Hardware-Assisted Privacy-Preserving Multi-Channel EEG Computational Headwear

Abdul Aziz*, Bhawana Chhaglani*, Amirmohammad Radmehr, Joseph Collins,
Jeremy Gummeson, Sunghoon Ivan Lee, Ravi Karkar, and Phuc Nguyen
University of Massachusetts Amherst, Amherst, MA, USA
*Co-primary Student Author

***Abstract*—EEG signals contain highly sensitive information about an individual's mental state, cognitive processes, and health conditions, making privacy preservation crucial. With the rise of commercial headwear capable of capturing EEG signals, developing robust mechanisms for ensuring privacy of such data is imperative. This work aims to protect EEG data privacy in cloud-based processing systems by sending intermediate output after neural network layer splitting to the cloud. We propose a novel holistic Combined Privacy Metric (CPM) that quantifies privacy leakage between raw EEG signals and intermediate outputs. Our study focuses on EEG-based seizure detection using a 1D CNN architecture, achieving accuracy of 96.25%. We evaluate various splitting configurations to optimize the trade-off between privacy preservation and computational efficiency. We find that splitting after the second convolutional layer achieves a CPM of 0.82 with a modest client-side model size of 509kB. This approach significantly enhances EEG data privacy while enabling effective cloud-based analysis, potentially facilitating wider adoption of secure EEG technologies in healthcare and research applications.**

***Index Terms*—EEG signals, neural networks, privacy leakage, computational headwear**

I. Introduction

The integration of sensors into head-worn devices has revolutionized our ability to capture brain signals non-invasively, significantly enhancing our understanding. This technological advancement has paved the way for a wide range of applications utilizing EEG headwear including Brain-Computer Interfaces (BCI), Human-Computer Interaction (HCI), cognitive load monitoring, and brain disorder surveillance. The field has seen remarkable innovations, attracting attention from both industry leaders and academic researchers. Industry giants like Apple have proposed ways to monitor brain activity through their AirPods [3], while researchers explore applications ranging from sleep pattern tracking [12] to monitoring neurological disorders such as seizure [11] or dementia [8].

Although exciting, head-worn wearables sense sensitive physiological signals such as EEG, EOG, EMG, HRV, and EDA, often leveraging external servers or the cloud for complex processing thereby introducing privacy and security vulnerabilities. EEG data, in particular, presents significant privacy concerns due to its highly sensitive nature [10]. Brain activity patterns can reveal mental workload, attention levels, and emotional responses. The increasing prevalence of wearable EEG devices has expanded data collection beyond clinical settings, raising concerns about user awareness and data control. When combined with other biometric data, EEG information can potentially enable individual identification, even if anonymized, heightening risks of unauthorized profiling or discriminatory practices.

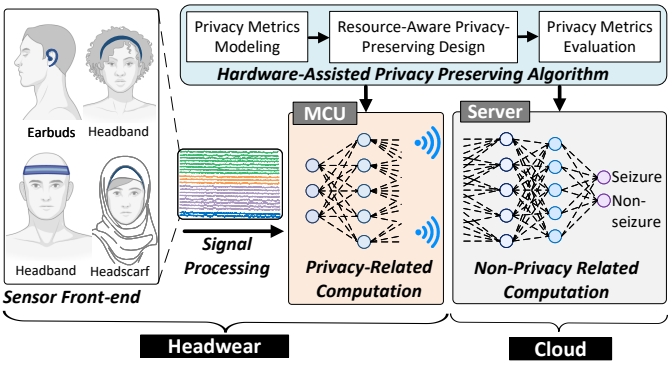

Fig. 1: We propose to investigate the computational offloading method where privacy signals are computed and erased immediately after sensing at the hardware level, and the remaining computations are performed on the cloud.

To address these privacy concerns, researchers have developed techniques to address privacy concerns such as encryption [4], differential privacy [7], secure multi-party computation [2]. Frameworks such as Generative Adversarial Networks (GANs) help generate and classify synthetic EEG data while preserving privacy [5]. Layer splitting in neural networks has been explored previously for privacy-preservation, but existing approaches lack a comprehensive method to determine the optimal split point to optimize privacy-efficiency trade-offs. Abuadbba et al. [1] found high privacy leakage in 1D CNN models, indicating inadequate protection for raw data in split architectures. They used correlation and DTW distance as privacy metrics, but a holistic metric is still lacking. Without a holistic metric informed decisions about the privacy-efficiency trade-off, critical in edge computing, are hindered. Malekzadeh et al. [9] proposed Salted DNNs allowing edge clients to control DNN output interpretation without revealing true outputs. However, there is still a lack of a robust framework to quantify privacy leakage across neural network layers.

