



A Residual Temporal Convolutional with Attention Neural Network for Electromyogram-Based Hand Gesture Recognition

Rachid Namane^{*}, Elhocine Boutellaa[®], Sif Eddine Salem[®], Yassine Babaci[®]

Institute of Electrical Engineering and Electronics, University M'Hamed Bougara, 35000 Boumerdes, Algeria

^{*} Correspondence: Rachid Namane (rachid.namane@univ-boumerdes.dz)

Received: 07-30-2025

Revised: 09-20-2025

Accepted: 09-25-2025

Citation: R. Namane, E. Boutellaa, S. E. Salem, and Y. Babaci, "A residual temporal convolutional with attention neural network for electromyogram-based hand gesture recognition," *Int. J. Comput. Methods Exp. Meas.*, vol. 13, no. 3, pp. 739–748, 2025. <https://doi.org/10.56578/ijcmem130320>.



© 2025 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: Electromyography (EMG)-based hand gesture classification is a developing core technology for designing intuitive and responsive human-computer interaction, notably for prosthetic control. EMG signals, which reflect muscle activity during contraction, offer a non-invasive and effective method for capturing user gestures. However, because of their natural variability, noise, and temporal richness pose significant hurdles to precise gesture recognition. In this paper, we investigate the use of causal convolutional layers, which are suitable for sequential data, to improve hand gesture recognition from raw EMG signals. We propose a deep neural network which bases on temporal convolutions and integrates residual connections and contextual attention in an end to end hand gesture recognition system. Furthermore, we apply multiple data augmentation techniques to mitigate intra-subject variability and enhance model generalization. Our approach is evaluated on the benchmark NinaProDB1 dataset. The proposed model show impressive classification performance with an average accuracy of 95.31% and where the majority of the gestures from various subjects were accurately recognized. These results demonstrate the effectiveness of causal convolutions and attention mechanisms for robust EMG-based gesture recognition.

Keywords: Electromyogram; Hand gesture recognition; Deep learning; Causal convolution; Attention mechanism

1 Introduction

Electromyography (EMG)-based hand gesture recognition has become a pivotal technology in developing intuitive and responsive human-computer interaction (HMI) systems. In particular, prosthetic control represents a major application domain in biomedical engineering. Surface EMG signals, which reflect underlying muscle activity during contraction, offer a non-invasive and effective way to decode user gesture intent, particularly for amputees [1]. The process involves classifying muscle activation patterns into discrete hand or wrist movements, enabling intuitive control over assistive devices. However, the use of EMG signal classification presents several challenges due to the non-stationary nature of the signal, inter-subject variability, muscle fatigue, and noise introduced by sensors or motion artifacts. Therefore, accurate EMG signal classification requires models that can robustly extract temporal and spatial features from multi-channel time-series data especially in real-world scenarios where generalization across users and sessions is required.

In recent years, deep learning (DL) has exhibited impressive capability in biomedical signal processing. In the context of electromyography (EMG)-based hand gesture recognition, DL has demonstrated remarkable promise, offering the potential to build more accurate and adaptive human-machine interfaces (HMIs) [2, 3]. Unlike traditional machine learning algorithms that rely on handcrafted features, which often restrict their scalability and generalization to unseen data, deep neural networks can automatically extract hierarchical and temporal features directly from raw EMG data [4, 5]. CNNs have been widely applied for capturing spatial representations, while RNNs and LSTMs are used for sequential modeling. These capabilities make them well-suited for handling dynamic, high-dimensional time-series data such as EMG. More recently, hybrid approaches combining CNNs, LSTMs, or attention mechanisms have improved recognition accuracy, yet they often struggle with training efficiency, long-term temporal modeling, and generalization across subjects.

Temporal Convolutional Networks (TCNs) have emerged as a compelling alternative to recurrent architectures, thanks to their use of causal convolutions and dilation. These properties allow TCNs to model long-range temporal

dependencies. Unlike standard CNNs, which are primarily designed for spatial feature extraction, TCNs utilize causal convolutions and dilation, enabling them to preserve sequence order while covering broad temporal contexts. This structure not only alleviates the vanishing gradient problem commonly associated with recurrent architectures but also allows for more efficient parallelization during training [6, 7]. Despite their promise, their application in EMG-based hand gesture recognition remains underexplored, particularly in combination with advanced techniques such as residual connections, attention mechanisms, and systematic data augmentation.

This study aims to address this gap by proposing a TCN-based model with residual connections and contextual attention for robust EMG-based hand gesture recognition. Unlike prior CNN or LSTM approaches, our model is designed to i) capture long-range temporal dependencies efficiently, ii) enhance discriminative focus through attention, and iii) mitigate overfitting via data augmentation. Our hypothesis is that this architecture can achieve superior accuracy and generalization compared to existing methods.

