Transfer Learning with ResNet-18 for Continuous Wavelet Transform (CWT) Based EMG Gesture Classification Under Leave-One-Subject-Out Evaluation

Author: Suleman Khan
AI-Robotics, KIST School, University of Science & Technology (UST), Seoul, Republic of Korea
Student Researcher, Intelligence and Interaction Research Center, Korea Institute of Science and Technology (KIST)
Academic Advisor: Dr. Kyung-Ryoul Mun
Associate Professor, University of Science and Technology (UST)
Principal Researcher, Intelligence and Interaction Research Center, Korea Institute of Science and Technology (KIST)

Course Instructor (Fundamentals of Deep Learning and PyTorch Programming): Dr. Donghoon Kang
Adjunct Professor, University of Science and Technology (UST)
Senior Researcher, Intelligence and Interaction Research Center, Korea Institute of Science and Technology (KIST)

Abstract

Surface electromyography (sEMG) provides a non-invasive way to monitor muscle activity and is widely used in gesture recognition and human–machine interaction. However, sEMG-based classifiers often struggle to generalize across subjects due to variations in physiology, electrode placement, and signal characteristics. This work examines whether transfer learning can improve subject-independent gesture classification when using Continuous Wavelet Transform (CWT) images and a ResNet-18 architecture.

Two training strategies were compared under a Leave-One-Subject-Out (LOSO) protocol: (1) training the model from scratch, and (2) a transfer-learning approach that first pretrains on a subset of the training subjects and then finetunes on the remaining data. Experiments were conducted for two LOSO folds (LOSO-1 and LOSO-2). Confusion matrices, averaged CWT images, and sample-level visualizations were used to assess model behavior and gesture separability.

Across both folds, transfer learning improved performance on unseen subjects and reduced confusion between gestures with similar activation patterns. The pretrained models learned more stable spectral–temporal representations, which made finetuning more effective and enhanced generalization.

1. Introduction

Surface electromyography (sEMG) provides a non-invasive measure of neuromuscular activation and is widely used in rehabilitation robotics, prosthetic control, and gesture-based human–machine interfaces. Although sEMG-based gesture classification has been studied extensively, achieving reliable performance across different users remains difficult. Physiological differences, variations in muscle anatomy, electrode placement, and skin impedance all contribute to subject-specific signal patterns, making cross-subject generalization a central challenge.

Deep learning models have improved EMG classification by learning features directly from data rather than relying on handcrafted descriptors. When sEMG signals are transformed into two-dimensional representations such as Continuous Wavelet Transform (CWT) images, convolutional neural networks (CNNs) can extract meaningful temporal–frequency structures. ResNet-18, in particular, is attractive for this task because its residual connections stabilize training and are effective for medium-sized biomedical datasets.

However, models trained on one group of subjects often perform poorly when evaluated on unseen individuals. For this reason, Leave-One-Subject-Out (LOSO) evaluation is commonly used to measure subject-independent performance. To address the generalization issue, transfer learning has been proposed as a way to leverage knowledge learned from a subset of subjects before adapting the model to new ones.

In this work, we compare two strategies using CWT-based ResNet-18 models across two LOSO folds (LOSO-1 and LOSO-2):

- **Baseline training:** the model is trained from scratch using only the available training subjects.
- **Transfer learning:** the model undergoes a pretraining stage on a subset of subjects (20% of training data), followed by finetuning on the remaining subjects.

Performance is analyzed using confusion matrices and qualitative visualizations, including averaged gesture images and sample-level CWT outputs. The results show that transfer learning improves robustness to subject variability and leads to better generalization on unseen users.

2. Methods

2.1 Dataset

The dataset contains multi-channel surface EMG recordings from several subjects performing nine gesture classes. Each gesture trial consists of short EMG segments sampled from forearm electrodes. For model compatibility, each segment was converted into a Continuous Wavelet Transform (CWT) image.

Gesture classes were treated as **numeric labels (0–8)**, following the dataset's original annotation. No anatomical gesture naming was used in training or evaluation period.

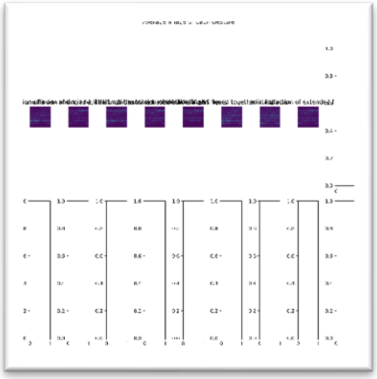
2.2 CWT Preprocessing

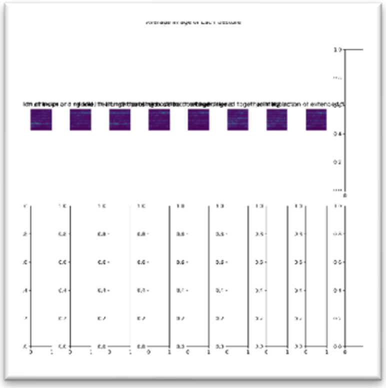
To obtain a time–frequency representation suitable for convolutional networks, each EMG segment was transformed using the Morlet Continuous Wavelet Transform. This produces a 2-D map where:

- the horizontal axis represents time,

- the vertical axis represents scale (related to frequency), and
- pixel intensities reflect local EMG energy.

CWT images were resized to the ResNet-18 input resolution and normalized across the dataset. For qualitative inspection, average CWT images were computed for the training, validation, and test sets under the LOSO-1 fold (Figures 1–3).