In this research, we aim to enhance the privacy of wearable sensors by maximizing on-device computation while addressing the limitations of existing privacy-preserving techniques. We propose a novel approach to offload the computation of

privacy-sensitive data to the edge, on the wearable device hardware, for privacy-efficiency EEG sensing and computing as illustrated in Figure 1. Note that existing works perform splitting to reduce computation on edge devices without considering privacy metrics. To be specific, we study current privacy practices and introduce a novel, comprehensive Combined Privacy Metric (CPM) that incorporates multiple statistical, geometric, and information-theoretic measures. We then used these metrics as the key criteria and developed an optimized Convolutional Neural Network (CNN) architecture, partitioned in two subnetworks: one to perform sensitive related computation deployed on Commercial Off-The-Shelf (COTS) EEG headwear and another to compute non-privacy-sensitive on external servers. This holistic approach allows us to quantify privacy leakage across different layers of the neural network, enabling informed decisions about where to offload the computation for optimizing both privacy protection and computational efficiency. In summary, this paper makes the following contributions:

- We explore statistical, geometric, and information-theoretic privacy metrics (CPM) to measure privacy leakage when transferring user data from edge to cloud.
- We integrate the above finding into a seizure detection CNN model based on EEG data from real patients.
- We propose a standardized approach to identify the optimal layer to split a machine learning model, maximizing privacy with minimal loss in accuracy.
- We discuss our findings related to the variation in privacy and efficiency based on different types of computational offloading.

Our results demonstrate that our approach enables systematic optimization of machine learning model deployment on wearable devices, enhancing privacy without compromising efficiency and accuracy in EEG-based seizure detection.

II. Methods

In our proposed approach, a machine learning model is split across a client and the server where: (i) the initial layers are deployed on the wearable device (client), and (ii) the remaining layers are deployed on the server. Only the intermediate activations, which are less sensitive and contain abstracted features rather than raw data, are transmitted between the two parts, significantly reducing the risk of exposing personal and sensitive information. We hypothesize that if the intermediate activations from the client output are significantly different from the raw inputs, they will not reveal any meaningful information about the raw data.

Analyzing Privacy: To evaluate the privacy preservation of our EEG signal processing system, we implemented a comprehensive set of privacy metrics aimed at quantifying the information leakage between the raw EEG signal and the intermediate output after layer splitting in our 1D CNN model. We utilized several statistical, geometric, and information-theoretic measures to provide a multifaceted analysis of privacy preservation. EEG data contains various types of information, including temporal patterns, amplitude variations, and frequency components and different metrics are sensitive to different aspects of this information. These metrics included cosine similarity (c_{sim}) to assess directional similarity, Pearson (p_{corr}) and Spearman correlations (s_{corr}) to evaluate linear and monotonic relationships, and mutual information (m_{info}) score to quantify the mutual dependence between the raw and intermediate signals. To account for potential temporal variations, we incorporated Dynamic Time Warping (DTW) (d_{dist}) distance, which measures the similarity between temporal sequences. Additionally, we used reconstruction error (r_{err}), calculated as mean squared error, to quantify how well the intermediate representation could reconstruct the original signal. For cosine similarity, correlations, and mutual information, lower values indicate better privacy preservation. Conversely, higher values of distances and reconstruction error suggest improved privacy.

$$
\begin{aligned}
p_{norm} &= 1 - \frac{|p_{corr}|}{\max(|p_{corr}|)}; s_{norm} = 1 - \frac{|s_{corr}|}{\max(|s_{corr}|)} \\
m_{norm} &= 1 - \frac{m_{info}}{\max(m_{info})}; c_{norm} = 1 - \frac{|c_{sim}|}{\max(|c_{sim}|)} \\
r_{norm} &= \frac{r_{err}}{\max(r_{err})}; d_{norm} = \frac{d_{dist}}{\max(d_{dist})}
\end{aligned} \tag{1}
$$

$$
CPM = \frac{1}{n}\sum_{i=1}^{n} M_i \tag{2}
$$

These metrics allow for evaluating how well the input is transformed over the client-side network and how much information a malicious party could glean, should they get access to the intermediate data during transmission. By combining these diverse metrics into our Combined Privacy Metric (CPM), we create a more comprehensive and robust measure of privacy. CPM is the average of normalized values of all the privacy metrics as shown in equation (2) where; where M_i is the ith privacy metric. For p_{corr}, s_{corr}, m_{info}, and c_{sim}, we subtract them from 1 because the lower values of these metrics reflect lower privacy leakage. So, the higher values of CPM would indicate lower privacy leakage. This approach allows us to capture a wider range of potential privacy leaks and provides a more nuanced understanding of the privacy-preservation capabilities of our system.

