2024 21st International Conference on Ubiquitous Robots (UR)
June 24 - 27, 2024, New York University, USA

A Hybrid CNN-LSTM Network with Attention Mechanism for Myoelectric Control in Upper Limb Exoskeletons

Paniz Sedighi, Amr Marey, Ali Golabchi, Xingyu Li, and Mahdi Tavakoli

Abstract— **This paper introduces a novel attention-based sequence-to-sequence network for predicting upper-limb exoskeleton joint angles, enhancing the control of assistive technologies for individuals with upper limb impairments. By integrating EMG and IMU signals, our model facilitates real-time decoding of user intentions, generating precise movement trajectories for a 3-DoF cable-driven upper-limb exoskeleton. The implementation of an attention mechanism within an encoder-decoder architecture allows for the dynamic prioritization of the most pertinent EMG features and historical angular positions. Our model significantly outperforms existing EMG-based hand motion prediction methods in terms of prediction accuracy and responsiveness as demonstrated in evaluations. This approach not only offers a tailored response to varying sequence lengths and compensates for sensor unreliability but also introduces a level of generalization and adaptability previously unattainable in robotic rehabilitation and assistive devices. The implementation of the attention mechanism allows for adaptive learning, focusing on the most relevant signals for each user, thereby enhancing the system's ability to learn and predict complex movement trajectories.**

I. Introduction

Robotic exoskeleton rehabilitation is widely used for patients with stroke or spinal cord injuries who experience upper limb motor impairments. This technology plays a crucial role in intensive rehabilitation therapy, aiding patients in regaining motor function [1]. However, the growing aging population and the escalating demand for therapy sessions present challenges. Often, there are not enough resources to provide adequate treatment. This shortage can lead to diminished therapy quality due to an extensive workload for the therapist [2]. Additionally, there's a growing need for more accurate and customized assistive technologies tailored to the unique needs of each patient. The effectiveness of assistive robotic exoskeletons is significantly influenced by the quality of human-robot interaction (HRI), which can be better assessed through the interpretation of biological signals [3]. Achieving intuitive and user-specific control in exoskeletons remains a challenge, underscoring the need for new approaches in myoelectric control systems that rely on more sophisticated processing of biosignals that define the user's interaction with an exoskeleton.

This work was supported by the Canada Foundation for Innovation, UAlberta Huawei-ECE Research Initiative (HERI), Government of Alberta, Natural Sciences and Engineering Research Council (NSERC) of Canada, Canadian Institutes of Health Research (CIHR), and Alberta Economic Development (Corresponding author: Paniz Sedighi.)

The authors are with the Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB T6G 1HQ, Canada (e-mail: sedighi1@ualberta.ca; amarey@ualberta.ca; agolabchi@ewiworks.com; xingyu@ualberta.ca; mahdi.tavakoli@ualberta.ca).

Surface Electromyogram (sEMG) signals, generated by subtle muscular contractions, have been extensively utilized in controlling myoelectric prosthetics and exoskeletons. By extracting meaningful information from EMG signals, it's possible to map these signals to intended movements [4]. Another significant advantage of using EMG is their capability to reflect motor intention approximately 200 milliseconds before the actual onset of joint movement [5].

Most studies in intention decoding with EMG sensors have historically centered on pattern recognition [6] and gesture classification [7], targeting discrete motions rather than continuous upper limb movements. This approach has limited applicability for applications requiring continuous control. The emerging field of proportional myoelectric control seeks to bridge this gap by focusing on predicting continuous movements in prosthetics. Traditional machine-learning (ML) methods rely on manually handcrafted and statistical features, which may not fully capture the nuanced information necessary for effective sensor fusion [8]. In contrast, recent advances have seen the adoption of deep learning (DL) and data-driven techniques capable of extracting comprehensive features from EMG sensors. When integrated with IMU data, these methods offer a more sophisticated approach to motion prediction, leveraging the full potential of sensor fusion inputs.