To validate our approach, the proposed model is trained and evaluated on a subset of the benchmark NinaPro-dataset, demonstrating competitive performance and promising generalization capabilities, which highlights its potential for real-time prosthetic control and human-machine interaction.

The remainder of this paper is organized as follows. Section 2 reviews related literature on deep learning for EMG-based gesture recognition. Section 3 presents the proposed methodology. Section 4 describes the experimental evaluation of the proposed approach, and Section 5 concludes the paper with a summary of contributions and directions for future improvement.

2 Related Work

A range of approaches have been explored for EMG-based hand gesture recognition, spanning classic machine learning, deep learning, and hybrid architectures. This section highlights key works, emphasizing temporal modeling methods and those tailored to the NinaPro dataset. Atzori et al. [8] leveraged a LeNet-based CNN on NinaPro DB2, achieving 78–85% accuracy across multiple gestures with feature-based SVM classifiers, a solid baseline for later DL models. Geng et al. [9] introduced the concept of instantaneous sEMG imaging and applied a CNN plus SVM on 8 gestures captured via high-density sEMG imaging, attaining 89.3% per-frame accuracy (and 99.0% after majority voting), illustrating the potential of single-frame classification. Du et al. [4] designed a custom LSTM-based model on NinaProDB2/DB3, achieving 85–90% accuracy on sequence recognition tasks, outperforming classical approaches.

While earlier studies demonstrated strong foundational performance using CNNs and LSTMs, more recent architectures have adopted TCNs and attention mechanisms to improve both accuracy and efficiency. Tsinganos et al. [10] employed TCN for EMG-based hand gesture recognition using the NinaPro DB1 dataset, achieving around 89.76% accuracy and demonstrating the strength of TCNs in capturing temporal dependencies in muscle activity signals. Zanghieri et al. [11] proposed TEMPONet-embedded TCN, A real-time embedded TCN model, achieving 93.7% accuracy on NinaPro enabled sessions with an ultra-low memory footprint.

Recent TCN and multi-stream architectures have further advanced the field. Rahimian et al. [12] proposed a lightweight TCN-attention model evaluated on 17 gesture classes, achieving 81.65% (300 ms windows) and 80.72% (200 ms windows), with $12 \times$ fewer parameters than benchmark models. A multi-stream deep architecture proposed by Shin et al. [13], integrating TCN, CNN, LSTM modules, and channel attention, was tested on NinaPro DB1 and DB9, attaining 94.3% and 98.96% accuracy, respectively, demonstrating the power of hybrid models for high-performance EMG classification.

Beyond these temporal models, recent research has introduced novel paradigms. Montazerin et al. [14] applied a Vision Transformer (ViT-HGR) to HD-sEMG, reporting 84.6% accuracy on 65 gestures with a remarkably compact parameter count, illustrating the promise of transformer-based methods. Zhong et al. [15] developed a Spatio-Temporal Graph Convolutional Network (named STGCN-GR) for HD-sEMG-based human-machine interfaces, which explicitly models electrode topology and reached 91.07% accuracy on 65 gestures, outperforming CNN-based baselines. Xiang et al. [16] introduced SE-DenseNet with channel attention, achieving 85.93% and 82.39% accuracy on NinaPro DB2 and DB4, respectively, highlighting the effectiveness of attention in EMG recognition.

To synthesize these advancements, Table 1 provides a comparative overview of representative studies. It summarizes the datasets used (including number of gestures), core architectures, and reported performance metrics, offering a consolidated view of how different methods perform across diverse experimental contexts.

Although our work employs temporal convolutions similar to those used in related studies, our proposed architecture distinguishes itself by incorporating multiple residual causal convolution blocks and a contextual attention mechanism placed at the top of the network to enhance feature discrimination. Moreover, unlike prior works that focus on reducing the number of layers to achieve lightweight models, our design leverages deeper residual structures to improve representational capacity. Additionally, we apply suitable data augmentation techniques to enhance learning and promote better model generalization.