Privacy-Efficiency Trade-offs: The privacy improves as more neural network layers are processed on the client side as less raw data is transmitted to the server. However, this increases local processing requiring more computational resources from the client device potentially affecting performance, especially on resource-constrained devices. Conversely, server-side processing reduces client burden improving efficiency and potentially allowing for more complex models, but at the cost of transmitting more sensitive data. Our study aims to identify an optimal split point balancing privacy protection with resource efficiency, considering client model size, computational complexity, privacy metrics, and system accuracy. To determine the optimal layer to split, we use client model size and CPM. We evaluate each layer of

the trained model to assess the privacy metrics between the raw input data and the intermediate activations. This score reflects the degree of information obfuscation achieved at each layer. The layer with the highest combined privacy score that still has a manageable client model size is the optimal point to split the network, ensuring maximum privacy preservation with minimal impact on computational complexity.

III. Implementation

Dataset: In our investigation, we used our own EEG dataset collected from real patients in a hospital setting to classify between seizures and non-seizures. Using a standard 21-channel scalp-EEG setup, we recorded data from 33 epilepsy patients aged between 19 and 74, with 17 biological males and 16 biological females represented, in the Epileptic Monitoring Unit (EMU). We collected 1320 hours of EEG recordings, in which 22m 35s are seizure events (rare seizure onsets are frequently observed and anticipated in seizure studies). To balance the dataset, we removed most non-seizure events and equalized the size of the resting, speaking, walking, and eating data. This results in 88 minutes of data for evaluation. Following American Clinical Neurophysiology Society guidelines [6], we segmented our data into 10s chunks and labeled them as seizure (1) if the chunk fell completely between seizure onset and offset times, or non-seizure (0) otherwise.

ML Architecture: We trained a 1D CNN network for our seizure classification task. The architecture consists of four 1D convolutional layers with 32, 64, 128, and 256 filters respectively. The kernel size was kept to 3 with padding and stride both set to 1 for all layers. We applied the ReLU activation function and used MaxPooling with a kernel size of 2 after each convolution. The convolutional layers were followed by three fully connected layers with 256, 128, and 64 units respectively. We also used a dropout layer with a probability of 0.5 after the first and second fully connected layers. The final fully connected layer outputs the binary classification outcome. We trained the model for 30 epochs using 80% of the data for training and 20% for testing. Out of the training set, we use 20% for validation.

Configurations: To study privacy preservation, we divided our model into client-side and server-side components at various points, ranging from the first convolutional layer to the first fully connected layer. This gave us five test configurations with each configuration representing a different trade-off between data protection and computational efficiency across different model architectures.

Privacy Metrics: To ensure dimensional compatibility between raw EEG signals and intermediate outputs for our privacy analysis, we used Pytorch's *nn.functional.interpolate* method with linear interpolation. This enabled direct comparisons and accurate calculation of privacy metrics at various stages of our neural network model. The privacy metrics were calculated using SciPy library functions, and a custom implementation of DTW. Each metric was calculated for individual data samples and then averaged across the entire test dataset to provide robust estimates. We also monitored client-side model size to evaluate the privacy-efficiency trade-off crucial in edge computing scenarios. By applying this comprehensive suite of metrics to different layer-splitting scenarios in our CNN architecture, we were able to conduct a thorough analysis of the privacy-utility trade-offs in our EEG-based seizure detection system.

IV. Results

Baseline: Our completed EEG-based seizure detection model achieves the test accuracy of 96.25% after 30 epochs with 100% precision, 88.75% recall, and 94.04% F1 score.

