Lightweight Deep Learning Models for sEMG-based Hand Gesture Classification on Low-Resource Devices

Israel Huentecura Rodríguez
Universidad Mayor
Santiago, Chile

Anthony D. Cho Lo, PhD
Universidad Mayor
Santiago, Chile

***Abstract*—Advanced myoelectric hand prostheses face significant real-world adoption challenges due to their prohibitive cost and high abandonment rates, often linked to control complexity and lack of functionality. This work, framed within the ProtoIA open-source project, addresses this issue by developing high-performance, computationally efficient neural network models for hand gesture classification from surface electromyography (sEMG) signals, specifically designed for low-resource hardware. Using the NinaPro DB1 dataset, we compare two literature-based models against four custom-designed deep learning architectures. Our experimental results show that all proposed models achieve over 98% classification accuracy. Notably, models like EMGHandNet-2D reached 99.78% accuracy while meeting the strict size constraint of less than 1.5 MB for microcontroller deployment. This study concludes that lightweight deep learning architectures are a viable and powerful solution for prosthetic control on affordable hardware, and by publishing the code and pre-trained models, we provide a tangible contribution to the democratization of assistive technologies.**

***Index Terms*—sEMG, deep learning, hand gesture recognition, prosthetics, edge AI, embedded systems, lightweight models**

I. Introduction

The amputation of a limb profoundly impacts an individual's daily life [1]. Advanced myoelectric prostheses, which use surface electromyography (sEMG) to interpret user intent from muscular activity [2], offer a path to restoring functionality. However, this technology is hindered by a dual challenge: accessibility and usability. Commercial prostheses are prohibitively expensive, with costs exceeding €54,000 for advanced models, a price that does not always translate to a better user experience [3]. Furthermore, systematic reviews indicate abandonment rates as high as 23% among adults, citing lack of functionality and control complexity as primary reasons [4].

In response, a global open-source movement aims to democratize assistive technologies [5]. This philosophy of collaborative and transparent development is the cornerstone of the *ProtoIA* project, which seeks to create a low-cost, open-source prosthetic hand. For such a prosthesis to be effective, its control model must be both accurate and efficient enough to run on affordable hardware. This leads to the domain of Edge AI, where algorithms are executed directly on the device [6]. This approach mitigates latency and dependence on internet connectivity, which are critical limitations for a prosthesis.

This work addresses the challenge of creating a deep learning model for gesture classification that is both highly accurate and compact enough for deployment on a low-cost microcontroller (ESP32). The guiding research question is: *Can a deep neural network architecture, with a deployment size under 1.5 MB, achieve over 95% average accuracy in classifying 12 hand gestures from the NinaPro DB1 dataset?*. A positive answer would provide a validated, high-performance software foundation for the next generation of open-source prostheses, with the pre-trained models released to the community to facilitate fine-tuning and personalization [7].

II. Methodology

A. Dataset and Preprocessing

This study utilized the NinaPro Database 1 (DB1) [2], focusing on "Exercise A," which consists of 12 finger movements compatible with the *ProtoIA* prosthesis design. The dataset contains sEMG signals from 27 able-bodied subjects, recorded from 10 bipolar electrodes at 100 Hz [2].

The preprocessing pipeline for each subject's data involved:

1) **Filtering:** A first-order Butterworth low-pass filter with a 1 Hz cutoff frequency was applied to remove baseline drift and low-frequency noise [8].
2) **Compression:** A μ-law companding algorithm ($\mu = 256$) was used to compress the dynamic range of the signals, enhancing the resolution of low-amplitude components.
3) **Segmentation:** Signals were segmented into 200 ms windows (20 samples) with a 19-sample overlap. A window was considered valid only if all its samples belonged to the same gesture class.
4) **Sequencing:** Valid windows were grouped into non-overlapping sequences of 5, representing 1 second of gestural activity.

B. Feature Extraction

For hybrid models like *DualStream-Lite* and *HyT-Net*, a set of 10 time-domain features was extracted from each 200 ms window for each of the 10 channels. This set was inspired by the work of Hudgins et al. [9] and extended with additional descriptors. The features included statistical measures like Standard Deviation (SD) and Root Mean Square (RMS), signal characteristics such as Zero Crossings (ZC) and Mean Absolute Value (MAV), and custom features like the maximum value in the first temporal quarter of the window.

C. Model Architectures

To perform a comprehensive analysis, six architectures were evaluated. This included two established models from the literature as benchmarks, and four novel lightweight architectures designed specifically for this study. All models were implemented in TensorFlow/Keras.

1) Reference Architectures:

- ***EMGHandNet-Original*:** Proposed by Karnam et al. [10], this model combines CNN layers for spatial feature extraction across sEMG channels with a Bi-LSTM layer to capture temporal dependencies.
- ***DualStream-Original*:** Developed by Zhang et al. [11], this architecture uses two parallel streams: one processes raw sEMG with a 1D-CNN, while the other processes handcrafted time-domain features. The outputs are then fused and fed to a Bi-LSTM.