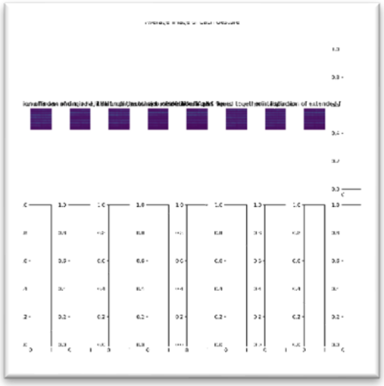


Figure 1. Average CWT image for training set (LOSO-1 baseline).
Figure 2. Average CWT image for validation set (LOSO-1 baseline).
Figure 3. Average CWT image for test set (left-out LOSO-1 subject).

2.3 Model Architecture: ResNet-18

ResNet-18 was used due to its stable optimization behavior and suitability for medium-sized biomedical datasets. The architecture includes:

- initial convolution, BatchNorm, and ReLU,
- four residual stages with skip connections,
- global average pooling,
- a fully connected output layer producing 9 class logits.

Residual connections help maintain gradient flow during training, which is beneficial when learning spectral–temporal structures from CWT images.

2.4 Training Procedure

Two training strategies were evaluated:

(A) Baseline Training (No Transfer Learning)

- Model initialized with random weights
- Trained on all available training subjects
- Evaluated directly on the left-out subject
- **30 training epochs**

(B) Transfer Learning (Pretrain → Finetune)

- Pretraining on **20% of the training subjects**
- The pretrained weights were then finetuned on the remaining subjects
- Model evaluated on the left-out subject
- **30 pretraining epochs + finetuning**

Both approaches used:

- **Optimizer:** Adam
- **Loss:** Cross-Entropy
- **Inputs:** 2-D CWT images

2.5 Leave-One-Subject-Out (LOSO) Evaluation

Generalization to unseen individuals was assessed using LOSO evaluation:

- **LOSO-1:** Subject 1 used as test subject
- **LOSO-2:** Subject 2 used as test subject

For each fold:

1. All other subjects were used for training.
2. A portion of the training data was used as validation.
3. The left-out subject served as the final test set.

This evaluation setup reflects real-world usage, where a trained EMG model must generalize to new users without retraining from scratch.

3. Results

Experiments were conducted under four configurations:
(1) LOSO-1 baseline, (2) LOSO-1 transfer learning, (3) LOSO-2 baseline, and (4) LOSO-2 transfer learning.
Performance was examined using averaged CWT gesture maps, sample-level visualizations, and confusion matrices.

3.1 LOSO-1 → Baseline Training (No Transfer Learning)

3.1.1 CWT Averages

The CWT averages for the training, validation, and test sets were already introduced in Section 2.2. These images illustrate the general spectral–temporal structure of the dataset and highlight differences between training and test subjects. They are not repeated here.

3.1.2 First Fifteen Sample Images (Qualitative Check)

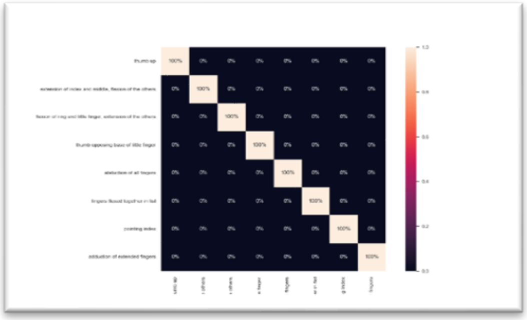
Figure 4 shows the first fifteen CWT images from the LOSO-1 test subject. These samples illustrate typical intra-class variation and subject-specific signal characteristics.

First Fifteen Test Images of Each Gesture
thumb up
extension of index and middle
flexion of ring and little finger, extension
thumb opposing base of little finger
abduction of all fingers
fingers flexed together in fist
pointing index
adduction of extended fingers

Figure 4. First fifteen CWT images from the LOSO-1 test subject (baseline).

3.1.3 Confusion Matrices

Figures 5–7 present the confusion matrices for the training, validation, and test sets. The baseline model shows strong performance on the training subjects but reduced accuracy on the held-out subject, reflecting limited cross-subject generalization.



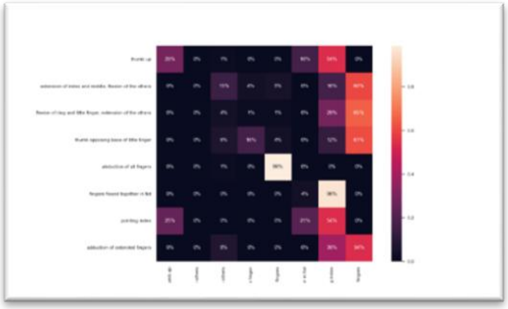


Figure 5. Training confusion matrix for LOSO-1 baseline.
Figure 6. Validation confusion matrix for LOSO-1 baseline.