Convolutional Neural Networks (CNNs) excel at hidden spatial feature extraction, making them particularly suited for EMG signal analysis due to their ability to handle high-dimensional data. This capability is especially beneficial in scenarios involving crosstalk within multi-degree-of-freedom exoskeletons. In their work, George *et al.* [9] employed a CNN-based method that consistently predicts joint torques across various levels of assistance. Furthermore, a novel hybrid CNN–SVM model, detailed in [10], was introduced for the identification of human locomotion modes. This model utilizes multi-channel inertial measurement unit (IMU) signals for its operations.

Nonetheless, it's crucial to consider the temporal aspect of EMG signals, as they are inherently sequential data. Hybrid architectures that combine CNNs with Recurrent Neural Networks (RNNs) have demonstrated superior performance compared to standalone CNN-based approaches, as evidenced in [11]. Similarly, Paniz *et al.* [12] proposed an innovative hybrid model for intention detection in upper-limb exoskeletons. However, the performance of these models could be improved by adding more flexibility and cross-subject variability. Despite the advancements in deep learning for myoelectric control, there remains a gap in models

979-8-3503-6107-0/24/$31.00 ©2024 IEEE

that can effectively integrate real-time adaptability and user-specific control, which this research aims to address.

Attention networks have significantly impacted sequence-to-sequence networks by enabling models to selectively focus on different segments of the input sequence for each part of the output [13]. Zhang *et al.* [14] proposed an attention-based Kalman Filter (KF) scheme to decode the upper body movements. Zhu *et al.* [15] used a hybrid CNN-LSTM model to estimate the joint positions for the lower-limb exoskeleton. In another study [16], an attention-based Deep CNN-BiLSTM Model was used to detect muscle fatigue.

This paper introduces a unique hybrid CNN-LSTM network with an attention mechanism for addressing the limitations of existing models by offering enhanced precision and adaptability in predicting upper limb movements. By integrating real-time EMG and IMU signal decoding, the model generates an optimal movement trajectory for a 3-DoF cable-driven soft upper-limb exoskeleton, factoring in historical joint position data. The model's encoder-decoder framework, enhanced with an attention mechanism, adeptly identifies and prioritizes the most crucial spatial and temporal EMG signal features for accurate position prediction. This is further refined through a PD position controller that incorporates feedback from rotary encoders.

The model's attention mechanism significantly increases its adaptability, generalizability, and flexibility, enabling it to efficiently handle input sequence variations—a critical feature for dynamic industrial tasks. It also remains effective when some sensors are offline, showcasing a robust design suitable for practical applications where sensor reliability may be an issue. Importantly, the model's ability to focus on user-specific data points enhances adaptability, ensuring a more intuitive and safe operation of the exoskeleton in rehabilitation and industrial environments.

This paper is arranged as follows: Section II introduces the framework of the encoder-decoder architecture, followed by a discussion of the position controller in Section III. The experimental methodology, designed to evaluate the effectiveness of our proposed method is explained in section IV, along with the resulting quantitative analyses and findings, which are presented subsequently in section V. The article concludes with Section VI.

II. Trajectory Prediction with Myoelectric Attention

In this section, we describe how we leverage a sequence-to-sequence model to generate and predict future joint positions of a 3-DoF robot within the context of the movement. The proposed model utilizes an encoder-decoder structure. The purpose of the encoder is to reduce the dimensionality of the raw EMG and IMU data and map it into a feature map. The purpose of the decoder is to generate the predicted future joint positions.

The attention scheme would allow the decoder to selectively focus on certain features of the input by selecting a subset of all the feature vectors. We denote vectors with bold font and matrices with capital letters.

