Are Edge Models Sufficient for sEMG-Based Gesture Recognition?

Mustapha Deji Dere, Sarfraz Ali, Saehyung Cheong, Ji-Hun Jo and Boreom Lee, *Member, IEEE*

***Abstract*—Recent advancements in electromyography (EMG)-based gesture decoding have provided a robust foundation for active rehabilitation device control. EMG literature has witnessed a proliferation of models, ranging from shallow to deep architecture, each offering unique trade-offs in terms of computational efficiency and accuracy. However, there is a notable trend among researchers adopting pre-existing computer vision-based models directly into EMG studies, despite their often computationally intensive nature, which hinders deployment to edge devices. The deployment of EMG decoders remains critical due to concerns related to privacy preservation and latency minimization. While edge models offer advantages in terms of deployment efficiency, deeper models typically deliver superior performance. However, the EMG literature currently lacks a definitive consensus on the comparative benefits of deep versus shallow/edge models for sEMG-based gesture classification. This study addresses this gap by investigating whether deep models are always necessary for sEMG-based gesture recognition. Specifically, we adopted lightweight models optimized for deployment on extreme edge devices and evaluated their performance relative to deeper models commonly employed in EMG studies. Our experimental results demonstrate that the edge model achieves comparable accuracy to deeper models while significantly reducing both parameters count and multiply-accumulate operations—features critical for real-time deployment on resource-constrained devices. These findings suggest that lightweight, shallow, and edge models represent a promising alternative for sEMG-based gesture classification, offering a practical balance between performance and computational efficiency. Consequently, we advocate for an increased focus on the development of optimized edge models within the EMG literature to better serve the needs of edge-device applications in active rehabilitation device development. The experiment can be replicated using: https://github.com/deremustapha/EdgeVsDeep**

Clinical Relevance— This study is of importance, as it investigates and recommends lightweight models in contrast to deep models for sEMG-based gesture recognition. This will help the BCI and rehabilitation device development research to focus on efficient edge models that offer efficiency in parameters and FLOPs, allowing for extreme edge deployment while protecting data privacy.

I. Introduction

Deep learning models are increasingly adopted for EMG-based gesture classification. Classical deep learning models without optimization are computational expensive for extreme edge deployment as shown in Fig 1. The large number of parameters and multiply-accumulate (MAC) operations in deep models necessitates deploying deep learning models on the cloud for efficient processing. Cloud deployment applications are subject to data privacy concerns in addition to inefficient latency compared to extreme new sensor deployment [1].

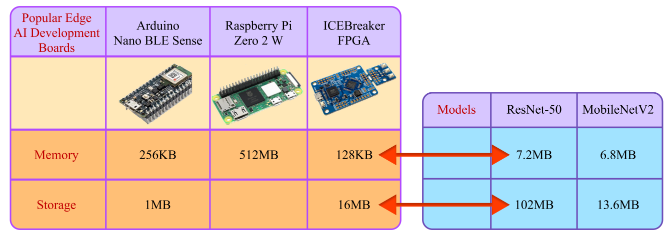

Fig. 1. Illustrating contrast between deep neural network and edge devices.

Active neuro-inspired rehabilitation device design requires decoding motor-intent from biosignals such as electroencephalography (EEG), electromyography (EMG), or other brain-related signals. Among these, EMG is preferred due to its higher signal-to-noise ratio (SNR) compared to EEG [2]. The decoding of gestures or motor-intention from surface electromyography (sEMG) has traditionally relied on decomposition algorithm-based approaches, and classical machine learning methods. However, there has been a paradigm shift toward deep learning methods for sEMG-based gesture decoding due to their high performance without requiring domain-specific feature extraction. While these deep learning decoders offer superior accuracy, they come with significant computational costs, which limits their deployment on extreme edge devices—a critical requirement for realizing an ideal active-rehabilitation device [2].

In recent EMG studies, there has been a notable trend towards adopting computationally intensive deep learning approaches inspired by architectures in computer vision and natural language processing, such as ResNets and Transformers (including Vision Transformers) [3]. These models have been proposed to enhance the accuracy of gesture classification from sEMG signals. However, existing studies primarily focus on offline performance metrics without adequately addressing the challenges associated with deploying these models on edge devices characterized by limited computational resources—a critical bottleneck for real-world applications. Additionally, distribution shifts resulting from factors such as electrode movement, muscle fatigue, intersubject variability, and intra-session differences require adaptive on-device model updates

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT), (No. RS-2023-00277220). (Corresponding author: Boreom Lee).