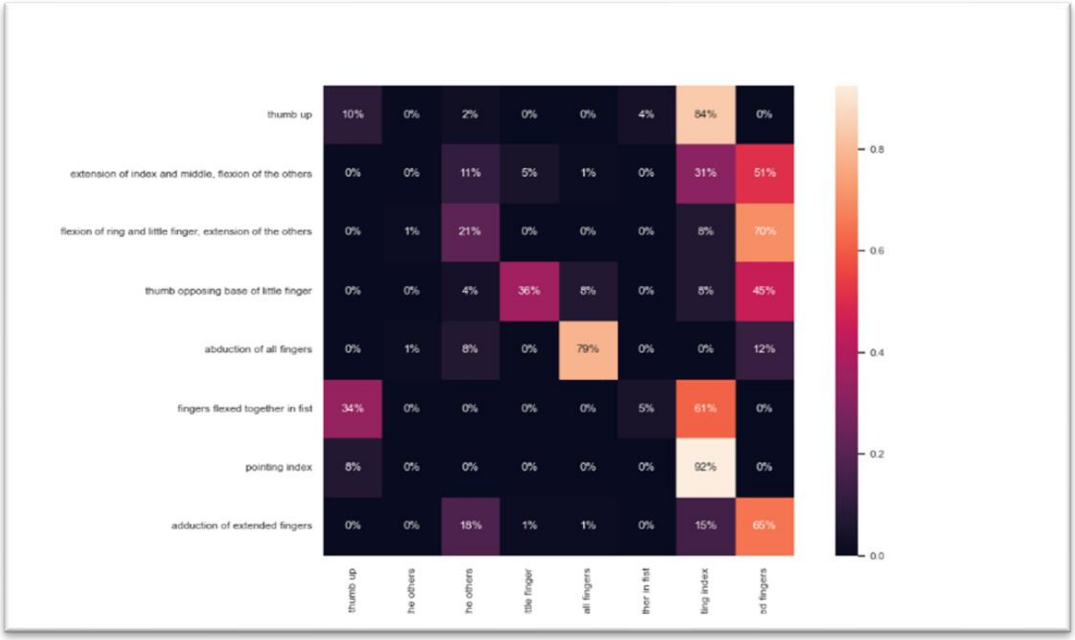
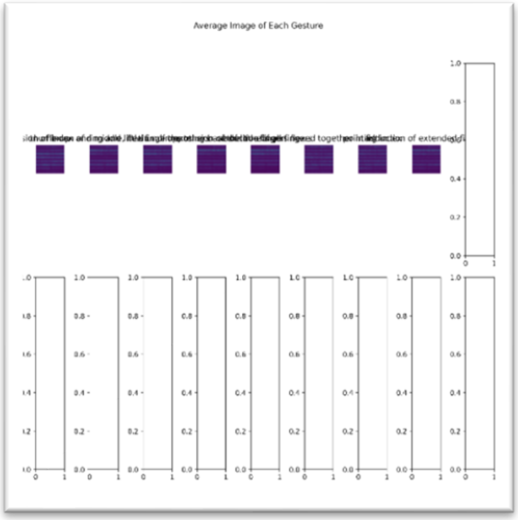


Figure 7. Test confusion matrix for LOSO-1 baseline.

3.2 LOSO-1 → Transfer Learning (Pretrain + Finetune)

3.2.1 CWT Averages

Figures 8–11 present the average CWT images for the pretraining, finetuning, validation, and test sets.
Compared to the baseline averages, these maps show clearer gesture-specific structures, suggesting improved feature consistency after transfer learning.



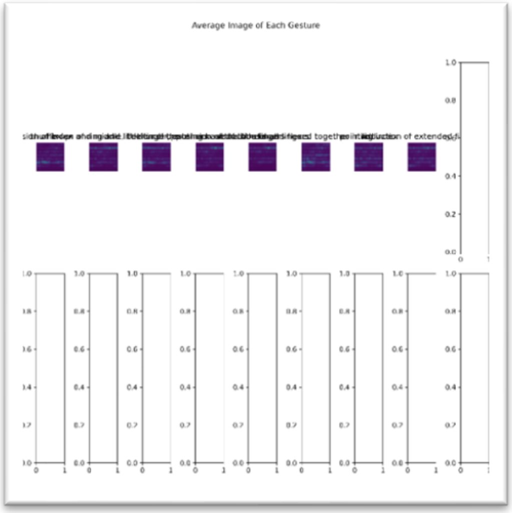
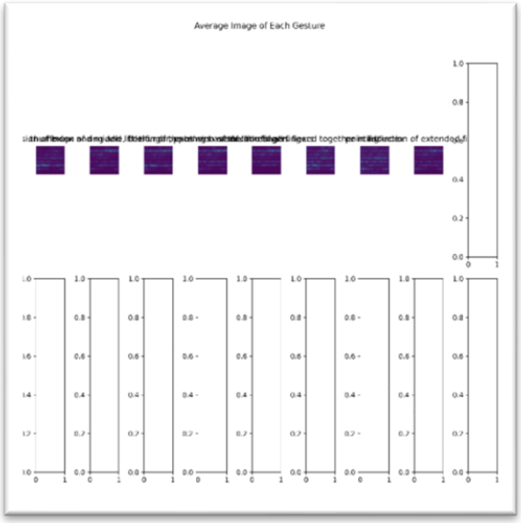


Figure 8. Average CWT images for LOSO-1 pretraining stage.
Figure 9. Average CWT images for LOSO-1 finetuning stage



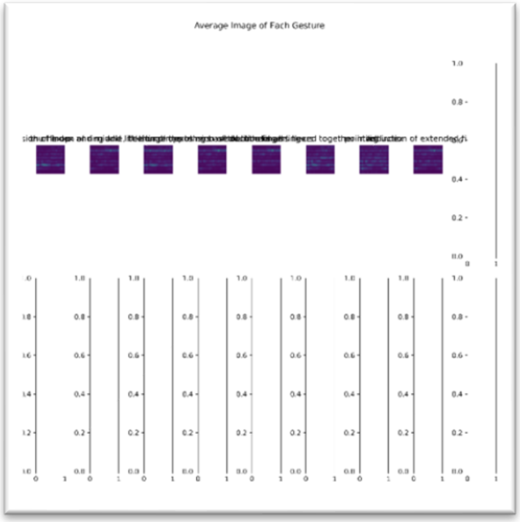
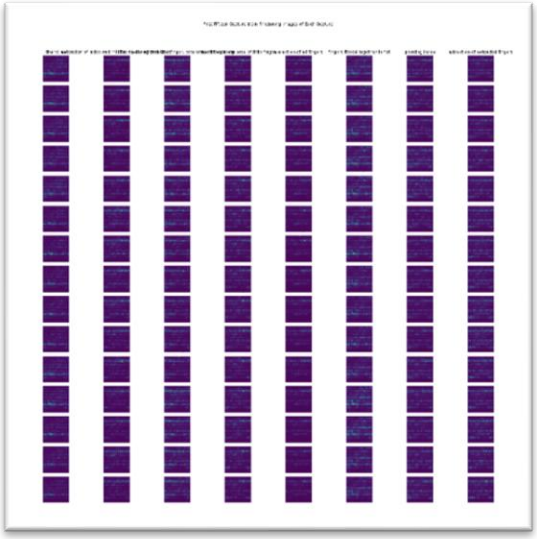


Figure 10. Average CWT images for LOSO-1 validation set.
Figure 11. Average CWT images for LOSO-1 test subject.