A. Encoder: Convolutional Features

The encoder section of our proposed seq2seq CNN-LSTM model shown in Fig. 1 starts with the input data x of a time window of length N across 21 channels, which it then maps onto the encoder state vectors. This model employs three 2D convolutional layers to encode the input into an X-dimensional feature map a with a length of K, as represented by Equation (1). Each feature vector $\mathbf{a}_j, j = 1, \ldots, K$ corresponds to the features extracted at different parts of the window.

$$a = \{\mathbf{a}_1, \ldots, \mathbf{a}_K\}, \mathbf{a}_j \in \mathbb{R}^X \tag{1}$$

The predicted output is a sequence of angular encoder positions with a length of T and a dimension of the DoF D.

$$y = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}, \mathbf{y}_i \in \mathbb{R}^D \tag{2}$$

The CNN layers are designed to explore correlations both temporally and across channels, thereby reducing the dimensionality of the feature map and the overall complexity of the model. The input channels are strategically arranged to place similar muscle groups -later explained in IV-A- in proximity, enhancing the model's ability to extract relevant features in the initial layers.

To mitigate the effect of noise in the feature map, average pooling is implemented in the shallower layers. The convolutional layers are followed by batch normalization and ReLU activation functions, ensuring non-linearity and efficient training. A dropout probability is applied before the feature map is relayed to the decoder, aiding in the prevention of overfitting and enhancing the model's generalization ability. This intricate arrangement of layers and functions within the CNN encoder not only optimizes the extraction of meaningful features from EMG and IMU data but also lays a robust foundation for the subsequent LSTM-based decoding process.

The first kernel size (3,30), is designed to encompass a broader context in the time dimension, allowing the model to integrate information over longer time sequences, which is essential for understanding the evolution of muscle signals. Conversely, the smaller dimensions of the subsequent kernel sizes, (3,5) and (4,2), allow the model to focus on more fine-grained and localized features within the EMG data. This hierarchical approach in kernel sizing—from capturing broad temporal patterns to zooming in on detailed features—enables the model to construct a comprehensive understanding of the EMG signals, ultimately leading to more accurate and responsive control of the exoskeleton.

Overall, the encoder's use of convolutional layers to process EMG and IMU data into a feature map is for reducing input dimensionality while preserving crucial spatial-temporal relationships. This prepares the model for more accurate prediction by focusing on the most relevant features for trajectory determination.

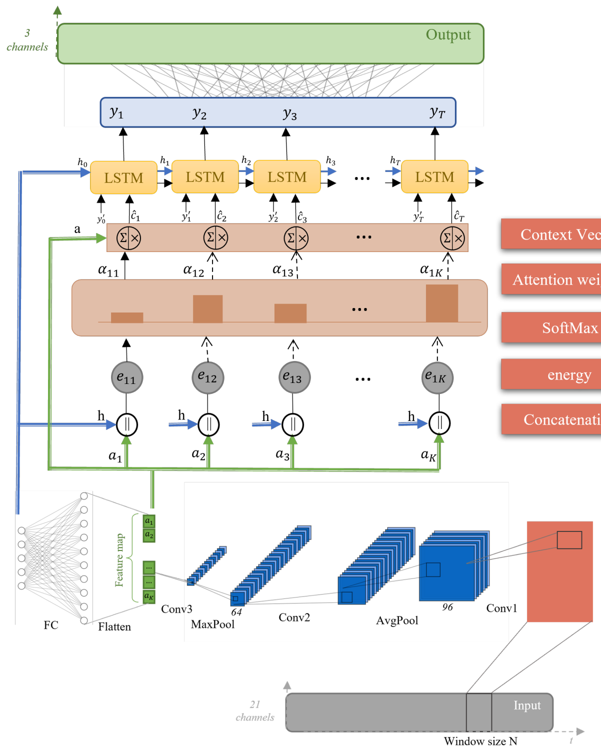

Fig. 1. Structure of the attention-based CNN-LSTM encoder-decoder network used for joint position prediction.