Table 1. Comparative overview of EMG-based hand gesture recognition models

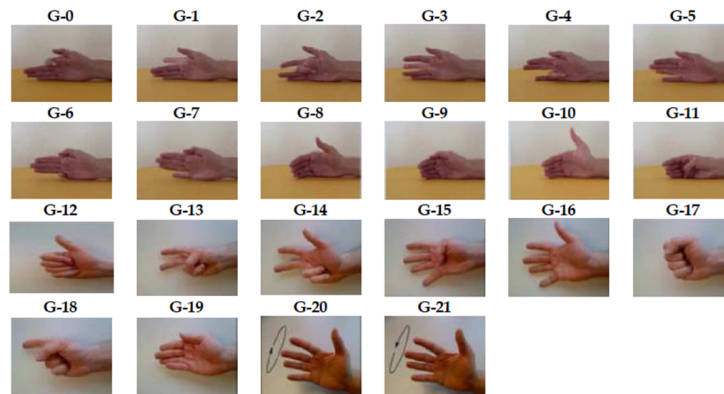
Study/Model	Dataset	Architecture	Accuracy
Atzori et al. [8]	NinaProDB2	CNN + SVM	78–85%
Geng et al. [9]	HD-sEMG	CNN + SVM	89.3% (frame), 99% (voting)
Du et al. [4]	NinaProDB2/DB3	Custom LSTM	85–90%
Tsinganos et al. [10]	NinaProDB1	TCN	89.76%
Zanghieri et al. [11]	NinaPro-enabled sessions	TCN on GAP8 processor	93.7% with ultra-low compute overhead
Rahimian et al. [12]	NinaProDB2	TCN + Attention (lightweight)	81% with few parameters
Shin et al. [13]	NinaProDB1/DB9	TCN + CNN + LSTM + Attention	94.3%/98.96%
Montazerin et al. [14]	HD-sEMG	ViT-HGR	84.6%
Zhong et al. [15]	HD-sEMG	STGCN-GR	91.07%
Xiang et al. [16]	NinaProDB2/DB4	SE-DenseNet	85.93%/82.39%

3 Methodology

3.1 Data and Preprocessing

There exists a number of publicly available sEMG hand gesture datasets. These data sets vary in terms of acquisition device, number of subjects, types of gestures, number of gesture repetitions, etc. In this context, NinaProdatabase [2] is considered to be one of the most comprehensive resources, today comprising 9 separate data sets collected from 204 healthy subjects and 15 amputee subjects. NinaPro contains a set of upper limb electromyographic, kinematic and dynamic data to allow the public to test machine learning algorithms for controlling hand prostheses and conducting other related researches. The 9 datasets include multiple repetitions of at least 50 hand movements. The latter were recorded with different acquisition protocols and configurations, to allow the recording of several multi-modal signals. Despite the fact that NinaPro only includes data from 15 hand amputees, it has been shown that data from healthy subjects can also be used as a proxy measure for amputees. This result justifies the use of healthy subjects in order to reduce the stress and pain that can be caused in amputees [17].

NinaPro DB1 contains sEMG data captured from 10 channels, obtained at a sampling frequency of 100 Hz. In our experiments, we use data of 10 subjects from NinaPro DB1. For classification, we consider a total of 22 gestures provided in Figure 1. Each gesture was repeated 10 times. The dataset also includes a "Restimulus" label, which represents the same movements, but with improved time labels that better match when the movements really happened. The raw sEMG signals were preprocessed and structured into a consistent format for training and evaluating the model. The preprocessing includes segmenting the EMG signal according to the repetition and windowing. This process is performed for each subject separately.

**Figure 1.** Hand gestures considered in our work

Inspired by established methods in time-series literatures [18, 19], we apply the following multiple data augmentation techniques, which attempt to simulate real-world EMG signal variations during training:

- Jittering: adds adjustable Gaussian noise to match a target SNR, mimicking real-world EMG noise from sensors or ambient electrical interference.

- Channel Shuffling: circularly rotates the EMG channel order by a random offset (± 2 channels). It models slight misplacement of electrodes, especially relevant for wearable EMG systems.
- Scaling: multiplies each channel by a random factor sampled from a normal distribution around 1 ($\sigma = 0.2$). It simulates inter-session variability like sensor gain or changes in muscle contraction strength.
- Permutation: divides the signal into segments and randomly shuffles them. It encourages the model to focus on local temporal features rather than fixed global ordering.

3.2 Proposed Model Architecture

As outlined in Figure 2, our proposed model, named Residual Temporal Convolutional Attention Neural Network (RTCAN), attempts to capture both short-term and long-term patterns in EMG signals. To this end, we use i) causal convolutions which do depend only on past data combined with ii) residual connections that let the signal jump over some layers to help train deeper networks, and iii) an attention mechanism to focus on important parts of the signal. The input of the model is a preprocessed EMG sample of size $n \times 10$, corresponding to the number of time steps and sensor channels. The output of the model is a discrete value specifying the recognized gesture.

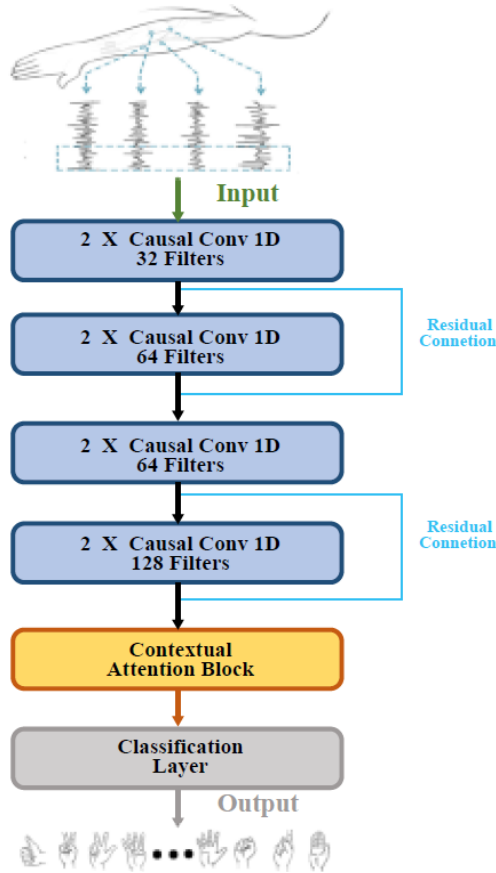


Figure 2. Proposed RTCAN model architecture

In the following, we discuss the details of each block of the proposed architecture.