3.2.2 First Fifteen Samples

Figure 12 shows the first fifteen finetuning samples, and Figure 13 shows the test samples. Compared with the baseline, the finetuned images appear more consistent within each class.



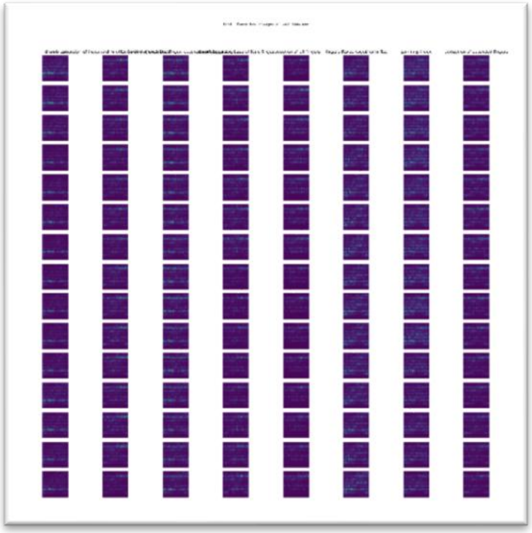


Figure 12. First fifteen CWT images from LOSO-1 finetuning stage.
Figure 13. First fifteen CWT images from LOSO-1 test subject after transfer learning.

3.2.3 Confusion Matrices (LOSO-1 Transfer Learning)

Figures 14–16 show the confusion matrices for the training, validation, and test sets. Transfer learning increases diagonal dominance in the test confusion matrix, indicating better recognition performance on unseen subjects.

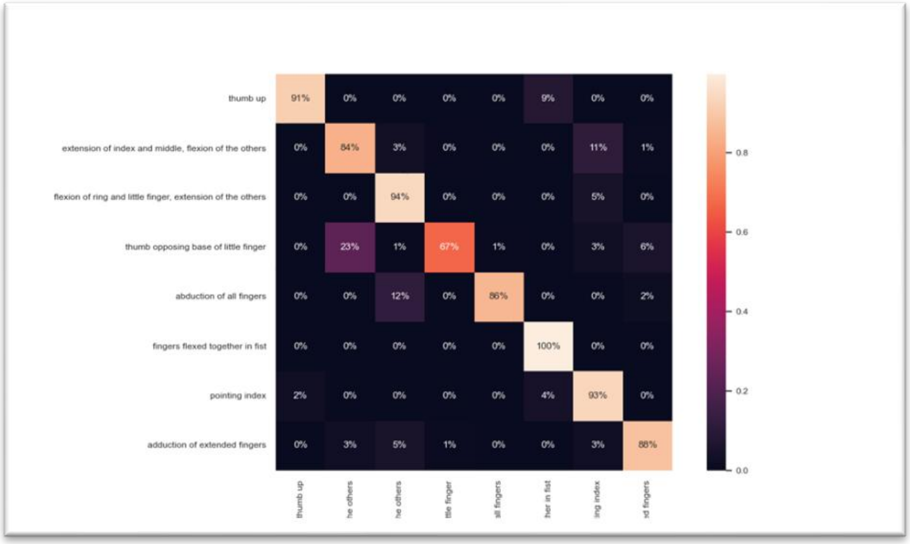


Figure 14. Training confusion matrix for LOSO-1 transfer learning.

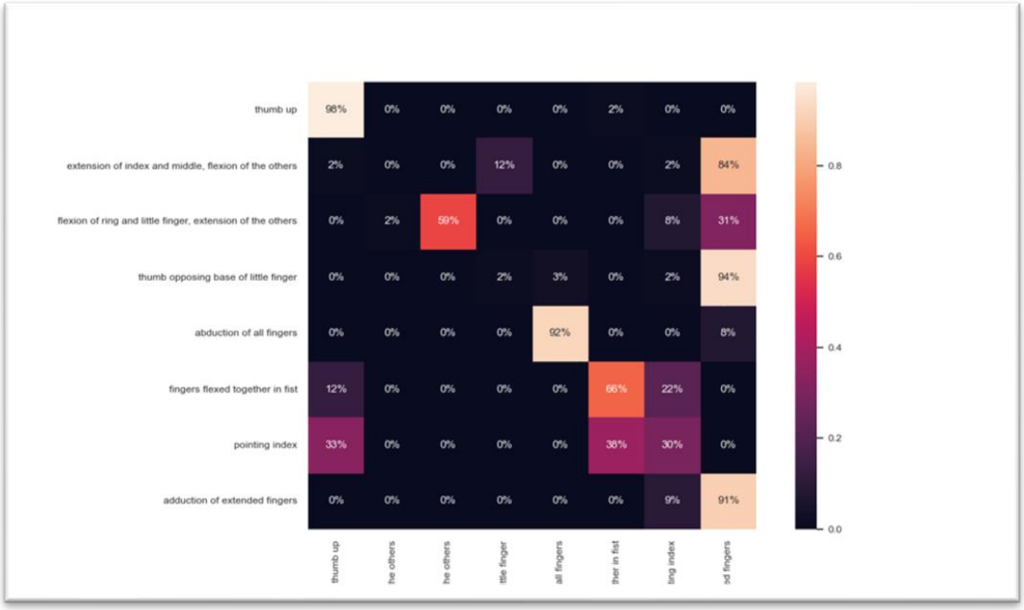


Figure 15. Validation confusion matrix for LOSO-1 transfer learning.