B. Decoder: Long-Short Term Memory

In this model, the decoder plays a critical role in processing the feature map generated by the Encoder, along with the previous hidden and cell states, to predict the angular position in a time sequence. The input to the decoder, denoted as $\mathbf{y}_{i'-1}$, can be either the target trajectory of the model or the previously generated trajectory. This input is initially transformed via a linear embedding $E\mathbf{y}_{i-1}$ to align it with the dimensions of the feature vectors. The single-layer LSTM then updates its hidden state $\mathbf{h}_i$ for each time step i using the formula:

$$\mathbf{h}_i = f\left(\mathbf{h}_{i-1}, \mathbf{y}_{i'-1}, \hat{\mathbf{c}}_i\right) \tag{3}$$

In (4), $\hat{\mathbf{c}}_i$ is the context vector for each embedding of y_i. This context vector plays a crucial role in the Decoder. It treats each feature vector as a token that encapsulates information about the entire input sequence, with a particular emphasis on the parts surrounding the j-th vector. As shown in Fig. 1, the context vector is calculated as a weighted sum of these feature vectors.

$$\hat{\mathbf{c}}_i = \sum_{j=1}^{K} \alpha_{ij}\mathbf{a}_j \tag{4}$$

The weights α_{ij} for each feature vector are generated by the attention mechanism, which determines the relevance or 'attention' each vector should receive when predicting the next embedding of y_i. The i-th context vector $\hat{\mathbf{c}}_i$ represents the expected energy over all the energies, weighted by the probabilities α_{ij}. For implementation, the *softmax* of the energy score is calculated.

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{K}\exp(e_{ik})} \tag{5}$$

The energy e_{ij} scores the similarity between the inputs around position j and the output at position i (6). Unlike traditional attention mechanisms that use dot products for this calculation, this model employs the ReLU of a feed-forward neural network, f_{att}, for determining the energy. The network's parameters are trained alongside the other components of the system.

$$e_{ij} = f_{\text{att}}\left(\mathbf{h}_{i-1}, \mathbf{a}_j\right) \tag{6}$$

Also, the RNN's 400 cell and hidden states were initialized with two separate multi-layer perceptrons by inputting the flattened encoder states before sending them to the decoder. Some related works have used the averaged encoder states instead.

$$\begin{aligned} \mathbf{c}_0 &= f_{\text{init, c}}\left(Dense(\mathbf{a})\right) \\ \mathbf{h}_0 &= f_{\text{init,h}}\left(Dense(\mathbf{a})\right) \end{aligned} \tag{7}$$

In our model, the teacher forcing method is employed to enhance the learning process, where the decoder is occasionally fed the actual target outputs instead of its predictions. This technique aids in stabilizing training and improves the model's ability to generate accurate sequences by aligning the training phase more closely with the inference phase, thus effectively mitigating discrepancies between the two.

In the decoder phase, the attention mechanism's role is critical for dynamically focusing on specific features within the extensive feature map generated by the encoder. By weighting these features based on their relevance to the current prediction task, the model achieves a higher level of precision in predicting joint positions, demonstrating the practical benefits of attention in managing sequential data.

C. Training

The extensive training process, utilizing a GPU for over 15 hours across 35 epochs, not only underscores the complexity of the model but also its efficiency in learning from the data. The optimization achieved through careful adjustment of the learning rate and dropout rates exemplifies the meticulous approach taken to enhance model performance. It utilized a batch size of 32 and a learning rate set at 3×10^{-4}, aiming for a future prediction span of 120 ms. The feature map was structured with dimensions [200, 5, 2], while the LSTM architecture featured a single layer containing 400 hidden units, complemented by an embedding size of 200. Experiments with the dropout rate for both the encoder and decoder ranged from 0.2 to 0.5, with optimal results achieved at a dropout rate of 0.2. Furthermore, the teacher-forcing ratio was determined to be most effective at 0.4.