Boreom Lee, Mustapha Deji Dere, Sarfraz Ali, Saehyung Cheong and Ji-Hun Jo are with the Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea, and with AI Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea (e-mail: leebr@gist.ac.kr)

[4]. Implementing optimization techniques to deploy deep neural networks (DNN) decoders effectively remains a significant challenge.

TABLE I
COMPARISON OF MODEL PARAMETERS AND MAC. MACS EVALUATED WITH DB2

Model	Param	MAC
Edge Models		
MCUNet [5]	0.584M	3.950M
ProxyLessNAS [6]	3.002M	7.616M
CTRLEMG [7]	0.413M	0.048M
Deep Models		
ViT [8]	3.236M	3.3289M
ResNet18 [3]	11.19M	28.93M

Studies have been conducted to recommend optimal electromechanical delays for real-time operation in sEMG-based applications [9]. Appropriate window sizes with respective overlaps have been proposed for sEMG use [10]. However, there is currently no study or consensus in EMG research to evaluate the effectiveness of edge models compared to deep learning models for sEMG-based gesture decoding. In this study, we aim to compare the performance between edge and deep learning models for sEMG-based classification. Our hypothesis is that shallow/light weight edge models could achieve comparable decoding accuracy to deep models. This finding will provide a clear pathway and guide future research in EMG studies, encouraging a shift towards lightweight architectures rather than focusing solely on complex deep learning implementations.

II. METHODOLOGY

A. Between Deep and Edge Models

The trade-off between edge and deep models lies somewhere between simplicity (shallow) and complexity (deep), as evidenced by their characteristics regarding parameters and multiply–accumulate (MAC) operations. In this study, we differentiate these models based on their resources requirements for deployment on edge devices such as Arduino Nano BLE Sense, Raspberry Pi, and Field-programmable gate array (FPGA) in Fig. 1. These edge devices were chosen due to their low-power consumption, form-size and suitability for real-world applications in active neuro-rehabilitation device design, where efficient computation, latency and form-size are crucial.

Models that exceed the capabilities of these constrained devices in Fig. 1 are classified as ”deep models,” while those within the device’s capacity are considered more light. This classification helps identify suitable algorithms for deployment on specific edge devices based on their computational needs. In this study, we consider ViT as deep model because of the quadratic computation cost.

B. Choice of Model

The edge models selected for investigation in this study include details such as the number of parameters and MAC operations. The choice of these edge models was informed by previous studies focused on deployment to edge devices. Evaluated within this study are the deep models presented in Table I. These deep models were previously utilized in EMG-based gesture classification studies, achieving state-of-the-art (SOTA) performance at the time of their respective publications. For fair comparison of the chosen models, we did not utilize pre-trained weights.

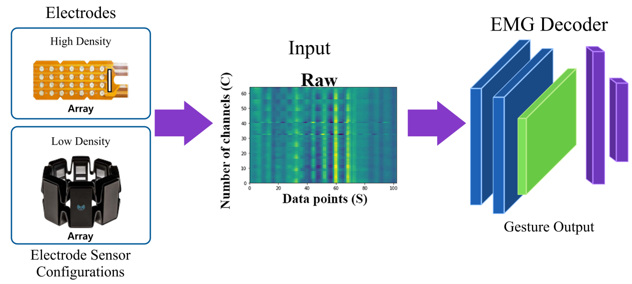

Fig. 2. Surface EMG-based Gesture Decoding Pipeline.

C. Dataset

In this study, we investigate the decoding performance on both low- and high-density electromyography (EMG). The high spatial selectivity provided by high-density EMG (HD-sEMG) suggests better decoding performance compared to its low-density counterpart. An illustration of the dataset and the decoding pipeline utilized in this study is provided in Fig. 2. The dataset is briefly described as follows:

1) Ninapro DB5: : The dataset comprises data acquired from 10 healthy participants performing 52 isometric and isotonic tasks of the upper limb using two Myo armbands (with a sampling frequency of 200 Hz). In this study, we refer to Ninapro DB5 as a low-density dataset, abbreviated as DB1.

