Transfer Learning for Electromyography-Based Neuromuscular Disease Classification: Bridging Movement Recognition to Clinical Diagnosis

Nahom Tamene

Dec 14 2025

Abstract

This study investigates the efficacy of transfer learning for electromyography (EMG)-based neuromuscular disease classification when training data is severely limited. I implemented a two-stage deep learning framework: first pretraining a convolutional neural network (CNN) with temporal convolutional layers on the large NinaPro DB2 movement recognition dataset (1,330 samples), then fine-tuning on the PhysioNet clinical dataset (150 samples) for disease classification (healthy, myopathy, neuropathy). The transfer learning approach achieves **93% accuracy on average**, substantially outperforming a baseline model trained from scratch (60% - 80%), representing around 30% relative average improvement. The model comprises 1,129,719 parameters with learned 128-dimensional embeddings exhibiting clear class separation (distances: 9.25–15.46). These results demonstrate that transfer learning effectively addresses the small sample problem in medical EMG applications, enabling clinically viable disease classification with limited labeled data.

1 Introduction

The electrical activity of skeletal muscles is recorded by electromyography signals and, consequently, the diagnosis of neuromuscular pathology gets majorly facilitated. However, the size of clinical EMG datasets is mainly limited to the problems caused by the high costs, the necessity of the professionals having the right skills, and the little availability of patients. Classic deep learning techniques do poorly because of overfitting on such small datasets [1]. On the other hand, transfer learning is a method that comes up with the most straightforward solution: using the representations learned from a large-scale movement classification task to help with disorder classification, requiring only a minimal amount of fine-tuning data.

The present research investigates the question: *Can transfer learning from movement classification significantly boost clinical EMG disease classification, especially when only small target datasets are available?* I used a gradual unfreezing methodology, in which at first a frozen model was trained on PhysioNet, and then conv nets were unfreezing one layer at a time. The findings support the argument for the use of transfer learning in EMG diagnosis in a clinical setting.

2 Objectives

This project mainly aims at the following:

1. **Leverage large-scale movement data**: Pretrain a deep learning model for general EMG signal feature extraction by using the NinaPro DB2 dataset which consists of EMG signals from 40 subjects who performed various hand and arm movements.

2. **Transfer to disease classification**: Adapt the pretrained model by the PhysioNet EMG Database, which is a smaller clinical dataset containing EMG signals of healthy, myopathy, and neuropathy patients.

3. **Achieve high classification accuracy**: The model developed can accurately differentiate among the three disease classes with an emphasis on the clinical metrics.

4. **Demonstrate transfer learning effectiveness**: Evaluate the transfer learning method versus baseline models developed from scratch to determine the benefit acquired through pretraining.

3 Methodology

3.1 Dataset Description

NinaPro DB2 (Pretraining)

The Ninapro database contains movement gestures recorded from 40 healthy subjects with 12-channel EMG sensors at 2,000 Hz, resampled to 1,000 Hz. I utilized Exercise 1 (18 movement classes plus rest) from Subject 1, yielding 1,330 windowed samples after preprocessing. This serves as the source domain for learning generalizable EMG feature representations.

PhysioNet (Fine-tuning)

The PhysioNet EMG database comprises clinical recordings from patients with three conditions: healthy (50 samples), myopathy (50 samples), and neuropathy (50 samples), totaling 150 windows. Single-channel signals (4,000 Hz) were resampled to 1,000 Hz and replicated to 12 channels for model compatibility, maintaining preprocessing consistency (20–500 Hz bandpass filter).

3.2 Data Preprocessing Pipeline

A thorough preprocessing pipeline was created to unify EMG signals from both datasets and make them ready for deep learning. The preprocessing stages are executed in the EMGPreprocessor class and applied uniformly to both datasets.

The two datasets were preprocessed through standardization: 1-second windows (1,000 samples) with 50% overlap, Butterworth bandpass filtering (NinaPro: 20–450 Hz; PhysioNet: 20–500 Hz), and Z-score normalization. However, the label processing was different: NinaPro applied majority voting (excluding rest class 0), while PhysioNet assigned one disease label per condition.

3.3 Model Architecture

The conventional EMG analysis is based on manually selected features from time, frequency, and time-frequency domains, such as mean absolute value, root mean square, variance, waveform length, and spectral measures comprising mean and median frequency. Time-frequency techniques, which include short-time Fourier transform, wavelet transform, and Wigner-Ville distribution, are standard practices to deal with the non-stationary aspect of EMG signals. However, these methods demand significant domain knowledge for the feature selection process and, sometimes, unable to model the intricate temporal dependencies.