3.2.1 Residual causal convolutional blocks

The role of the Residual Causal Convolutional Blocks is to identify specific patterns over time for robust gesture classification. The basic building blocks are temporal convolution layers. In contrast to the usual convolution where the output at time t depends on past, present, and future inputs ($x(t - k/2)$ to $(t + k/2)$, k is the filter size), in the temporal convolution, the output depends only on the past time steps. Another advantage of the temporal convolution is the use of dilated convolutions to increase their receptive field without drastically increasing the number of parameters.

$$y(t) = \sum_{i=0}^{k-1} f(i) \cdot x(t - d \cdot i) \quad (1)$$

where, $y(t)$ is the output at time step t , $f(i)$ is the filter of size k , $x(t - d.i)$ is the input sequence, and d is the dilation factor. This last controls the convolution receptive field.

Residual layers are added to maintain gradient flow, and the overall transformation can be described by:

$$H(x) = F(x) + x \quad (2)$$

where, $H(x)$ is the output, $F(x)$ is the function learned by the skipped convolution layers, and x is the input that gets passed through via the residual connection.

Our model comprises four causal convolution blocks strengthened by two residual connections. The first is from the output of the first causal convolution block to the input of the third one, and the second connection is from the output of the third causal convolution block to the input of the contextual attention block. Each causal convolution block has specific parameters as follows: [2, 2, 2, 1] and [[32, 32], [32, 32], [64, 64], [128]] for the number convolution layers and their corresponding number of filters, respectively. The ReLU activation function is applied to all convolution layers. To prevent model overfitting L2 regularization is applied to the kernel weights and each convolution is followed by a dropout layer. A summary of the causal convolutional blocks' parameters is provided in Table 2.

Table 2. Causal convolutional blocks parameters

Block	# Layers	Filter(s)	Purpose
Block 1	2	32, 32	Initial temporal feature extraction
Block 2	2	32, 32	Reinforce base temporal patterns
Block 3	2	64, 64	Intermediate pattern abstraction
Block 4	1	128	High-level temporal representation

The residual causal convolution blocks help the model learn how EMG signals change over time, while keeping the order of the signals, before being processed by the contextual attention.

3.2.2 Contextual attention block

The attention mechanism implemented in this work builds upon the hierarchical attention networks proposed by Yang et al. [20], adapting it for temporal EMG signal analysis. Unlike traditional sequence-to-sequence attention that operates between encoder and decoder states, this implementation processes a single temporal sequence to identify the most discriminative time steps for classification. Its mathematical formulation is as follows. Given an input sequence representation $X \in \mathbb{R}^{B \times T \times F}$, where B is the batch size, T is the number of time steps, F is features' dimension, the attention mechanism computes importance weights through three learned components:

Transformation Matrix (W): Projects the input features into the attention space:

$$U = \tanh(+XW + b) \quad (3)$$

where, $W \in \mathbb{R}^{F \times F}$ and $b \in \mathbb{R}^F$ are learnable parameters.

Context Vector (u): Learns a fixed query vector that determines which time steps are relevant:

$$a' = Uu \quad (4)$$

where, $u \in \mathbb{R}^F$ is learned during training.

Normalized Attention Weights: Computed via softmax with masking support:

$$\alpha_t = \frac{\exp(a'_t)}{\sum_{i=1}^T \exp(a'_i)} m_t \quad (5)$$

where, $m_t \in \{0, 1\}$ is the mask value for step t .

The final context vector is the weighted sum:

$$c = \sum_{t=1}^T \alpha_t x_t \quad (6)$$

3.2.3 Classification layer

The top of the model is the classification layer implemented as a fully connected layer with softmax activation and its size the number of classes (recognized gestures).

4 Experimental Evaluation

4.1 Experimental Setup

The experimental setup defines the configuration encompassing the dataset, the model architecture, and the training strategy.

To ensure exclusive data samples between train and test, for each subject, data is portioned into disjoint sets based on repetitions. Specifically, for each subject signals from repetitions $\{1, 3, 4, 6, 8, 9, 10\}$ are used for training and those from repetitions $\{2, 5, 7\}$ are used for test. This setting allows fair model evaluation under unseen test data. For training, data augmentation is applied with a noise level of 25 dB SNR, along with magnitude warping and time warping, each set to 0.2. The batch size for training is 32, and samples are weighted based on their class frequency.