- **Ninapro DB5**: The dataset comprises data acquired from 10 healthy participants performing 52 isometric and isotonic tasks of the upper limb using two Myo armbands (with a sampling frequency of 200 Hz). In this study, we refer to Ninapro DB5 as a low-density dataset, abbreviated as DB1.
- **MyoArmband Dataset [11]**: The data was acquired from 28 male and 12 female participants using the Myo Armbands from Ninapro DB5. In this study, we refer to the Myo Armband Dataset as DB2.
- **Hyser Dataset [12]**: The Hyser dataset was acquired using a 16 × 8 electrode array, resulting in 256 channels, with a sampling frequency of 2048 Hz. The dataset includes pattern recognition, maximum voluntary contraction, one degree of freedom (DoF), n-DOF, and random tasks. However, this study utilizes only the pattern recognition dataset, referred to as DB3. The dataset comprises 34 types of gestures acquired from 20 participants.
- **FlexEMG Dataset [13]**: A custom-designed 64-channel HD-sEMG system was used to acquire EMG signals from five subjects performing 13 static gestures. The HD-sEMG platform referred to as DB4 in this study operates at a sampling frequency of 1000 Hz.

TABLE II
HYPERPARAMETER SETTINGS USED IN THE STUDY

Hyperparameter	Value
Learning Rate	0.001
Batch Size	32
Optimizer	Adam
Loss Function	Cross-Entropy
Number of Epochs	30
Dropout Rate	0.25

D. Preprocessing

We used traditional preprocessing pipelines from EMG studies as shown in Fig. 2. The datasets underwent notching, bandpass filtering, and division into segments with overlapping windows. DB1 and DB2 underwent notch filtering at 50.0Hz/60.0Hz followed by bandpass filtering using a 5th-order Butterworth filter with cutoff frequencies of 5.0Hz–99.00Hz. DB3 and DB4 underwent bandpass filtering with cutoff frequencies of 10Hz–4500Hz.

A window size of 200 ms was used for both DB1 and DB2 to maintain consistency with established preprocessing methods employed in similar EMG studies [14]. This choice ensures that the data acquisition process does not affect the upper limit of the electro-mechanical delay. For high-density datasets (DB3 and DB4), smaller window sizes were utilized: 32 ms for DB3 and 64 ms for DB4. This adjustment was made to ensure consistent data sampling within windows, particularly when comparing low-density and high-density EMG datasets. Khushaba and Nazarpour [10], reported a window size of 32 ms as been appropriate for HD-sEMG. An overlap of 60% was chosen for all datasets to maintain sufficient data consistency across preprocessing steps.

E. Training and Evaluation

We conducted model training on two desktop computers equipped with NVIDIA GeForce RTX 3090 and RTX 4090, respectively. To ensure fair comparison between the models, we performed preliminary testing on each dataset to select global hyperparameters reported in Table II for training both the edge and deep models. This step is particularly crucial for ensuring reproducibility, achieved alongside setting appropriate random numbers.

As proposed by authors in [3], we adopted train-test splits for time series (TSTS) methods of evaluation. TSTS reflects real-world settings where preceding data is not used for evaluation within the same trial. For this approach, we utilized 80% of the trials as training data and reserved the remaining 20% as test data from each respective dataset. The performance of the EMG-to-gesture decoders was evaluated using classification accuracy (CA) (1).

$$\mathrm{CA} = \left(\frac{\text{Number of correct predictions}}{\text{Total number of true events}}\right) \times 100 \qquad (1)$$

We carried out the Shapiro-Wilk test to assess both the homogeneity and normality of the results. Subsequently, we performed the Two-Way ANOVA test to examine the impact of model types and dataset types on CA while also investigating the interaction effects between model type and dataset type. Additionally, we conducted Paired t-Test tests for each dataset to determine whether there were significant differences in performance between edge and deep models. We set the significance level at $\alpha = 0.05$. All statistical analysis were done in R.