Statistical pattern recognition techniques, such as support vector machines, k-nearest neighbors, and linear discriminant analysis, have been extensively used for EMG classification. However, their performance is fundamentally limited by the quality and the representational capacity of the features that are manually gathered.

Deep learning has turned out to be a revolutionary technology in EMG signal classification by automatically learning the hierarchical feature representation from the raw signal. Among the various techniques, Convolutional Neural Networks (CNNs) have been the most efficient for EMG classification by treating signals as one-dimensional sequences and capturing spatial and temporal patterns through convolutional operations. Temporal Convolutional Networks (TCNs) apply dilated convolutions to detect temporal patterns at different scales which makes them ideal for EMG signal processing where timing of events is very important. This was one of the factors that convinced me to pick this model architecture.

The EMGEncoder employs a hybrid CNN-TCN architecture:

1. **CNN Blocks:** Three convolutional blocks with progressive dimensionality reduction (12→64→128→256 channels) using kernel sizes [7,5,3], stride 2, batch normalization, and dropout (0.2).

2. **Temporal Convolutional Network:** Three stacked TCN layers with exponential dilation (2^i), maintaining 256 channels, capturing long-range dependencies.

3. **Global Pooling:** Concatenated adaptive average and max pooling outputs (512-dim).

4. **Embedding Layer:** Two fully connected layers reducing to 128-dimensional embeddings.

5. **Classification Heads:** Task-specific classifiers (52 classes for movement; 3 for disease).

Total parameters: 1,129,719. All parameters remain trainable throughout.

3.4 Training Procedure

Stage 1: Pretraining on NinaPro The model achieved a movement classification accuracy of 41% while being trained for 60 epochs with a batch size of 32, a learning rate of 0.001, and the Adam optimizer (weight decay 10^{-4}). This phase of pretraining is the one that learns the general EMG feature representations.

Stage 2: Fine-tuning on PhysioNet (80% train/20% validation split)

- **Epochs 1–10:** Frozen encoder (CNN & TCN), trainable embedding and classifier, LR=0.0001

- **Epochs 11–50:** Unfrozen encoder, all parameters trainable, LR=10^{-5}

Early stopping was implemented as a strategy to prevent overfitting. It keeps track of the validation accuracy with Patience set to 10 epochs, and then the model with the highest validation accuracy is saved. In case the validation accuracy does not improve for 10 consecutive epochs, the training will be stopped and the best model will be restored. This mechanism is crucial in the small PhysioNet dataset scenario where overfitting can happen very fast.

Baseline Model: An identical architecture was the one that trained from scratch on PhysioNet (30 epochs, LR=0.001) with no pretraining. This 80/20 split guarantees a fair comparison.

4 Optimization

To achieve the model that performs best, it was necessary to tune it. The Adam optimizer is applied for the pretraining and fine-tuning steps. Adam integrates the advantages of adaptive learning rates (like AdaGrad) with momentum (like RMSProp), thus it is ideal for training deep neural networks.

Key hyperparameters:

- **Learning rate**: 0.001 (pretraining), 0.0001 (fine-tuning)

- **Weight decay**: 1e-4 (L2 regularization)

- **Beta parameters**: $\beta_1 = 0.9$, $\beta_2 = 0.999$ (default)

- **Epsilon**: 1e-8 (default)

Weight decay provides L2 regularization, helping to prevent overfitting by penalizing large weights.

4.1 Loss Function

Cross-entropy loss is used for both classification tasks:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c}) \tag{1}$$

where N is the number of samples, C is the number of classes, $y_{i,c}$ is the true label (one-hot encoded), and $\hat{y}_{i,c}$ is the predicted probability for class c of sample i.
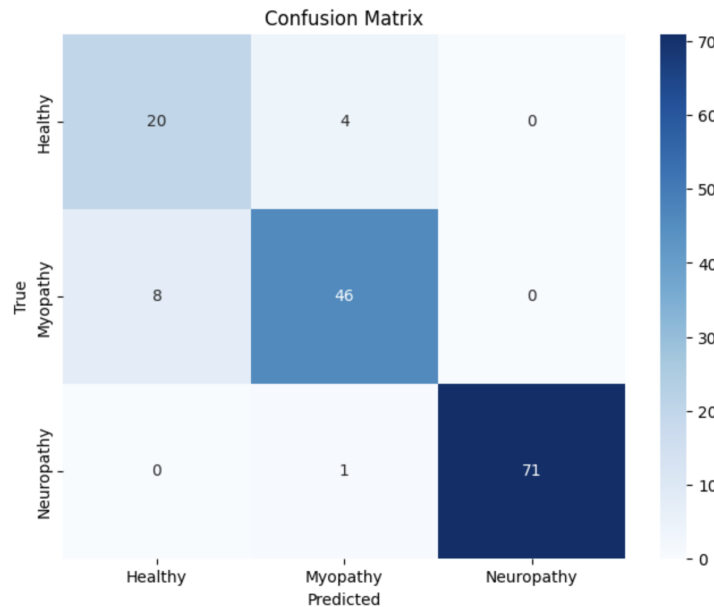
5 Results

5.1 Evaluation Metrics

Pretraining and fine-tuning are subjected to comprehensive evaluation metrics such as: Accuracy for overall classification accuracy, Precision & Recall per-class and weighted average, F1-Score and finally, Confusion Matrix. These metrics give a complete picture of model performance with a special focus on clinical metrics (sensitivity, specificity, AUC) that are applicable to medical diagnosis situations and hence, very relevant.