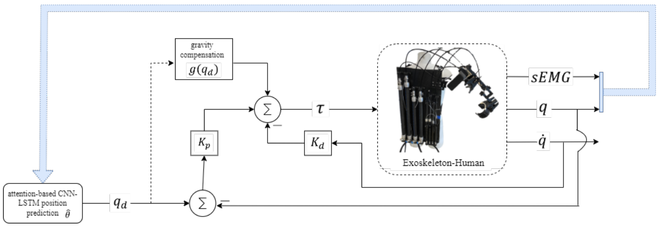

Fig. 2. The upper-limb exoskeleton's intention-based model control schematic, which uses proportional derivative (PD) for low-level control and the encoder-decoder network for high-level control while gravity compensated.

In summary, the integration of an attention mechanism within a hybrid CNN-LSTM framework marks a substantial advancement in trajectory prediction for upper limb exoskeletons. This model not only enhances prediction accuracy and system responsiveness but also introduces a degree of adaptability and personalization previously unattainable, setting a new standard for myoelectric control technologies.

III. Robot-assisted Control

For the trajectory tracking controller, we aim to follow the predicted position, accounting for challenges like delays and internal friction inherent to the pneumatic cable-driven system. Employing a Proportional-Derivative (PD) control scheme, as depicted in Fig. 2, we ensure the robotic exoskeleton accurately follows the desired joint angles. The controller uses a PD control scheme in 8 with proportional K_P and derivative K_D gains to track the desired joint angle q_{d_i} at each time step, complemented by feedforward gravity compensation τ_{GCi} that leverages the system's known kinematics for enhanced control performance.

$$\tau_{ri} = K_P(q_{d_i} - q_i) - K_D\dot{q}_i + \tau_{GCi} \tag{8}$$

IV. Experiments

A. Hardware Setup

The myoelectric control neural network alongside the gravity compensation neural network was implemented using Python (Python Software Foundation, USA). The angular positions of the exoskeleton joints were captured through an Arduino Mega 2560 using the Pyserial library, which allows for bidirectional data communication between the computer running Python and the Arduino. The angular position data was sampled at a frequency of 500 Hz. The EMG and IMU data were sampled through the Delsys Trigno Research+ system at a sampling frequency of 2000 Hz. After the input data has been captured through the different data streams, the data would inputted to the control and gravity compensation neural networks. These networks would regulate the pressure of the pneumatic soft actuators attached to the exoskeleton through the usage of another Arduino Mega 2560 communicating with Python.

The fluidic muscles DMSP-20-RM-CM from Festo Corporate in Esslingen, Germany, powered the joints of the exoskeleton. Omega electro-pneumatic transducers EP211-X120-10V from Omega Engineering Inc. in the USA were employed to regulate the pressure of these pneumatic soft actuators. Quadrature optical encoders (HEDM-5500 B12, Broadcom Inc., US) were attached to the shoulder and elbow joints to measure the exoskeleton's position. Three Trigno Avanti Sensors were positioned along the 1- biceps brachii, 2- anterior deltoid, and 3- medial deltoid to gather surface EMG and IMU data from each muscle. Six Trigno Avanti Sensors were positioned along the 4- brachialis, 5- brachioradialis, 6- pectoralis major (Clavicular Head), 7- triceps, 8- deltoideus posterior, and 9- trapezius descendens muscles to gather surface EMG data only [17], [18].

B. Experimental Setup

In this study, our attention-based network was trained using a dataset collected from seven individuals, aged 21 to 35 (including 3 females), over the course of a week. All participants were able-bodied and provided written informed consent. Prior to the experiments, we thoroughly explained the procedures to each participant.

The experimental protocol involved seven types of upper-limb movements, each integrating a combination of three joint actions. The first three movements comprised ascending and descending the elbow, shoulder abduction/adduction, and shoulder flexion/extension. The next three involved two of the same three joint actions, and the last one involved all three at the same time. To add complexity, these movements were performed under two conditions: without load and with a 1kg load, resulting in a total of 14 distinct tasks. Each task was repeated five times to ensure consistency and to introduce natural variations in muscle activity due to fatigue.