The training process uses the Adam optimizer with an initial learning rate of 0.001. A constant learning rate schedule is applied throughout the 100 training epochs. To promote model generalization on unseen data, we employ i) weight decay with a regularization strength of 0.0001; ii) a dropout rate of 0.4 after convolutions; iii) L2 regularization with a coefficient of 0.0005.

4.2 Results

Inspired by related works, we perform a subject-specific evaluation. Thus, for each subject, we train and evaluate the model with his own data ensuring that the test and train data subsets are disjoint.

Table 3. Per subject train and test performance

Subject #	Training Time (s)	Test Time (s)	Train Accuracy (%)	Test Accuracy (%)
1	543.57	0.2177	99.63	92.19
2	560.14	0.3490	98.96	93.75
3	822.87	0.4937	99.48	98.44
4	837.55	0.5068	99.27	89.06
5	1087.82	0.6797	99.58	100.00
6	927.05	0.3588	99.58	92.19
7	730.95	0.3680	99.22	96.88
8	645.26	0.4350	99.48	96.88
9	1515.46	0.4797	97.92	95.31
10	676.32	0.3479	98.75	98.44
STD	292.13	0.13	0.53	3.46
Average	834.70	0.4236	99.19	95.31



Figure 3. Learning graphs for Subject #4

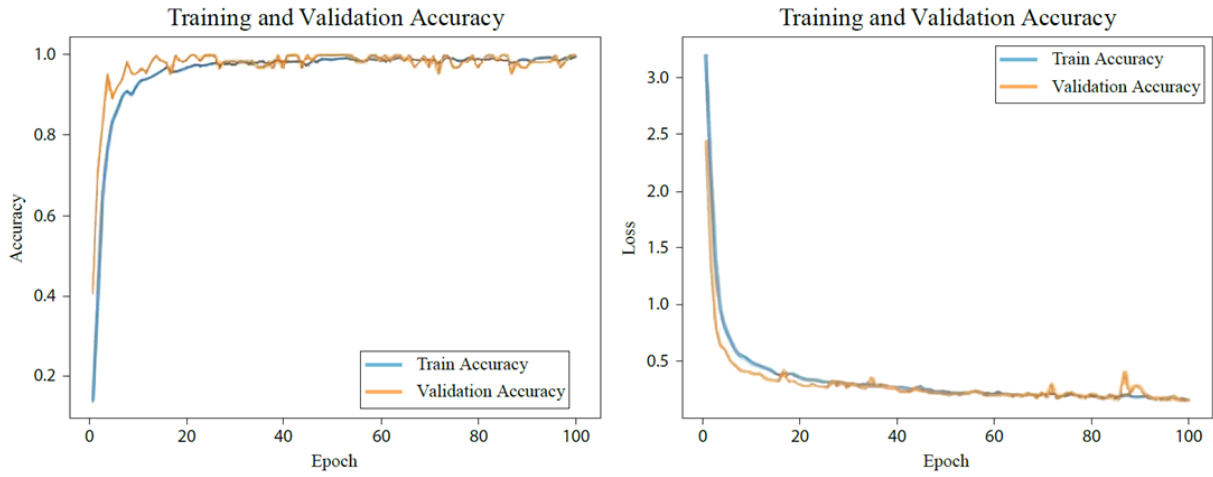


Figure 4. Learning graphs for Subject #5

In Table 3, we report train and test times and classification accuracy for each subject. These results demonstrate that the model achieves consistent and robust accuracy, indicating its suitability for gesture recognition tasks using sEMG signals from NinaProDB1. Most of the accuracies are satisfactory, with a mean test accuracy over subjects of 95.31%. The worst model is that of Subject #4 with an accuracy of 89.06%. The best model is that of Subject #5 with a perfect test classification. Overall, the models generalize well to unseen gestures, as indicated by the small difference between train and test accuracies of each subject.