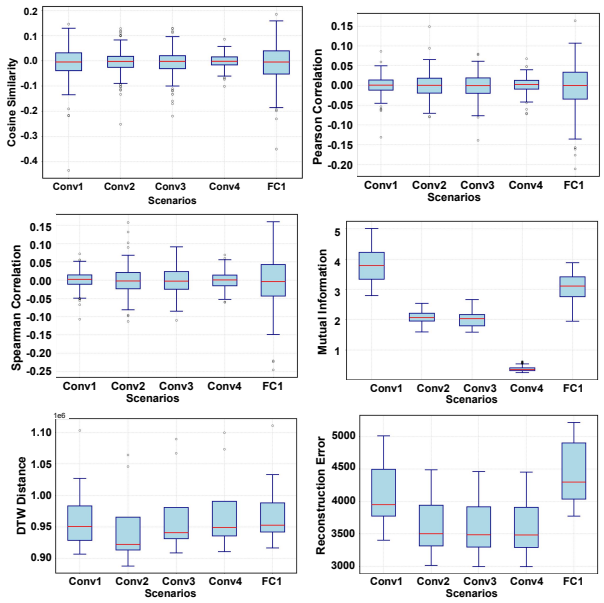

Fig. 2: Variation of privacy metrics for various layer splitting

Privacy-Preserving Performance: Figure 2 shows how different privacy metrics change when breaking the model at various layers: Conv1, Conv2, Conv3, Conv4, and FC1. For cosine similarity, Pearson correlation, and Spearman correlation, lower values indicate better privacy preservation. The boxplots show that these metrics slightly decrease as we move from Conv1 to Conv4, suggesting improved privacy. For instance, the mean cosine similarity decreases from Conv1 (0.021) to Conv4 (0.018), Pearson correlation from Conv1 (0.015) to Conv4 (0.012), and Spearman correlation from Conv1 (0.018) to Conv4 (0.015). FC1 shows a slight increase in these metrics, indicating a potential privacy trade-off at this layer. The mutual information score, which quantifies the dependence between the input and intermediate representations, shows a clear decreasing trend from Conv1 (mean: 3.82) to Conv4 (mean: 0.42). This decrease suggests that later layers retain less information about the original input, enhancing privacy. Interestingly, FC1 shows an increase in mutual information, further supporting the observation of a privacy trade-off at this layer. The Dynamic Time Warping distance shows an increasing trend from Conv1 (mean: 952,234) to Conv4 (mean: 957,347), with a slight decrease at FC1 (mean: 956,789). Higher DTW distances suggest greater

dissimilarity between the original and transformed signals, implying improved privacy preservation in deeper layers. Interestingly, the reconstruction error exhibits a decreasing trend from Conv1 (mean: 4,021) to Conv4 (mean: 3,498), followed by an increase at FC1 (mean: 4,326). It's important to note that these trends are not uniformly linear across all metrics. For instance, the reconstruction error trend differs from the others, highlighting the complexity of privacy dynamics in neural networks and the value of using multiple metrics for a comprehensive assessment. These nuanced trends across different metrics underscore the importance of a multi-faceted approach to privacy assessment in split learning scenarios.

Privacy-Efficiency Performance: Figure 3 provides a holistic view of the privacy-efficiency trade-off by plotting the Combined Privacy Metric (CPM) against the client model size. The CPM, which aggregates all six privacy metrics, shows that Conv4 achieves the highest privacy score (0.82) but at the cost of a larger client model size (509.62 kB). Conv2 presents an interesting balance, with a relatively high CPM (0.70) and a modest model size (28.12 kB). Conv1, while having the smallest model size (3.88 kB), achieves a CPM (0.66) comparable to Conv2. Notably, the FC1 configuration, despite having the largest client model size (20990.12 kB), shows the lowest CPM (0.36). This result underscores the non-linear relationship between model complexity and privacy preservation in split learning scenarios. These findings suggest that splitting the model at Conv2 or Conv4 could offer optimal trade-offs between privacy preservation and computational efficiency on the client device. The choice between these two configurations would depend on the specific requirements of the application, balancing the need for privacy with the computational constraints of the wearable EEG device.

Key Takeaways: Sending intermediate output after convolutions layers significantly improves privacy over sending raw data to the cloud while retaining the model accuracy. We can use CPM and model size to identify the best layer to split to maintain privacy and efficiency. Privacy preservation generally improves as the split point moves deeper into the network (from Conv1 to Conv4), as evidenced by the trends in various privacy metrics. However, splitting at the FC1 layer shows a decrease in privacy, indicating a potential trade-off. The CPM analysis reveals that Conv4 achieves the highest privacy score (0.82), while Conv2 offers a balanced trade-off between privacy (0.70) and client model size (28.12 kB), suggesting these two configurations as optimal choices depending on the specific application requirements and device constraints.