2) Proposed Lightweight Architectures:

- ***DualStream-Lite*:** A simplified version of *DualStream-Original*, using shallower networks and a single Bi-LSTM layer to reduce complexity.
- ***EMGHandNet-2D*:** Inspired by *EMGHandNet-Original*, this model treats the multi-channel sEMG input as a 2D image, using `Conv2D` layers to learn spatio-temporal patterns efficiently before a final Bi-LSTM layer.
- ***CRNN-Attn*:** A novel, efficiency-focused model combining depthwise separable convolutions with a lightweight Bi-GRU recurrent layer and an attention mechanism to focus on relevant temporal features.
- ***HyT-Net*:** A novel Hybrid Transformer Network that replaces recurrence with a Transformer encoder, using a multi-head self-attention mechanism [12] to model global dependencies in the signal sequence.

All developed models and their weights are available in a public repository [13].

D. Experimental Protocol

A rigorous evaluation protocol was established. The full dataset was first split into a development set (80%) and a hold-out test set (20%) using stratified sampling. A 10-fold stratified cross-validation was performed on the development set. In each fold, models were trained for up to 100 epochs using the Adam optimizer. Callbacks were used for early stopping (patience of 15 epochs on validation loss), learning rate reduction on plateau, and saving the best model based on validation accuracy. The final performance was reported as the mean and standard deviation of accuracy across the 10 saved models when evaluated on the unseen test set.

III. Results

The evaluation focused on two key metrics: classification accuracy and model size, to determine the feasibility for on-device deployment.

TABLE I
MODEL PERFORMANCE (MEAN ± STD. DEV. ACCURACY)

Model	Train Acc	Val Acc	Test Acc
HyT-Net	0.9998 ± 0.0000	0.9999 ± 0.0000	0.9983 ± 0.0002
DualStream-Orig.	0.9998 ± 0.0001	0.9999 ± 0.0000	0.9982 ± 0.0003
EMGHandNet-Orig.	0.9993 ± 0.0001	0.9997 ± 0.0001	0.9980 ± 0.0002
EMGHandNet-2D	0.9993 ± 0.0001	0.9997 ± 0.0001	0.9978 ± 0.0003
DualStream-Lite	0.9975 ± 0.0003	0.9988 ± 0.0003	0.9941 ± 0.0003
CRNN-Attn	0.9920 ± 0.0011	0.9961 ± 0.0008	0.9855 ± 0.0015

TABLE II
MODEL SIZE COMPARISON

Model	Parameters	TFLite Size (MB)
DualStream-Lite	274,700	1.01
CRNN-Attn	291,021	1.17
EMGHandNet-2D	341,516	1.35
HyT-Net	846,348	3.34
EMGHandNet-Orig.	1,637,132	6.44
DualStream-Orig.	3,039,148	11.93

A. Classification Performance

All six evaluated architectures successfully surpassed the 95% accuracy target on the test set. The results, summarized in Table I, show two distinct performance tiers.

The top-tier models (*HyT-Net*, *DualStream-Original*, *EMGHandNet-Original*, and *EMGHandNet-2D*) achieved nearly flawless classification, with test accuracies around 99.8%. The proposed *HyT-Net* and *EMGHandNet-2D* architectures performed on par with the more complex reference models. The ”efficient” tier models, *DualStream-Lite* and *CRNN-Attn*, also demonstrated excellent performance, albeit slightly lower. An analysis of the confusion matrix for *DualStream-Lite* revealed that most of its errors were concentrated in distinguishing between subtle thumb movements, such as ”Thumb Adduction” and ”Thumb Extension”.

B. Model Size and Complexity

The second objective was to produce a model smaller than 1.5 MB to ensure it could be deployed on a microcontroller. Table II compares the number of parameters and the final TFLite file size for each architecture.

Three of the proposed architectures successfully met the size constraint: *DualStream-Lite* (1.01 MB), *CRNN-Attn* (1.17 MB), and *EMGHandNet-2D* (1.35 MB). This result validates their architectural design for memory-constrained environments. In contrast, the reference models and *HyT-Net* were significantly larger, making them unsuitable for the target hardware.

IV. Discussion

The results confirm that it is possible to achieve an exceptional balance between high classification accuracy and the efficiency needed for deployment on low-cost hardware. The choice of the optimal model, however, depends on the target deployment scenario.

A. Scenario 1: Single-Board Computer (e.g., Raspberry Pi)

For a platform with more generous resources like a Raspberry Pi, maximizing accuracy is the priority. In this context, ***HyT-Net*** is the superior choice. It matches the state-of-the-art accuracy of the reference models but with a dramatically more efficient parameterization, being 3.5 times smaller than *DualStream-Original*.