Each task was completed within an average time range of 4 to 17 seconds. Participants were encouraged to perform these tasks in a manner they found most ergonomic and comfortable. This approach allowed for individual variations in task execution; for instance, the participants were allowed to choose the pace or lower any joint earlier than others.

The training and testing datasets were partitioned randomly according to movements, adopting an 11:3 ratio for the split. This procedure entailed conducting five distinct training sessions, during each of which three randomly chosen movements from all users were allocated for testing. This approach, termed combined training, utilized the entire dataset inclusively. Alternatively, in a different evaluation strategy, we individually segmented each participant's dataset using an 80:20 ratio for training and testing purposes, respectively. This alternative method also involved executing five training iterations, but in this scenario, we randomly selected one out of the seven participants for each iteration.

Four deep learning models—Bi-LSTM, CNN-Bi-LSTM, CNN-LSTM, and the attention-based CNN-LSTM—were trained with input sequence lengths of 150, 250, and 390. The Bi-LSTM model enhances the capabilities of the unidirectional LSTM by simultaneously learning from both forward and backward contextual relationships in the activities and the input surface electromyography (sEMG) signals. Ma *et al.* [19] utilized a Bi-LSTM network for predicting upper limb joint angles, utilizing data from three EMG sensors. However, a combined framework of CNN and Bi-LSTM can learn not only bi-directional temporal relationships but also spatial correlations. Karnam *et al.* [20] adopted this combined approach in their work on hand gesture recognition, employing EMG signal classification.

In our study, we adapt the structure of their model to assess its applicability in our context. Additionally, we conducted a comparative analysis with our previously developed CNN-LSTM model, which was trained on a dataset from a single individual. The training was conducted in two scenarios: first, using data from a single random participant, and second, using combined data from all participants. In both cases, model performance was evaluated based on inference accuracy for the selected individual.

In the second phase of our experiments, we evaluated the performance of an attention-based CNN-LSTM model trained on a composite dataset, for predicting the exoskeleton's real-time joint positions. A predefined trajectory shown in Fig. 3, designed to engage all three joints of the exoskeleton and displayed on a computer screen alongside the robot's end effector (the user's index finger), was followed in real time at an adjusted speed. The participant was instructed to follow the online generated trajectory without deviating a tolerance boundary of a 2 cm radius. Data from EMG, IMU, and angular encoders were input into the model in windows of 250 by 21 (with a stride of 30). Participants attempted to trace this trajectory 20 times with all sensors available, 20 times with one sensor disabled, and an additional 20 times with two sensors disabled.

Apart from the initial three primary sensors detailed in Section IV-A, one or two of the subsequent six sensors were randomly disabled during each repetition's inference phase. The primary three sensors were exempt from this because they provide crucial information about the intended movement, upon which the model heavily relies for accurate prediction. This strategy tested the model's robustness to random sensor failure.

V. EXPERIMENTAL RESULTS

The experimental results are divided into two sections. In the first section, we assess the performance of four different deep learning models, focusing on their accuracy across various sequence lengths. Following this analysis, the model demonstrating the best performance is selected for further evaluation. This second phase involves testing the chosen model's inference capabilities on a robotic exoskeleton, along with an examination of its modularity and flexibility.

A. Performance Evaluation

The models were trained once on combined data from all participants and once on a randomly selected participant's data. The loss is calculated on a test data set from the randomly selected individual using Mean Squared Error (MSE). Due to computational complexity and limitations in GPU resources, we were unable to test the IBPA and CNN Bi-LSTM models with a 390 ms sequence length.

TABLE I
EVALUATION RESULTS OF FOUR DEEP LEARNING MODELS ACROSS THREE SEQUENCE LENGTHS.