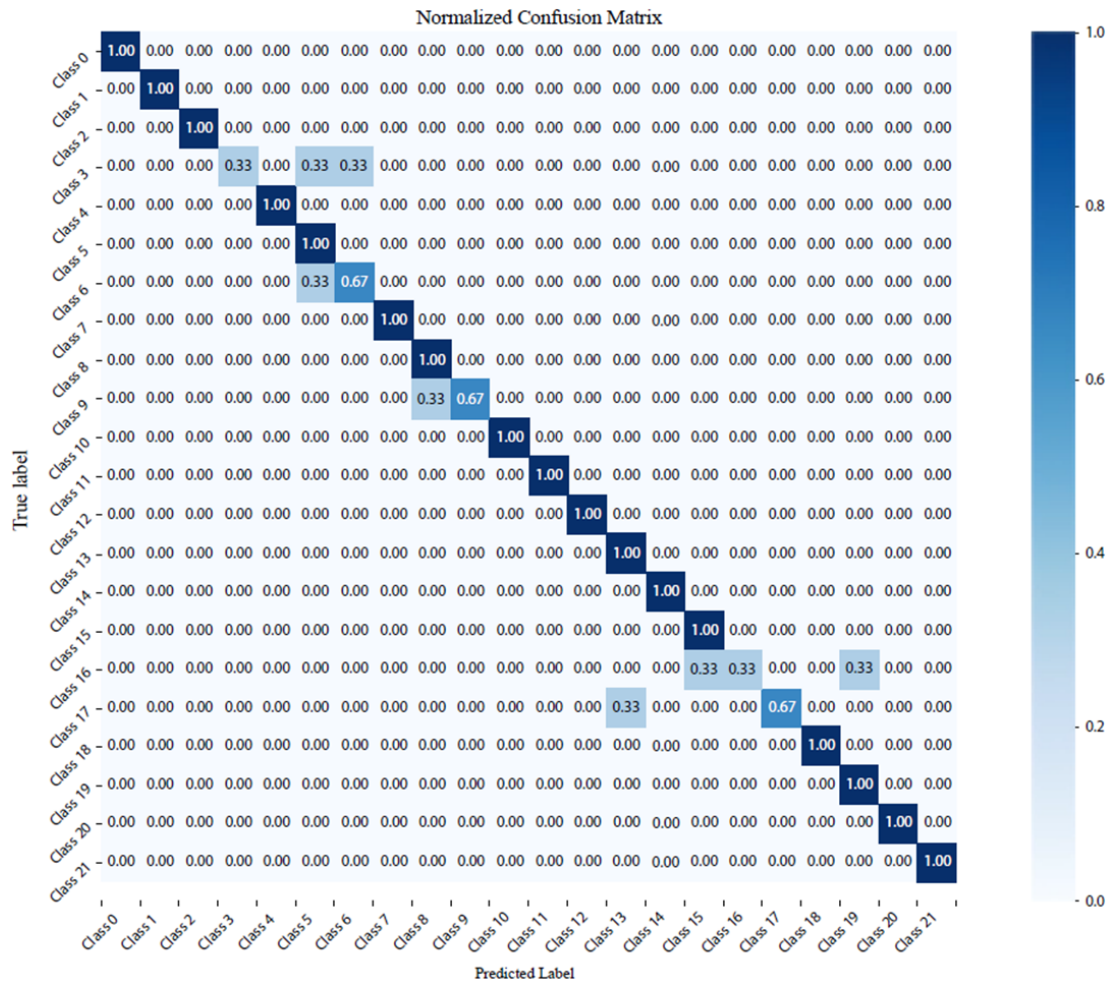


Figure 5. Confusion matrix for Subject #4

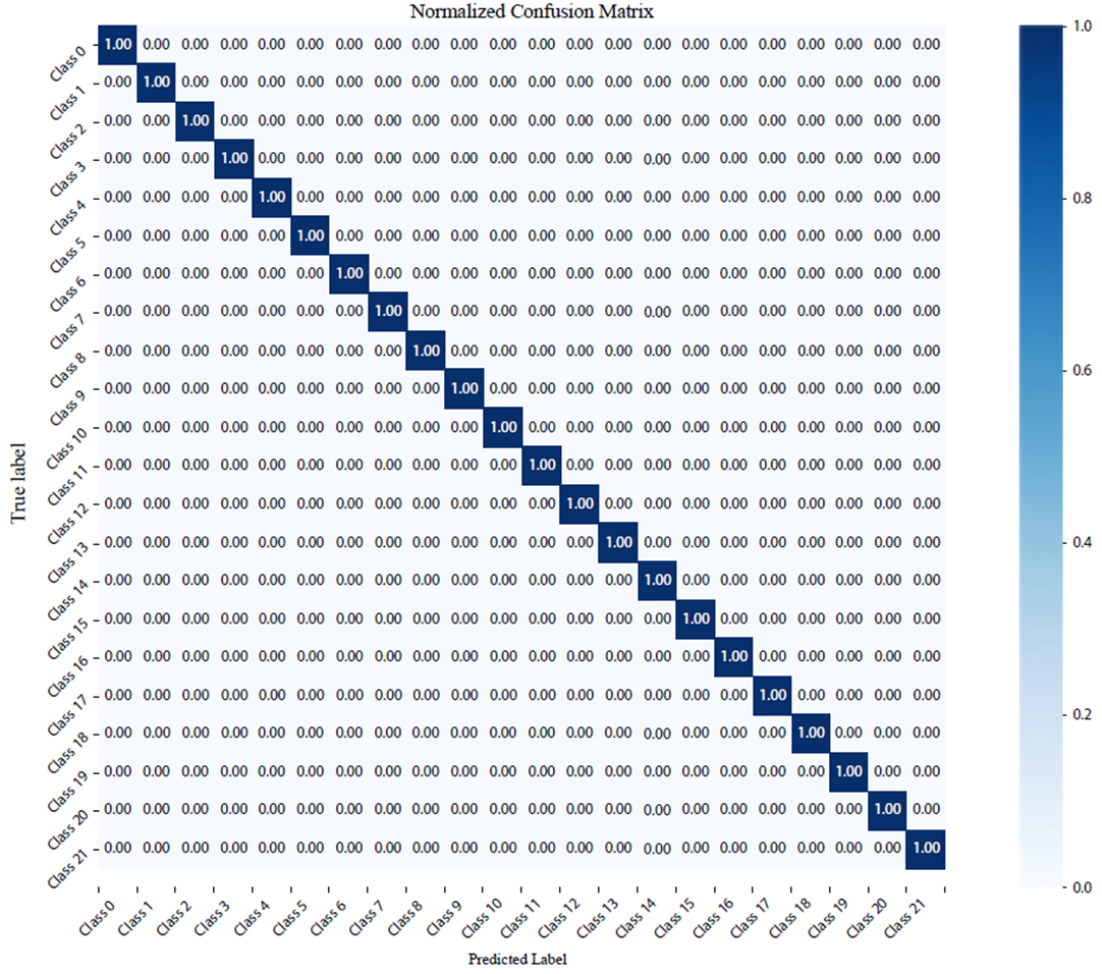


Figure 6. Confusion matrix for Subject #5

In average a model takes a bit less than 15 minutes (834.70s) for training and requires almost 0.43 seconds for recognizing a hand gesture during test. Moreover, our developed model is lightweight, with a total of 84.758 parameters, all of which are trainable, occupying just 331.09KB storage memory. This compact design is well-suited for embedding, enabling real-life applications such as controlling myoelectric prosthesis. To this end, however, deeper investigation should be conducted in this context.

To gain a deeper insight into the experimental results, we depict the training graphs for the best and worst performing subjects in Figure 3 and Figure 4 as well as their corresponding confusion matrices in Figure 5 and Figure 6. The learning graphs of the worst subject (Figure 3) demonstrate an overfitting, as depicted in the divergence between the train and validation curves for both the loss and the accuracy. In contrast, the learning graphs of the best subject (Figure 4) demonstrate the generalization of the model, as depicted by an almost perfect match between the train and validation curves.

Since the test accuracy of the best subject is 100%, a perfect confusion matrix is achieved as shown in Figure 5. The classification of a number of gestures is mismatched in case of the worst subject as shown in Figure 4, reducing the overall classification accuracy to 89.06%. Specifically, some EMG signals of five gestures (3, 7, 9, 16 and 17) were misclassified with different rates, while the remaining gestures are perfectly recognized.

4.3 Discussion

Directly comparing EMG-based hand gesture recognition (HGR) models remains challenging due to differences in experimental settings, including the number of subjects, gesture classes, recording sessions, sensor configurations, and preprocessing pipelines. These factors can significantly influence reported performance, making absolute comparisons less straightforward. Despite these challenges, when evaluated against the most relevant works reviewed in Section 2, our approach achieves an average classification accuracy of 95.31% on the NinaPro DB1 dataset, which is competitive with and in several cases surpasses recent baselines.