B. Scenario 2: Microcontroller (e.g., ESP32)

This is the core scenario for the *ProtoIA* project, where memory constraints are paramount. Even with quantization (converting weights from FP32 to INT8 to reduce size by 75%) [14], the larger models are infeasible as their memory footprint would exceed the available SRAM. The choice is between the three models that met the size requirement:

- *DualStream-Lite* is the most memory-frugal option, making it ideal for the most constrained systems.
- *CRNN-Attn* is discarded due to its higher performance instability (larger standard deviation).
- ***EMGHandNet-2D*** emerges as the most balanced and promising candidate. It achieves the highest accuracy (99.78%) among the lightweight group, nearly matching the top-tier models. Its end-to-end nature, which does not require manual feature extraction, simplifies the inference pipeline and reduces CPU load, a critical advantage on a microcontroller.

Therefore, *EMGHandNet-2D* is the recommended model for the low-cost, embedded version of the prosthesis.

V. Conclusion and Future Work

This study has successfully demonstrated that lightweight deep learning architectures can achieve state-of-the-art accuracy for sEMG-based gesture classification while adhering to the strict size constraints of low-resource hardware. We identified ***EMGHandNet-2D*** as the most suitable model for a microcontroller-based prosthesis, offering the best trade-off between its 99.78% accuracy and a 1.35 MB deployment size. The proposed ***HyT-Net*** model stands as the best option for less constrained systems like a Raspberry Pi.

By releasing the models and source code, we provide a validated, high-performance foundation for the open-source community, enabling a practical path for the personalization of prosthetic controls via fine-tuning.

Future work will focus on:

1) **Hardware Validation:** Deploying the quantized *EMGHandNet-2D* and *DualStream-Lite* models on an ESP32 to measure real-world inference speed, RAM usage, and power consumption.
2) **Gesture Expansion:** Evaluating the architectures on more complex gesture sets (Exercises B and C) from the NinaPro dataset.
3) **Clinical Validation:** Conducting a pilot study to validate the fine-tuning of the *EMGHandNet-2D* model on data from users with amputations to confirm its adaptability.

Acknowledgment

The authors wish to thank the members of the ProtoIA project for their invaluable collaboration and insights. This research was supported by the academic environment and resources provided by Universidad Mayor. The guidance of Dr. Anthony D. Cho Lo was instrumental in the development of this study.

References

[1] C. L. McDonald, S. Westcott-McCoy, M. R. Weaver, J. Haagsma, and D. Kartin, "Global prevalence of traumatic non-fatal limb amputation," *Prosthetics and orthotics international*, vol. 45, no. 2, pp. 105–114, 2021.

[2] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific data*, vol. 1, no. 1, pp. 1–13, 2014.

[3] N. Kerver, E. Karssies, P. F. Krabbe, C. K. van der Sluis, and H. Groen, "Economic evaluation of upper limb prostheses in the netherlands including the cost-effectiveness of multi-grip versus standard myoelectric hand prostheses," *Disability and Rehabilitation*, vol. 45, no. 25, pp. 4311–4321, 2023.

[4] P. Capsi-Morales, C. Piazza, L. Sjoberg, M. G. Catalano, G. Grioli, A. Bicchi, and L. M. Hermansson, "Functional assessment of current upper limb prostheses: An integrated clinical and technological perspective," *Plos one*, vol. 18, no. 8, p. e0289978, 2023.

[5] E. S. Raymond, *The Cathedral the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly Media, 1999, este libro documenta la 'Ley de Linus' y atribuye la cita 'Given enough eyeballs, all bugs are shallow' a Linus Torvalds.

[6] X. Wang and W. Jia, "Optimizing edge ai: A comprehensive survey on data, model, and system strategies," *arXiv preprint arXiv:2501.03265*, 2025.

[7] 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.

[8] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng, "A multi-stream convolutional neural network for semg-based gesture recognition in muscle-computer interface," *Pattern Recognition Letters*, vol. 119, 2019.

[9] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE transactions on biomedical engineering*, vol. 40, no. 1, pp. 82–94, 1993.

[10] N. K. Karnam, S. R. Dubey, A. C. Turlapaty, and B. Gokaraju, "Emghandnet: A hybrid cnn and bi-lstm architecture for hand activity classification using surface emg signals," *Biocybernetics and biomedical engineering*, vol. 42, no. 1, pp. 325–340, 2022.

[11] K. Zhang, F. J. Badesa, Y. Liu, and M. Ferre Pérez, "Dual stream long short-term memory feature fusion classifier for surface electromyography gesture recognition," *Sensors*, vol. 24, no. 11, p. 3631, 2024.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 5998–6008. [Online]. Available: http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

[13] I. Huentecura, "openprosthesis-emg-protoIA: Modelos abiertos de IA para prótesis de mano," May 2024. [Online]. Available: https://github.com/IsraelHuentecura/openprosthesis-emg-protoIA

[14] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2704–2713.