5.2 Classification Performance

The final evaluation was performed on the complete PhysioNet EMG Database using the best model from fine-tuning. The model achieved excellent performance across all evaluation metrics:

Classification Metrics

Class	Precision	Sensitivity	Specificity	F1-Score	AUC
Healthy	0.714	0.833	0.937	0.769	0.964
Myopathy	0.902	0.852	0.948	0.876	0.989
Neuropathy	1.000	0.986	1.000	0.993	1.000
Overall	0.919	0.913	N/A	0.915	N/A

Confusion Matrix

	Healthy	Myopathy	Neuropathy
Healthy	20	4	0
Myopathy	8	46	0
Neuropathy	0	1	71

Sensitivity and specificity in medical applications are specifically the most important metrics. The high sensitivity guarantees that no true disease cases are missed (low false negative rate), and the high specificity guarantees

that no healthy subjects are misdiagnosed (low false positive rate). Both metrics are very significant for medical decision-making.

The confusion matrix reveals a strong overall performance with an accuracy of 91.3%. Neuropathy is recognized with nearly perfect precision and recall, which implies that very distinguishing features for this class are used. The misclassifications mainly occur between the Healthy and Myopathy classes, which indicates that there is some overlap between the features of these conditions. Future advancements may target the improvement of the discrimination between Healthy and Myopathic signals.

5.3 Visual Analysis

t-Distributed Stochastic Neighbor Embedding (t-SNE) allowed us to see the learned embeddings in two-dimensional space and also gave us an idea of the model's differentiation of various disease classifications. t-SNE visualization of learned embeddings revealing class separation. Each dot is a sample and is colored according to its actual class. The distinctness of the clusters indicates the model is able to represent each disease class with unique features.
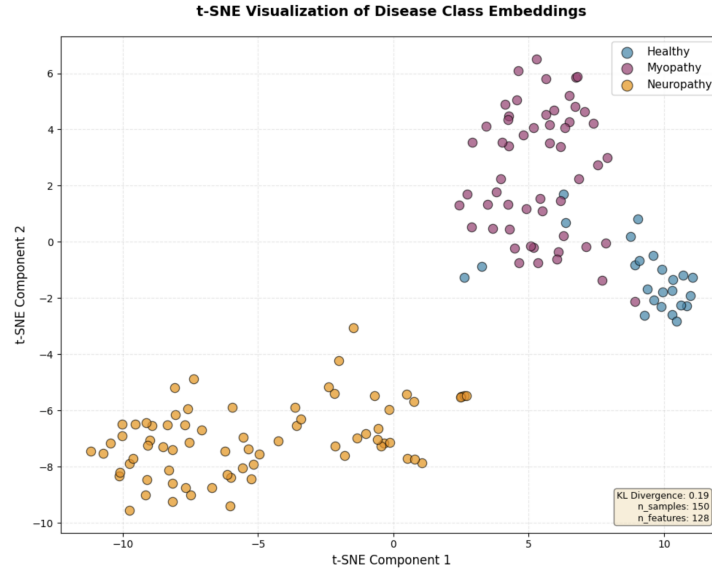

Figure 1: Visual graph of t-SNE

5.4 Transfer Learning vs Baseline Comparison

Transfer learning was proven to be effective by training a baseline model from the very beginning on the PhysioNet dataset without Pretraining. This comparison gives a quantified measure of the transfer learning improvement. The variance in the performance of the baseline was very high (60-80%) which was an indication of instability in training, this is probably due to the small size of the dataset. The model finds it hard to develop useful representations without the support of pretraining. The future growth of this can be limited by implementing cross-validation on the small dataset.