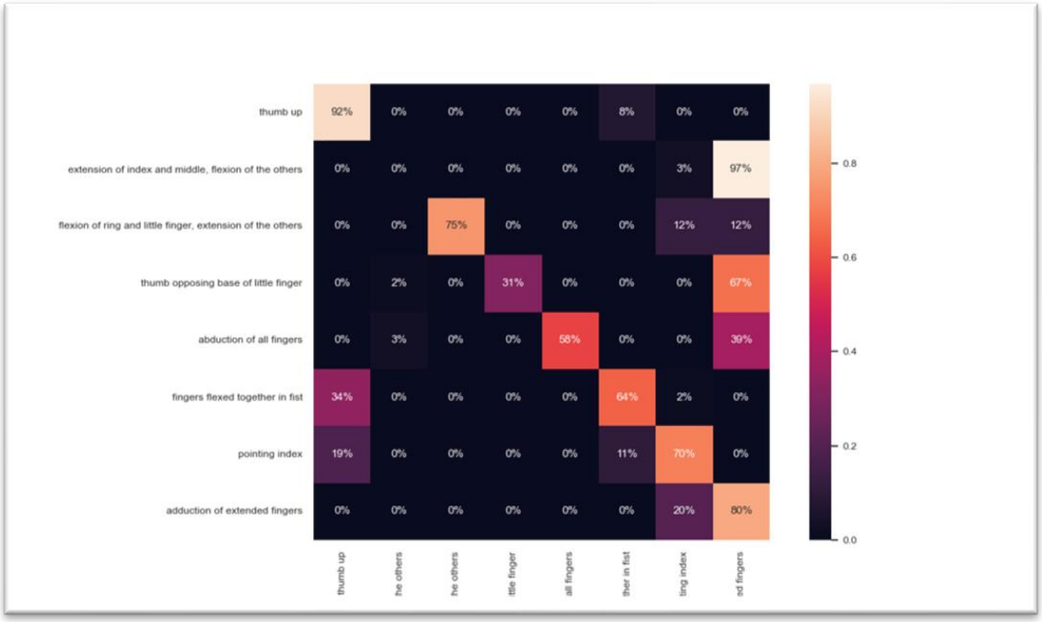
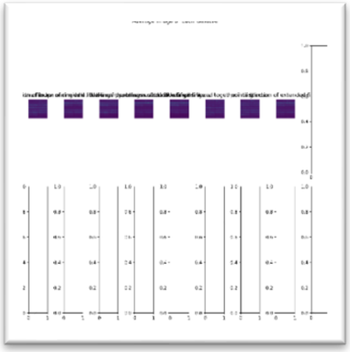


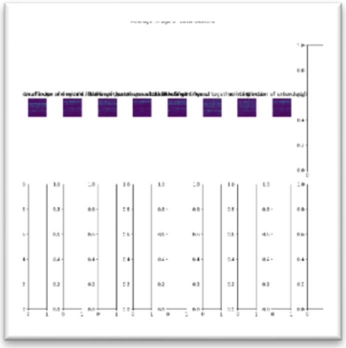
Figure 16. Test confusion matrix for LOSO-1 transfer learning.

3.3 LOSO-2 → Baseline Training (No Transfer Learning)

3.3.1 Average Gesture Images

Figures 17–19 show the averaged CWT maps for the LOSO-2 training, validation, and test sets.





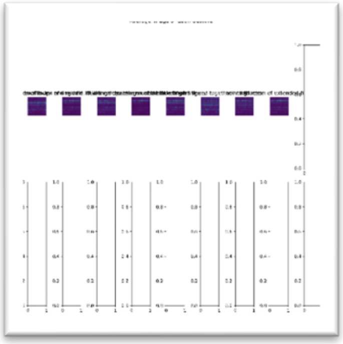


Figure 17. Average CWT images for LOSO-2 training set.
Figure 18. Average CWT images for LOSO-2 validation set.
Figure 19. Average CWT images for LOSO-2 test subject.

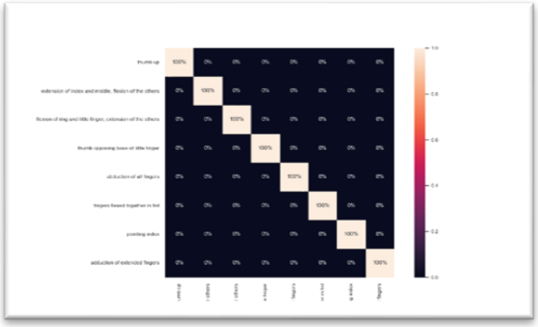
3.3.2 First Fifteen Samples

Figure 20 shows the first fifteen CWT samples from the LOSO-2 test subject.

Figure 20. First fifteen CWT images from the LOSO-2 test subject (baseline).

3.3.3 Confusion Matrices

Figures 21–23 show the confusion matrices for LOSO-2 training, validation, and test sets. As in LOSO-1, training accuracy is high while test accuracy remains low, indicating strong overfitting.