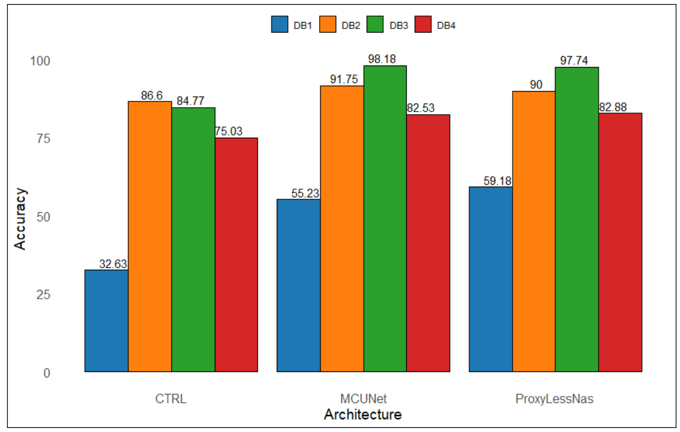

Fig. 3. Average Classification of Edge Models Accuracy Across all Datasets

III. RESULTS AND DISCUSSION

A. Performance of Edge Model

The Edge Model, as shown in Fig. 3, achieved competitive CA across all datasets except DB1, where performance was notably lower. The reduced accuracy in DB1 can be attributed to the high number of gestures (52) and the low sampling rate of the MyoBand device, which impacts data acquisition efficiency. To address this issue, we increased the training epochs and adjusted the window size during segmentation to 400 ms, 600 ms, 800 ms, and 1 second. These changes did not result in a significant reduction in classification error. Prior work [3] reported a CA of 42.1% for DB1 using a pretrained ResNet18 model without fine-tuning. The best-performing model within our edge models for DB1, ProxyLessNAS, achieved a 14.08% improvement over theirs, despite having fewer parameters (2.9 million compared to ResNet18's 11 million). Additionally, DB2 demonstrated superior performance due to its reduced gesture count, resulting in higher CA. Across all datasets, the edge Model maintained an average CA of 78.04%. We excluded ProxyLessNAS from our analysis since it exhibited similar parameters and MAC values (average accuracy of 75.84%). This decision was to allow for non-bias paired t-test statistical analysis.

B. Performance of the Deep Model

The ResNet18, characterized by its significantly high number of trainable parameters, outperformed ViT on DB3 (achieving a 10.57% improvement) and DB4 (a 11.40% improvement). Conversely, ViT demonstrated superior performance on array-based low-density EMG. Our results from Fig. 4 suggest that deep networks can be advantageous for the electrode-based EMG with high spatial selectivity and volume conduction capabilities. However, this observation

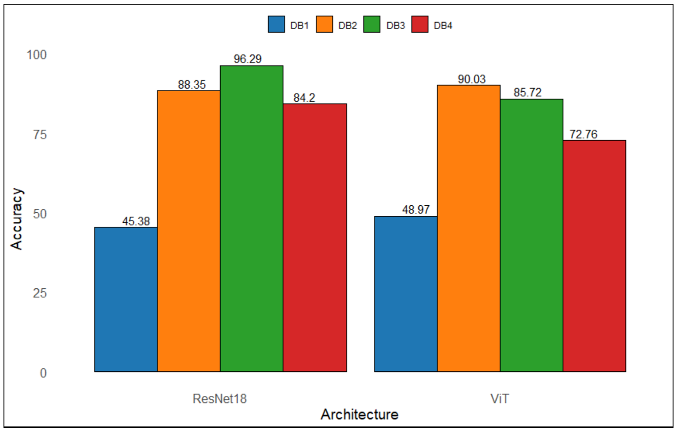

Fig. 4. Average Classification of Deep Models Accuracy Across all Datasets

does not represent a universal law. We found that edge models with sufficient capacity can achieve high CA, as evidenced by ProxyLessNAS and MCUNet outperforming ResNet18 by 1.45% and 1.89%, respectively.

TABLE III
TRAINING TIME (SECONDS/EPOCH) AND POWER CONSUMPTION (WATTS/EPOCH) FOR EACH MODEL ACROSS ALL DATASET

Models	Training Time	Power Consumption
MCUNet	11.75	192.15
CRTL	1.775	146.84
ProxyLessNas	8.69	172.60
ViT	2.87	153.22
ResNet18	5.06	185.34

C. Comparison between Edge and Deep Model

We conducted Two-Way ANOVA to examine the impact of model types (Edge vs. Deep) and dataset types on CA. The analysis yielded an F-value of 0.203 with a p-value of 0.660 for the comparison between edge and deep models, indicating no significant difference in performance between these two model types. Furthermore, across all datasets analyzed, our study found that the p-values were consistently less than 0.05, suggesting that dataset characteristics significantly influence CA. However, when assessing interaction effects between model type and dataset type, a p-value of 0.996 was obtained, indicating no significant interaction effect. These findings lead to the conclusion that the effect of model type on CA is independent of the dataset type. This aligns with the initial hypothesis, which posits that edge models are a viable alternative compared to computation and memory-intensive deep models, as they achieve similar classification accuracies without the associated computational overhead.

