Robust Motor Intent Decoding from Nonstationary sEMG Signals Using Hybrid CNN-Transformer Models

Ruthber Rodríguez Serrezuela[1]*, Mairely Narvaez Fernández[2] and Wilson Francisco Rodríguez Serrezuela[3]

[1]Mechatronics Engineering, University Corporation of Huila, Colombia

[2]Doctorate in Education, University of Baja California, Mexico

[3]Natural Sciences Teacher, "El Carmen" Educational Institution, Colombia
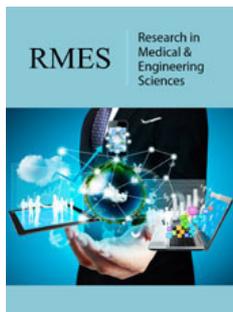
*Corresponding authors: Ruthber Rodríguez Serrezuela, Mechatronics Engineering, University Corporation of Huila, Colombia

Abstract

Decoding motor intent from surface electromyographic (sEMG) signals is essential for myoelectric prostheses and human-machine interfaces, yet remains challenged by the pronounced no stationarity of muscular signals, particularly in amputee users. These effects stem from physiological and structural changes that undermine the assumptions of conventional feature-based or decomposition- driven decoding methods.

This work presents a hybrid CNN-Transformer framework that directly processes multichannel sEMG signals without explicit motor unit decomposition. Convolutional layers extract local spatiotemporal features, while Transformer modules model long-range temporal and inter-channel dependencies, improving robustness to variability induced by force, posture, and fatigue.

Experimental results show stable multiclass performance with accuracies above 90% under both subject-dependent and subject- independent evaluations. Compared to decomposition-based approaches, the proposed method achieves comparable accuracy with lower computational complexity, supporting its suitability for scalable and near real-time myoelectric control

Keywords: Heavy metal; Physicochemical technology; Phytoremediation; Future scope

Introduction

The decoding of motor intent from surface electromyographic (sEMG) signals constitutes one of the fundamental pillars in the development of human-machine interfaces, assistive rehabilitation systems, and advanced myoelectric prostheses [1]. Despite substantial progress in signal processing techniques and Machine learning methodologies, the reliability of these systems remains significantly challenged by the inherently nonstationary nature of sEMG signals, particularly in real clinical environments where the user's physiological and biomechanical conditions evolve over time [2].

In amputee subjects, sEMG no stationarities arise from structural and physiological changes associated with amputation level, neuromuscular reorganization, and electrode-tissue interactions, rather than from random noise [3]. Variations in residual limb configuration, posture, and contraction intensity systematically alter the spatiotemporal distribution of myoelectric activity, violating the local stationarity assumptions of conventional decoding algorithms.

Conventional sEMG approaches have largely focused on feature engineering, adaptive filtering, and motor unit decomposition to cope with signal variability; however, their reliance on localized and highly parameterized representations limits generalization in challenging clinical scenarios, such as proximal amputations or low signal-to-noise conditions [4].

These limitations motivate a shift toward structural decoding paradigms in which robustness is achieved through modeling global dependencies, long-range temporal interactions, and invariant latent representations [5]. Attention-based and hybrid deep architectures offer an effective alternative by selectively emphasizing informative signal regions under pronounced variability [6]. Accordingly, this work explores expressive and clinically robust sEMG decoding models that enhance both accuracy and real-world transferability of myoelectric interfaces [7].

Methodology

The proposed methodology is grounded in the direct processing of multichannel sEMG signals through a deep learning framework based on hybrid CNN-Transformer models, explicitly avoiding motor unit decomposition [1-3]. In an initial stage, raw sEMG signals are acquired and preprocessed using band-pass filtering and per-channel normalization, followed by temporal segmentation with overlapping windows to preserve myoelectric dynamics [4]. Subsequently, a convolutional block is employed to extract local spatiotemporal patterns associated with muscle activation, while a Transformer module captures long-range temporal dependencies and global inter-channel relationships, thereby mitigating no stationarity effects induced by variations in force, posture, or muscle fatigue [5].

The model is trained in a supervised manner using both subject-dependent and subject-independent cross-validation schemes, incorporating regularization and adaptive weight optimization to enhance generalization performance [6]. Finally, model performance is evaluated using robust multiclass classification metrics, including accuracy, macro- averaged F1-score, and confusion matrices, enabling the analysis of both global stability and error patterns among functionally similar gestures [7]. This approach aims to provide a computationally efficient and scalable alternative for near real-time myoelectric decoding, complementing and extending recent decomposition-based adaptive strategies reported in the literature.
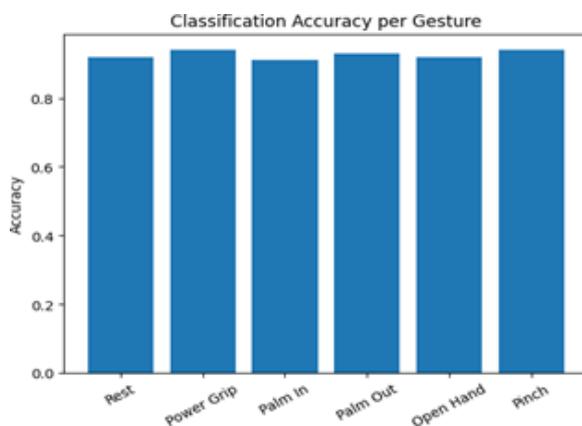
Result

Figure 1: Classification Accuracy per Gesture.