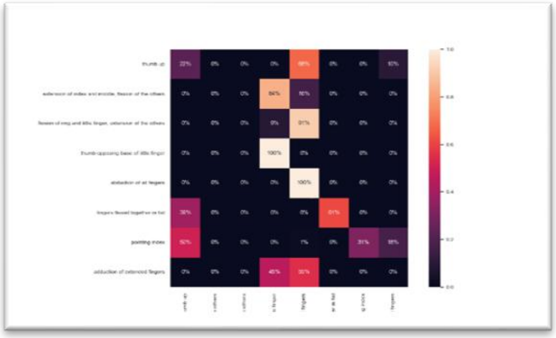


Figure 21. Training confusion matrix for LOSO-2 baseline.
Figure 22. Validation confusion matrix for LOSO-2 baseline.

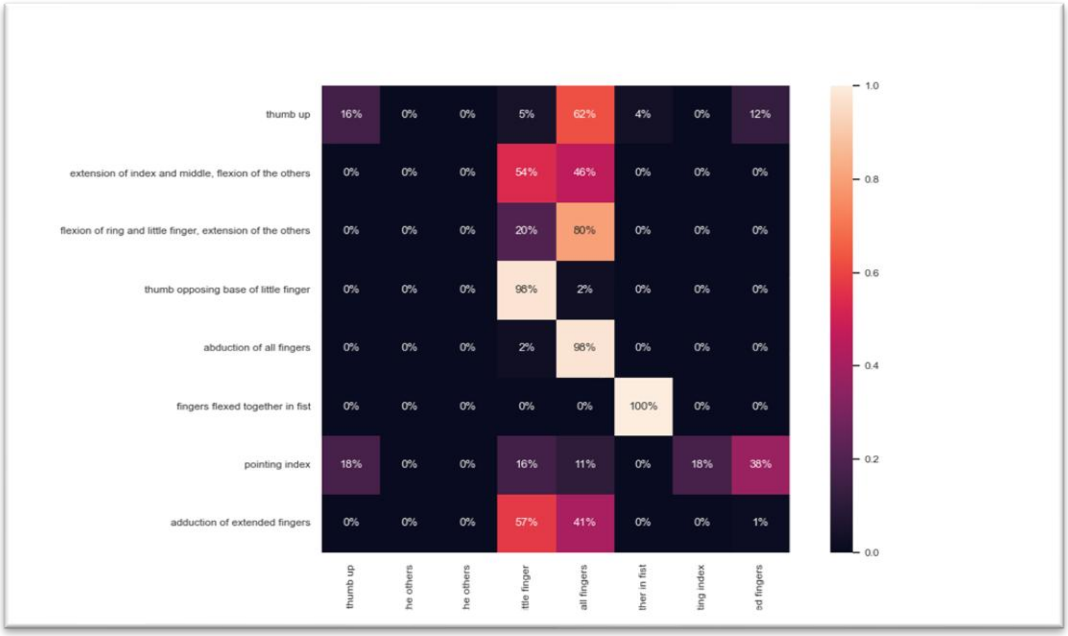
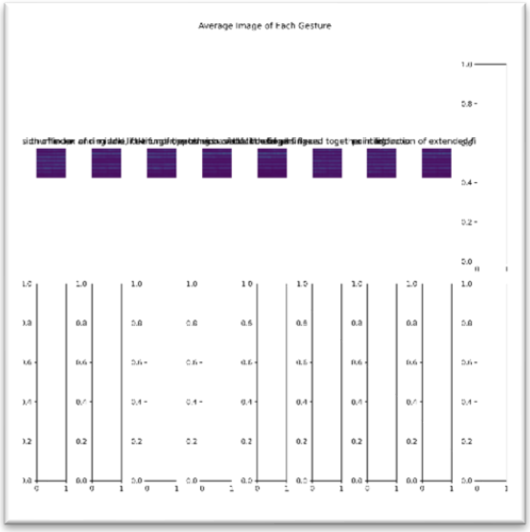


Figure 23. Test confusion matrix for LOSO-2 baseline.

3.4 LOSO-2 → Transfer Learning (Pretrain + Finetune)

3.4.1 CWT Averages

Figures 24–27 show the CWT averages for the pretraining, finetuning, validation, and test sets. Compared with baseline training, gesture clusters appear more distinct.



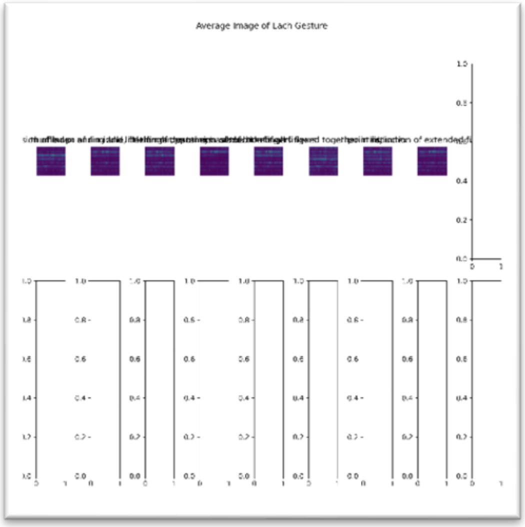
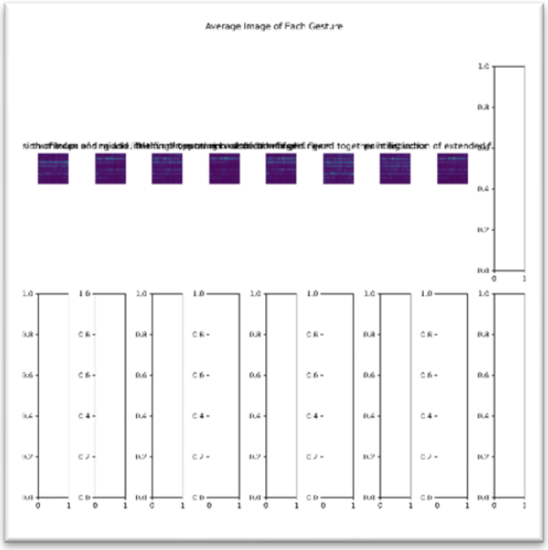


Figure 24. Average CWT images for LOSO-2 pretraining stage.
Figure 25. Average CWT images for LOSO-2 finetuning stage.