Generalizability: The current study focuses on EEG signal processing for seizure detection, yet the insights gained can be extended to other domains involving privacy-sensitive data. Further research should be conducted to ensure the proposed approach's effectiveness across different datasets, tasks, architectures, and applications, but the CPM introduced in this study offers a versatile approach to quantifying privacy in split learning scenarios. In addition, our CPM also provides a standardized method to assess privacy-efficiency trade-offs when splitting neural networks at different layers. Although this study focuses on a 1D CNN for EEG-based seizure detection, the CPM framework can be readily adapted to other model architectures and data types.

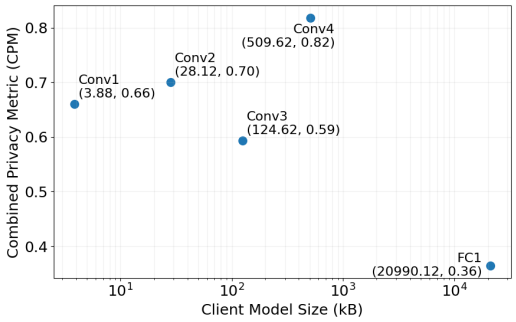

Fig. 3: CPM and model size deployed on wearable devices across different layer splitting

V. CONCLUSION

This paper presented a computational offloading approach for privacy-preserving EEG-based seizure detection. We also introduce a Combined Privacy Metric (CPM) to quantify privacy preservation. Our experiments demonstrated an optimal balance between privacy, efficiency, and performance by splitting the network after the second convolutional layer. This configuration achieved a CPM of 0.82, with high seizure detection accuracy (94.04% F1 score) and a modest client model size of 509 kB. These results show the feasibility of protecting EEG data privacy without compromising analysis accuracy, potentially enabling wider adoption of secure EEG-based technologies in various applications.

REFERENCES

[1] Sharif Abuadbba et al. Can we use split learning on 1d cnn models for privacy preserving training? In *ACM ACCCS*, pages 305–318, 2020.

[2] Anisha Agarwal et al. Protecting privacy of users in brain-computer interface applications. *IEEE NSRE*, 27(8):1546–1555, 2019.

[3] Erdrin Azemi et al. Biosignal Sensing Device Using Dynamic Selection of Electrodes, July 2023.

[4] AJ Bidgoly, HJ Bidgoly, and Z Arezoumand. Towards a universal and privacy preserving eeg-based authentication system. sci rep 12, 2531.

[5] Ahmed G Habashi, Ahmed M Azab, Seif Eldawlatly, and Gamal M Aly. Generative adversarial networks in eeg analysis: an overview. *Journal of NeuroEngineering and Rehabilitation*, 20(1):40, 2023.

[6] Lawrence J. Hirsch et al. American Clinical Neurophysiology Society's Standardized Critical Care EEG Terminology: 2021 Version. *Journal of clinical neurophysiology*, 38(1):1–29, January 2021.

[7] Nusrat Jahan. Eeg data privacy enhancement using differential privacy in wgan-based federated learning.

[8] José Juez et al. Development of a wearable system with In-Ear EEG electrodes for the monitoring of brain activities: An application to epilepsy. In *2021 IEEE ICBEB*, pages 1–4, October 2021.

[9] Mohammad Malekzadeh and Fahim Kawsar. Salted inference: Enhancing privacy while maintaining efficiency of split inference in mobile computing. In *ACM HOTMOBILE*, pages 14–20, 2024.

[10] Lubin Meng, Xue Jiang, Jian Huang, Wei Li, Hanbin Luo, and Dongrui Wu. User identity protection in eeg-based brain-computer interfaces: Supplementary material. *IEEE NSRE*, 2023.

[11] Jonas Munch Nielsen, Ivan C. Zibrandtsen, Paolo Masulli, Torben Lykke Sørensen, Tobias S. Andersen, and Troels Wesenberg Kjær. Towards a wearable multi-modal seizure detection system in epilepsy: A pilot study. *Clinical Neurophysiology*, 136:40–48, April 2022.

[12] Nhat Pham et al. WAKE: a behind-the-ear wearable system for microsleep detection. In *ACM MobiSys*, June 2020.