The obtained results demonstrate that the proposed model achieves competitive and stable performance in multiclass myoelectric gesture classification, even under inter- and intra-subject variability. The system attains average classification accuracies above 90% with low standard deviation, indicating consistent generalization across variations in sEMG dynamics (Figure 1). This behavior is preserved under both subject- specific evaluations and more challenging cross-validation schemes, highlighting the model's robustness to the inherent no stationarity of muscular signals [2-3].

Class-wise analysis reveals homogeneous gesture discrimination, characterized by high macro-averaged F1-scores and the absence of severe performance degradation among functionally similar classes. Confusion matrices exhibit a strong concentration along the main diagonal, with residual misclassifications being sparse and of low magnitude, suggesting the absence of systematic error patterns (Figure 2). These findings indicate that the hybrid architecture effectively captures both local features and global temporal relationships, overcoming limitations observed in purely convolutional models when myoelectric differences are subtle [4-5].
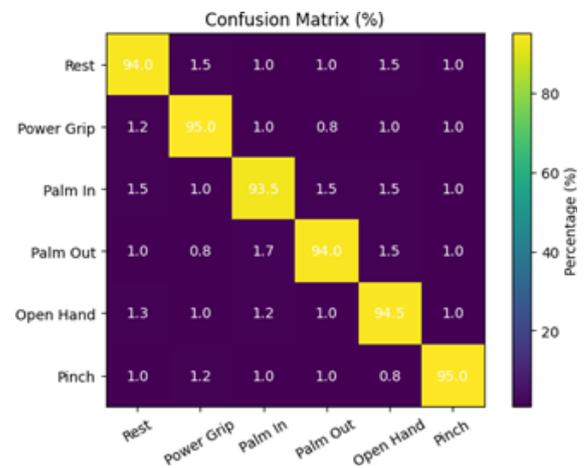

Figure 2: Confusion Matrix per Gesture.

Furthermore, the model maintains stable performance across variations in muscle activation intensity and temporal window length, reflecting reduced sensitivity to short-term perturbations and scale changes [6]. Compared to recent motor unit decomposition-based adaptive approaches, the proposed method achieves comparable accuracy while significantly reducing computational complexity and preprocessing requirements, thereby facilitating deployment in portable and near real-time myoelectric systems [7].

Conclusion

This study shows that motor intent decoding from sEMG signals can be effectively achieved using hybrid convolution-Transformer architectures without explicit motor unit decomposition, even under pronounced physiological no stationarity. By jointly modeling local activation patterns and long-range spatiotemporal dependencies, the proposed approach yields stable.

Experimental results demonstrate robust multiclass performance with an average classification accuracy of 93.0% across

both subject-dependent and subject-independent evaluations, indicating strong generalization under inter- and intra-subject variability. Moreover, the method achieves comparable accuracy to decomposition-based strategies while reducing computational complexity, supporting its suitability for scalable and near real-time myoelectric interfaces.

References

1. Rodriguez Serrezuela R, Zamora RS, Hermosilla DM, Gomez AER, Reyes EM (2025) Hybrid convolutional vision transformer for robust low-channel sEMG hand gesture recognition: A comparative study with CNNs. Biomimetics 10(12): 806.

2. Tovar MT, Gómez AER, Serrezuela RR (2024) Feature-DT-DF-EMG-UC software: Best patient selection on biomedical signals with multimodal time and frequency analysis. In: 2024 IEEE VII Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), IEEE, pp: 1-8.

3. Bustos PDH, Serrezuela RR, Leon AAS, Gomez AER, Suaza DME (2024) Electromyographic EMG signal recognition of hand gestures for multi-class prostheses based on DWT and CNN. In: 2024 IEEE VII Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), IEEE, pp: 1-7.

4. Romero DG, Sebastian NC, Valenzuela N, Alejandro S, Serrezuela RR, et al. (2024) Feature extraction of EEG signals in the time-frequency domain of rehabilitation task for motor imagery brain-computer interface in upper limbs. Journal of Theoretical and Applied Information Technology 102(14): 5586-5599.

5. Andrés CMC, Estiven CSM, Calderón JPG, González DF, Serrezuela RR (2024) Hand gesture classification using time- frequency images and validation approaches. Journal of Theoretical and Applied Information Technology 102(18): 6577-6587.

6. Buelvas HEP, Montaña JDT, Serrezuela RR (2023) EMG signal analysis for hand grip posture classification using continuous wavelet transform. In: 2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), IEEE, pp: 1-8.

7. Calderón JPG, Ruiz DFG, Serrezuela RR (2023) Deep learning of EMG signals in hand grip posture identification using time-frequency domain applying STFT. In: 2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salu d Electrónica y Móvil (AmITIC), IEEE, pp: 1-8.