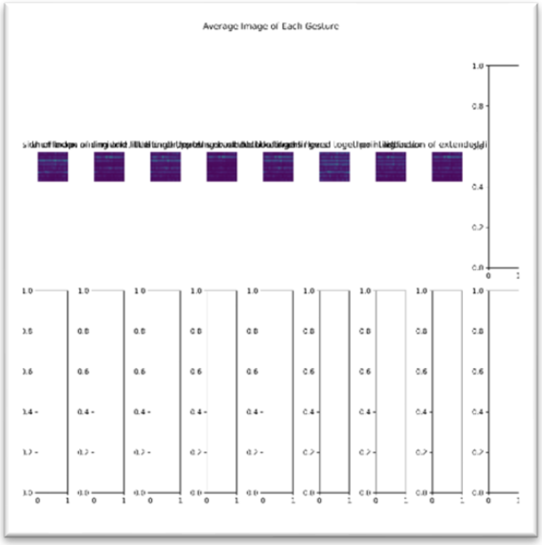
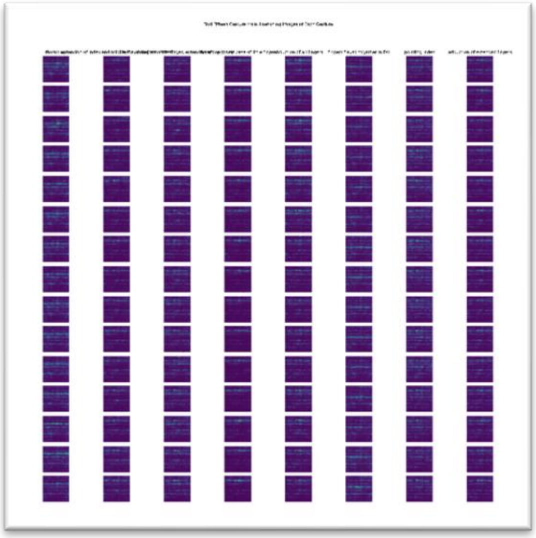


Figure 26. Average CWT images for LOSO-2 validation set.
Figure 27. Average CWT images for LOSO-2 test subject.

28.4.2 **First Fifteen Samples**

Figure 28 shows finetuning samples, and Figure 29 shows the test samples. Transfer learning produces more stable patterns across gesture classes.



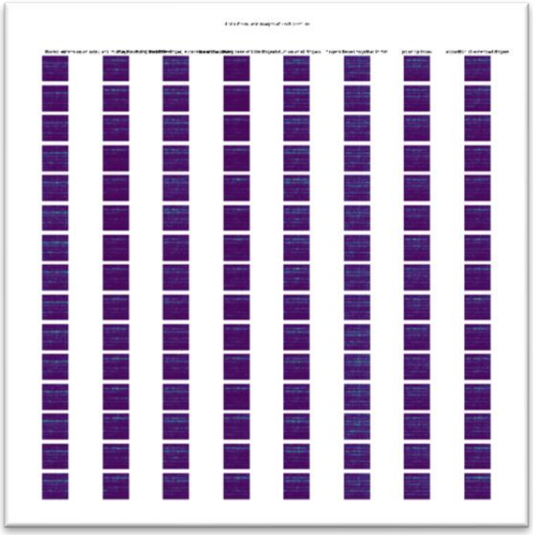
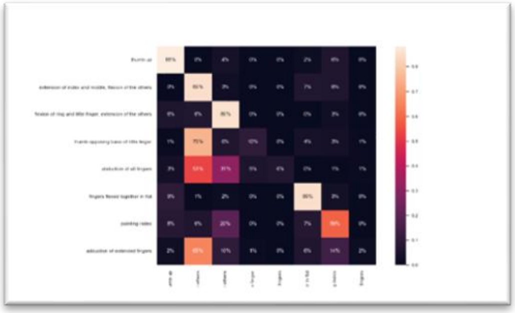


Figure 28. First fifteen CWT images from LOSO-2 finetuning stage.
Figure 29. First fifteen CWT images from LOSO-2 test subject after transfer learning.

3.4.3 Confusion Matrices

Figures 30–32 show the confusion matrices for the training, validation, and test sets. The LOSO-2 transfer-learning model demonstrates the highest test performance among all configurations.



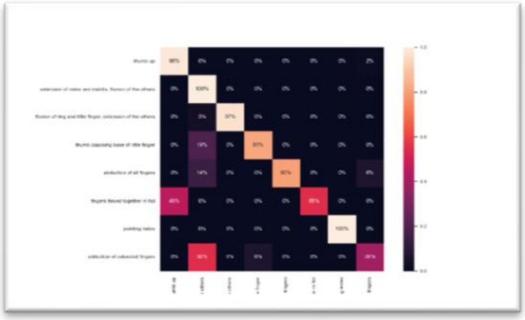


Figure 30. Training confusion matrix for LOSO-2 transfer learning.
Figure 31. Validation confusion matrix for LOSO-2 transfer learning.