The combination of Temporal Convolutional Networks with residual connections, attention mechanisms, and extensive data augmentation appears to be particularly effective in enhancing temporal feature modeling and gener-

alization. In addition to accuracy gains, our framework demonstrates greater robustness across different subjects, as reflected by reduced variance in inter-subject performance compared to baselines such as CNN and RNN-based models. This stability is largely attributable to residual connections and extensive data augmentation, which mitigate overfitting and improve generalization, an essential property for practical deployment in prosthetic control and human-computer interaction.

5 Conclusions

In this paper, we explored the development of a deep learning framework tailored for surface EMG-based hand gesture recognition, with a particular focus on the use of Causal Convolutional, which present the advantage of its ability to model long-term dependencies, maintain temporal dependence over layers, and support efficient parallel training. We applied a set of preprocessing steps to structure the dataset for optimal learning and introduced our proposed architecture, RTCAN. RTCAN integrates residual causal blocks with contextual attention to enhance the extraction of meaningful EMG temporal features for robust hand gesture recognition. We evaluated our proposed approach on the NinaPro DB1 dataset, where the obtained results demonstrate strong classification performance, with most gestures correctly recognized across different subjects. In the future, we intend to conduct deeper analysis of the misclassified gestures and low performing subjects. Moreover, we will investigate subject independent evaluation of our proposed system. Extending the number of gestures and subjects will inform how our system scales. Another promising area is the implementation of the model as a real-time systems and testing its performance on embedded and low-power devices to determine its practicality for wearable prosthetic applications.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgment

The authors would like to thank the NinaPro project team for creating and maintaining the NinaPro DB1 dataset and for making it publicly available for research purposes. The authors are also grateful to the contributors whose open-source code and tools facilitated the development and evaluation of this work.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Jaramillo-Yáñez, M. E. Benalcázar, and E. Mena-Maldonado, “Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review,” *Sensors*, vol. 20, no. 9, p. 2467, 2020. <https://doi.org/10.3390/s20092467>
- [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 noninvasive naturally controlled robotic hand prostheses,” *Sci. Data*, vol. 1, p. 140053, 2014. <https://doi.org/10.1038/sdata.2014.53>
- [3] M. A. Oskoei and H. Hu, “Support vector machine-based classification scheme for myoelectric control applied to upper limb,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, 2008. <https://doi.org/10.1109/TBME.2008.919734>
- [4] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, “Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation,” *Sensors*, vol. 17, no. 3, p. 458, 2017. <https://doi.org/10.3390/s17030458>
- [5] T. Varrecchia, C. D’Anna, M. Schmid, and S. Conforto, “Generalization of a wavelet-based algorithm to adaptively detect activation intervals in weak and noisy myoelectric signals,” *Biomed. Signal Process. Control*, vol. 58, p. 101838, 2020. <https://doi.org/10.1016/j.bspc.2019.101838>
- [6] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv preprint arXiv:1803.01271*, 2018. <https://arxiv.org/abs/1803.01271>
- [7] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, “Temporal convolutional networks for action segmentation and detection,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*, 2017, pp. 1003–1012. <https://doi.org/10.1109/CVPR.2017.113>
- [8] M. Atzori, M. Cognolato, and H. Müller, “Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands,” *Front. Neurobot.*, vol. 10, p. 9, 2016. <https://doi.org/10.3389/fnbot.2016.00009>
- [9] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, “Gesture recognition by instantaneous surface EMG images,” *Sci. Rep.*, vol. 6, p. 36571, 2016. <https://doi.org/10.1038/srep36571>

- [10] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Improved gesture recognition based on sEMG signals and TCN," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, U.K., 2019, pp. 1169–1173. <https://doi.org/10.1109/ICASSP.2019.8683239>
- [11] M. Zanghieri, S. Benatti, V. Kartsch, A. Burrello, F. Conti, and L. Benini, "Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 244–256, 2020. <https://doi.org/10.1109/TBCAS.2019.2959160>
- [12] E. Rahimian, S. Zabihi, A. Asif, D. Farina, S. F. Atashzar, and A. Mohammadi, "Hand gesture recognition using temporal convolutions and attention mechanism," in *ICASSP 2022—IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, 2022, pp. 1196–1200. <https://doi.org/10.1109/ICASSP43922.2022.9746174>
- [13] J. Shin, A. S. M. Miah, S. Konnai, I. Takahashi, and K. Hirooka, "Hand gesture recognition using sEMG signals with a multi-stream time-varying feature enhancement approach," *Sci. Rep.*, vol. 14, p. 22061, 2024. <https://doi.org/10.1038/s41598-024-72996-7>
- [14] M. Montazerin, S. Zabihi, E. Rahimian, A. Mohammadi, and F. Naderkhani, "ViT-HGR: Vision transformer-based hand gesture recognition from high-density surface EMG signals," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Glasgow, U.K., 2022, pp. 5115–5119. <https://doi.org/10.1109/EMBC48229.2022.9871489>
- [15] W. Zhong, Y. Zhang, P. Fu, W. Xiong, and M. Zhang, "A spatio-temporal graph convolutional network for gesture recognition from high-density electromyography," in *2023 29th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, Queenstown, New Zealand, 2023, pp. 1–6. <https://doi.org/10.1109/M2VIP58386.2023.10413402>
- [16] Y. Xiang, W. Zheng, J. Tang, Y. Dong, and Y. Pang, "Gesture recognition from surface electromyography signals based on the SE-DenseNet network," *Biomed. Eng./Biomed. Tech.*, vol. 70, no. 3, pp. 207–216, 2025. <https://doi.org/10.1515/bmt-2024-0282>
- [17] F. Palermo, G. Rossi, G. D. Marchis, P. Artemi, F. Giovacchini, and N. Vitiello, "Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data," in *2017 International Conference on Rehabilitation Robotics (ICORR)*, London, U.K., 2017, pp. 1154–1159. <https://doi.org/10.1109/ICORR.2017.8009405>
- [18] Y. J. Luwe, C. P. Lee, and K. M. Lim, "Wearable sensor-based human activity recognition with hybrid deep learning model," *Informatics*, vol. 9, no. 3, p. 56, 2022. <https://doi.org/10.3390/informatics9030056>
- [19] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances," *Sensors*, vol. 22, no. 4, p. 1476, 2022. <https://doi.org/10.3390/s22041476>
- [20] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, San Diego, CA, USA, 2016, pp. 1480–1489. <https://doi.org/10.18653/v1/N16-1174>