xin Wang, and Xiaolin Hou. Lightweight neural network with knowledge distillation for csi feedback, 2024.

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.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

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.

Mohamed EL Menshawy, Abdelghani Benharref, and Mohamed Serhani. An automatic mobile-health based approach for eeg epileptic seizures detection. *Expert Systems with Applications*, 42(20):7157–7174, 2015. ISSN 0957-4174. https://doi.org/https://doi.org/10.1016/j.eswa.2015.04.068. URL https://www.sciencedirect.com/science/article/pii/S0957417415003103.

Emma Farago, Andrew J Law, Sujoy Ghosh Hajra, and Adrian DC Chan. Blink and saccade detection from forehead eeg. In *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2022.

Wolfgang Fuhl, Susanne Zabel, Theresa Harbig, Julia-Astrid Moldt, Teresa Festl Wietek, Anne Herrmann-Werner, and Kay Nieselt. One step closer to eeg based eye tracking. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, pages 1–7, 2023.

Dandan Gao and Dengwen Zhou. A very lightweight and efficient image super-resolution network. *Expert Systems with Applications*, 213:118898, 2023. ISSN 0957-4174. https://doi.org/https://doi.org/10.1016/j.eswa.2022.118898. URL https://www.sciencedirect.com/science/article/pii/S0957417422019169.

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.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015.

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, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini. 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)*, pages 2958–2965, 2020. https://doi.org/10.1109/SMC42975.2020.9283028.

Houtan Jebelli, Mohammad Mahdi Khalili, and SangHyun Lee. Mobile eeg-based workers' stress recognition by applying deep neural network. In Ivan Mutis and Timo Hartmann, editors, *Advances in Informatics and Computing in Civil and Construction Engineering*, pages 173–180, Cham, 2019. Springer International Publishing. ISBN 978-3-030-00220-6.

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, Martyna Beata Płomecka, Damián Pascual, Lukas Wolf, Victor Gillioz, Roger Wattenhofer, and Nicolas Langer. Eegeyenet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. *arXiv preprint arXiv:2111.05100*, 2021a.

Ard Kastrati, Martyna Beata Płomecka, Damián Pascual, Lukas Wolf, Victor Gillioz, Roger Wattenhofer, and Nicolas Langer. Eegeyenet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction, 2021b.

Ard Kastrati, Martyna Beata Plomecka, Joël Küchler, Nicolas Langer, and Roger Wattenhofer. Electrode clustering and bandpass analysis of eeg data for gaze estimation. In *Annual Conference on Neural Information Processing Systems*, pages 50–65. PMLR, 2023.

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.

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.

Sachin Mehta and Mohammad Rastegari. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer, 2022a.

Sachin Mehta and Mohammad Rastegari. Separable self-attention for mobile vision transformers, 2022b.

Ashish Ranjan Mishra, Rakesh Kumar, Vibha Gupta, Sameer Prabhu, Richa Upadhyay, Prakash Chandra Chhipa, Sumit Rakesh, Hamam Mokayed, Marcus Liwicki, Foteini Simistira Liwicki, et al. Signeeg v1. 0: Multimodal electroencephalography and signature database for biometric systems. *bioRxiv*, pages 2023–09, 2023.

Eric Modesitt, Ruiqi Yang, and Qi Liu. Two heads are better than one: A bio-inspired method for improving classification on eeg-et data. In *International Conference on Human-Computer Interaction*, pages 382–390. Springer, 2023.

Eric Modesitt, Haicheng Yin, Williams Huang Wang, and Brian Lu. Fusing pretrained vits with tcnet for enhanced eeg regression, 2024.

Nathan Koome Murungi, Michael Vinh Pham, Xufeng Dai, and Xiaodong Qu. Trends in machine learning and electroencephalogram (eeg): A review for undergraduate researchers. In *International Conference on Human-Computer Interaction*, pages 426–443. Springer, 2023a.

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*, 2023b.

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.

Tim Rolff, H Matthias Harms, Frank Steinicke, and Simone Frintrop. Gazetransformer: Gaze forecasting for virtual reality using transformer networks. In *DAGM German Conference on Pattern Recognition*, pages 577–593. Springer, 2022.

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.

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.

Congzhong Sun and Chaozhou Mou. Survey on the research direction of eeg-based signal processing. *Frontiers in Neuroscience*, 17, July 2023. ISSN 1662-453X. `https://doi.org/10.3389/fnins.2023.1203059`. URL `http://dx.doi.org/10.3389/fnins.2023.1203059`.

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.

Shakti N. Wadekar and Abhishek Chaurasia. Mobilevitv3: Mobile-friendly vision transformer with simple and effective fusion of local, global and input features, 2022.

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.

Xuduo Wang and Ziji Wang. Cnn with self-attention in eeg classification. In *International Conference on Human-Computer Interaction*, pages 512–526. Springer, 2022.

Lukas Wolf, Ard Kastrati, Martyna Beata Płomecka, Jie-Ming Li, Dustin Klebe, Alexander Veicht, Roger Wattenhofer, and Nicolas Langer. A deep learning approach for the segmentation of electroencephalography data in eye tracking applications. *arXiv preprint arXiv:2206.08672*, 2022.

Brian Xiang and Abdelrahman Abdelmonsef. Vector-based data improves left-right eye-tracking classifier performance after a covariate distributional shift. In *International Conference on Human-Computer Interaction*, pages 617–632. Springer, 2022.

Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross Girshick. Early convolutions help transformers see better, 2021.

Xu Xu, Lei Yang, Yan Yan, and Congsheng Li. Cmfs-net: Common mode features suppression network for gaze estimation. In *Proceedings of the 2023 Workshop on Advanced Multimedia Computing for Smart Manufacturing and Engineering*, AMC-SME '23, page 25–29, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400702730. `https://doi.org/10.1145/3606042.3616449`. URL `https://doi.org/10.1145/3606042.3616449`.

Ruiqi Yang and Eric Modesitt. Vit2eeg: Leveraging hybrid pretrained vision transformers for eeg data, 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.