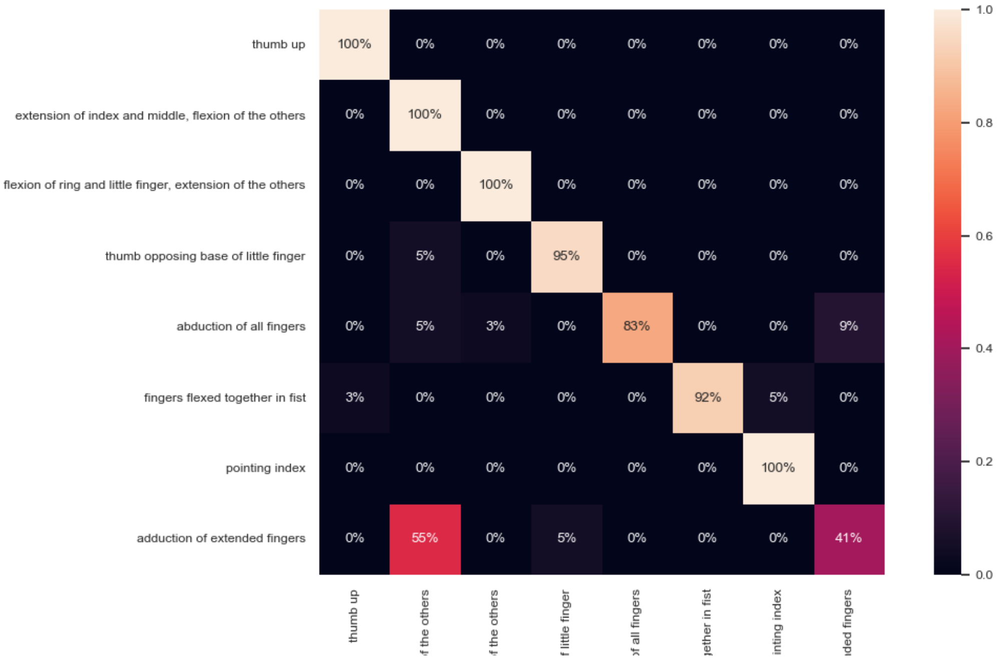


Figure 32. Test confusion matrix for LOSO-2 transfer learning.

4. Discussion

The experiments show that transfer learning consistently improves cross-subject EMG gesture classification compared with training from scratch. This section summarizes the main observations derived from the LOSO-1 and LOSO-2 results.

4.1 Effect of Pretraining on Gesture Separability

In both folds, the transfer-learning model produced confusion matrices with clearer diagonal structure, indicating reduced confusion between gesture classes. The baseline models, on the other hand, showed frequent misclassification, particularly for gestures with similar activation patterns. This suggests that pretraining helps the network learn more stable feature representations before adapting to new subjects.

4.2 Learned Spectral–Temporal Features

The CWT representation exposes meaningful time–frequency information, and the pretrained ResNet-18 model appeared to make better use of this structure. During finetuning, the network adapted more efficiently than models trained from scratch, which tended to overfit to the training subjects. This supports the idea that pretraining extracts reusable low-level patterns that generalize well across subjects.

4.3 Stability During Finetuning

The sample visualizations show that gesture-specific CWT patterns become more consistent after the finetuning stage. This improvement is particularly clear in LOSO-2, where the baseline model produced unstable class boundaries and relatively **low test accuracy (≈42%)**. In contrast, the transfer-learning model achieved substantially **higher performance (≈89%)** and displayed more coherent gesture patterns. These results suggest that finetuning helps the network adapt its decision boundaries to subject-specific variations while preserving the general spectral–temporal features learned during pretraining.

4.4 Impact of Subject Variability

The differences observed between LOSO-1 and LOSO-2 confirm that subject variability remains a major challenge. Factors such as electrode placement, arm posture, and muscle activation strategy produce distribution shifts that limit the performance of models trained from scratch. Transfer learning mitigates some of these effects but does not eliminate them entirely.

4.5 Practical Considerations

The improvements achieved through pretraining and finetuning demonstrate a practical benefit: effective cross-subject generalization can be achieved without collecting large calibration datasets for each new user. This is particularly relevant for wearable or assistive systems, where rapid setup and minimal user-specific data are important constraints.

5. Conclusion

This work examined subject-independent EMG gesture classification using CWT-based inputs and a ResNet-18 model under a Leave-One-Subject-Out (LOSO) evaluation framework. Two training strategies were compared: baseline training from scratch and a transfer-learning approach that included pretraining followed by finetuning.

Across both LOSO folds, transfer learning provided clear gains over baseline training. The test confusion matrices showed fewer misclassifications, and the sample visualizations suggested more stable gesture representations after finetuning. In contrast, baseline models tended to overfit to the training subjects and generalized poorly to unseen users. These findings confirm that subject variability remains a major difficulty in EMG classification, and that pretraining helps mitigate this issue by learning more transferable spectral–temporal features.

Overall, transfer learning was effective in improving cross-subject robustness in EMG gesture recognition. The approach reduces the amount of user-specific data required and is suitable for applications such as wearable interfaces and assistive devices, where rapid adaptation to new users is important. Future work may explore architectures that combine CNN and Transformer components, domain-adaptation techniques, or expanded datasets to further improve generalization across subjects.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.

[2] K. Englehart and B. Hudgins, “A robust, real-time control scheme for multifunction myoelectric control,” *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, 2003.

[3] M. Atzori et al., “Electromyography data for non-invasive naturally-controlled robotic hand prostheses,” *Sci. Data*, vol. 1, pp. 1–13, 2014.

[4] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[5] J. T. Ingram et al., “Cross-subject EMG gesture recognition via deep learning and domain adaptation,” *IEEE Sensors J.*, vol. 21, no. 2, pp. 1767–1778, 2021.

[6] M. Zia ur Rehman et al., “Multichannel EMG signal processing using spectrotemporal analysis for prosthetic control,” *Biomed. Signal Process. Control*, vol. 45, 2018.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016