We further performed a paired t-test to determine whether there was a statistically significant difference in accuracy between edge models and deep models. The results showed that, across all four datasets (DB1, DB2, DB3, DB4), the p-values were greater than 0.05, indicating no statistically significant difference in performance between edge models and deep models. Based on these findings, we infer that model type does not have a significant impact on classification accuracy for low-density and high-density EMG data. However, it is noteworthy that edge models offer practical advantages in terms of CA, computational efficiency, fast training times, and lower memory utilization, as evidenced by the results presented in Table III. The high training time observed for MCUNet and ProxyLessNAS was attributed to an outlier encountered during the training process on DB3. Excluding DB3 from the analysis yielded improved performance metrics: MCUNet achieved 3.93 seconds per epoch and consumed 167.42 watts per epoch, while ProxyLessNAS recorded lower values with 3.62 seconds per epoch and 166.52 watts per epoch in terms of power consumption across the RAM, CPU, and GPU. We considered raw input for gesture decoding in this study; however, we plan to investigate other feature extraction techniques in the future.

IV. CONCLUSION

In this study, we evaluated the benefits of edge models compared to deep models in sEMG-based gesture recognition. Through extensive experiments using efficient edge models and deep models across both low-density and high-density EMG datasets, our results indicated that edge models offer computational efficiency, faster training times, and lower memory utilization with minimal accuracy loss compared to Deep models. This study serves as a wake-up call for researchers and industry experts leveraging EMG to develop active rehabilitation devices or human-computer interaction applications. We urge them to prioritize the development of lightweight models that enable edge deployability over deep models without clear-cut benefits, especially when computational resources are limited.

REFERENCES

[1] B. Plancher and V. Janapa Reddi, "Tinymledu: The tiny machine learning open education initiative," in *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*, 2022, pp. 1159–1159.
[2] A. Holobar and D. Farina, "Noninvasive neural interfacing with wearable muscle sensors: Combining convolutive blind source separation methods and deep learning techniques for neural decoding," *IEEE signal processing magazine*, vol. 38, no. 4, pp. 103–118, 2021.
[3] J. Yang, M. J. Soh, V. Lieu, D. J. Weber, and Z. Erickson, "Emgbench: Benchmarking out-of-distribution generalization and adaptation for electromyography," in *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024.
[4] K. R. Pyun, K. Kwon, M. J. Yoo, K. K. Kim, D. Gong, W.-H. Yeo, S. Han, and S. H. Ko, "Machine-learned wearable sensors for real-time hand-motion recognition: toward practical applications," *National science review*, vol. 11, no. 2, p. nwad298, 2024.
[5] J. Lin, W.-M. Chen, Y. Lin, C. Gan, S. Han *et al.*, "Mcunet: Tiny deep learning on iot devices," *Advances in neural information processing systems*, vol. 33, pp. 11 711–11 722, 2020.
[6] H. Cai, L. Zhu, and S. Han, "Proxylessnas: Direct neural architecture search on target task and hardware," in *International Conference on Learning Representations*, 2018.
[7] C. labs at Reality Labs, D. Sussillo, P. Kaifosh, and T. Reardon, "A generic noninvasive neuromotor interface for human-computer interaction," *bioRxiv*, pp. 2024–02, 2024.
[8] M. D. Dere and B. Lee, "A novel approach to surface emg-based gesture classification using a vision transformer integrated with convolutive blind source separation," *IEEE Journal of Biomedical and Health Informatics*, 2023.
[9] C. R. Carvalho, J. M. Fernández, A. J. Del-Ama, F. Oliveira Barroso, and J. C. Moreno, "Review of electromyography onset detection methods for real-time control of robotic exoskeletons," *Journal of neuroengineering and rehabilitation*, vol. 20, no. 1, p. 141, 2023.

[10] R. N. Khushaba and K. Nazarpour, "Decoding hd-emg signals for myoelectric control-how small can the analysis window size be?" *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8569–8574, 2021.

[11] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 4, pp. 760–771, 2019.

[12] X. Jiang, X. Liu, J. Fan, X. Ye, C. Dai, E. A. Clancy, M. Akay, and W. Chen, "Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1035–1046, 2021.

[13] A. Moin, A. Zhou, A. Rahimi, A. Menon, S. Benatti, G. Alexandrov, S. Tamakloe, J. Ting, N. Yamamoto, Y. Khan *et al.*, "A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition," *Nature Electronics*, vol. 4, no. 1, pp. 54–63, 2021.

[14] M. D. Dere, J.-H. Jo, and B. Lee, "Event-driven edge deep learning decoder for real-time gesture classification and neuro-inspired rehabilitation device control," *IEEE Transactions on Instrumentation and Measurement*, 2023.