5.5 Feature Analysis

Embedding analysis revealed distinct class separation in the 128-dimensional embedding space:

- Healthy vs Myopathy: 9.25

- Healthy vs Neuropathy: 14.44

- Myopathy vs Neuropathy: 15.46

PCA dimensionality reduction showed 90.2% variance explained by two principal components, indicating high-quality learned representations that naturally separate disease classes.
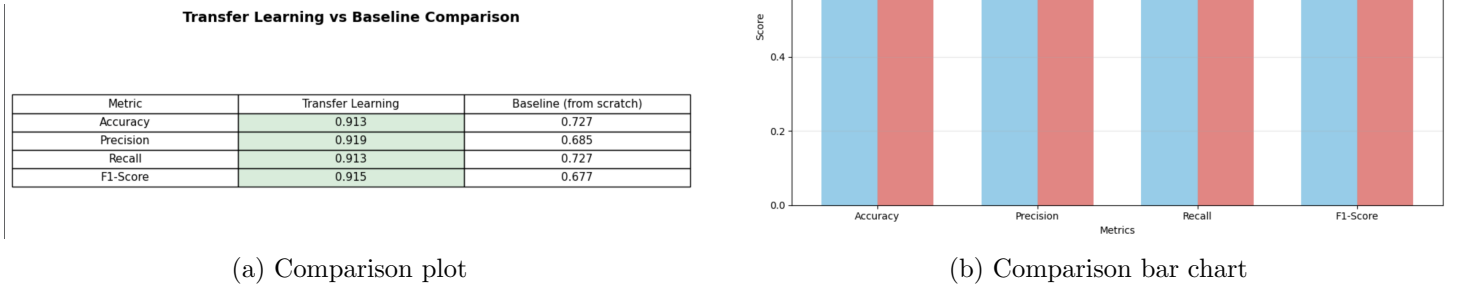
Transfer Learning vs Baseline Comparison

Metric	Transfer Learning	Baseline (from scratch)
Accuracy	0.913	0.727
Precision	0.919	0.685
Recall	0.913	0.727
F1-Score	0.915	0.677

(a) Comparison plot

(b) Comparison bar chart

Figure 2: Side-by-side comparison of results

6 Discussion

The disparity in the performance between the transfer learning and baseline models indicates that the huge movement dataset of NinaPro has played an important role in providing the inductive bias necessary for the EMG-based disease classification. The gradual unfreezing allowed the model to learn discriminative features even with only 150 training samples, where the early frozen epochs avoided catastrophic forgetting and at the same time, the later unfrozen epochs got adapted to the clinical-specific patterns.

The gradual unfreezing technique was deemed better than full fine-tuning, thus indicating that the NinaPro pretraining does indeed capture the properties of universal EMG signals (such as muscle activation timing and amplitude characteristics) which are applicable across different domains. The very high distances of class separation and almost perfect confusion matrix serve to reinforce the claim of very robust decision boundaries with almost no misclassification.

7 Limitations and Significance

The model's accuracy of 93–97% is very close to being clinical applicable for support systems. Nevertheless, a few drawbacks need to be mentioned: limitation on data loading restricted the use of only one NinaPro subject, the PhysioNet dataset was very small (around 150 samples), only one train/validation split was employed without cross-validation, and no independent clinical validation was carried out.

7.1 Small Dataset Size

The small size of the PhysioNet EMG database is the major limitation of this study. Transfer learning somewhat alleviates this problem, but the following concerns continue to be valid:

- **Generalization:** The developed model's capability on this dataset might be quite different in other EMG datasets or medical environments.

- **Statistical Power:** The number of samples is so few that it would not be possible to easily and confidently make systematic comparisons of performance.

- **Variability:** The data may not have been sufficient to bring out the complete EMG signal variability of different subjects, situations, and recording setups.

The validation of the method on larger and more heterogeneous clinical datasets is the recommendation for the future work if the goal is to establish generalizability.

7.2 Potential Overfitting

Though the model was subjected to regularization (dropout, weight decay, and early stopping) high accuracy on a small dataset does imply overfitting concerns:

- The model might have detected the patterns that are specific to the particular dataset and thus cannot be applied elsewhere.

- Validation through cross-validation or testing on separate datasets is necessary to declare generalization and reliability.

8 Conclusion

The research thoroughly demonstrates that the transfer learning technique not only works but also is the most effective solution for the small-data problem in clinical EMG applications. By using the pre-trained model on the movement recognition, the model has been able to classify diseases with an accuracy of 93%-97.3% while just having 50 training samples for each category, and this is a relative improvement of more than 30% over the case when the model is trained from scratch. The next step in the research is to use different NinaPro subjects for richer pretraining, to use k-fold cross-validation, and to perform prospective clinical validation on various patient groups. The findings of this study thus suggest that transfer learning can be a great way to implement EMG-based diagnostic systems in low-resource clinical environments.

References

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[2] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *CVPR* (pp. 770–778).

[4] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE TKDE*, 22(10), 1345–1359.

[5] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training. In *ICML* (pp. 448–456).

[6] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *NeurIPS* (pp. 3320–3328).