Models	Sequence length	Combined dataset	Individual dataset
Bi-LSTM [19]	150ms	1.72 ± 0.11	1.68 ± 0.09
	250ms	1.55 ± 0.34	1.40 ± 0.17
	390ms	1.42 ± 0.10	1.29 ± 0.26
CNN Bi-LSTM [20]	150ms	1.36 ± 0.14	1.23 ± 0.18
	250ms	1.36 ± 0.16	1.34 ± 0.22
	390ms	N/A	N/A
IBPA (parallel CNN-LSTM) [12]	150ms	1.50 ± 0.15	1.28 ± 0.14
	250ms	1.45 ± 0.17	1.36 ± 0.19
	390ms	N/A	N/A
Attention-based CNN-LSTM	150ms	1.28 ± 0.09	1.20 ± 0.12
	250ms	0.97 ± 0.22	0.89 ± 0.15
	390ms	1.15 ± 0.25	1.08 ± 0.09

Table I distinctly illustrates that the attention-based CNN-LSTM model achieves the highest overall accuracy when utilizing a 250 ms sequence length for input. This superior performance is maintained when the model is trained exclusively on data from a single user and then applied to the same individual. This model also demonstrates robust performance when trained on a dataset comprising multiple users but tested on some unseen movements from the same users, thereby affirming its generalized and flexible nature.

Regarding the impact of sequence length, while the attention mechanism does enhance the model's capability to process longer sequences, this does not automatically translate into higher accuracy for longer sequences as opposed to shorter ones. In our findings, a 250 ms sequence length emerges as the optimal choice for the attention-based model.

When the models are trained using a combined dataset from multiple users, a noticeable decline in accuracy is observed across all four model types. However, it is evident that the attention mechanism is more adept at managing variations in users' muscle activities compared to the other models, showcasing its relative robustness in this aspect.

B. Modularity

We conducted tests to assess the models' adaptability and robustness under sensor failure conditions. Specifically, the trained attention-based CNN-LSTM model, with data from all participants, was evaluated. During these tests, participants were instructed to follow a predefined trajectory, displayed on a computer screen. Deviation beyond this boundary was marked as a failure. Each movement was repeated 20 times at a consistent speed with all sensors working. We then conducted this experiment in two scenarios: once with the output from one EMG sensor set to zero (simulating sensor failure), and once with two sensors deactivated. On average, each task repetition lasted 9.5 seconds, with a standard deviation of 0.01 seconds.

Figure 3 presents the Root Mean Square (RMS) path for each scenario, averaged across all repetitions and illustrated with dashed lines, alongside the predicted path. With

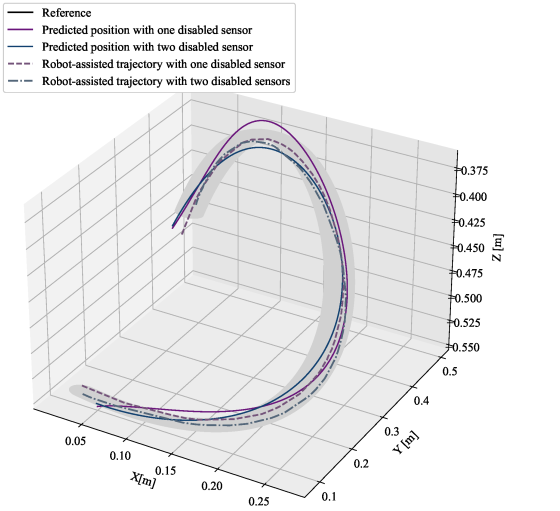

Fig. 3. The reference trajectory in the solid black line shown to the participant to follow during the second part of the experiments. Dashed lines represent the robot-assisted paths traveled by the user root mean squared over all the repetitions with purple representing the first and blue representing the second scenario.

TABLE II
COMPARISON OF TWO SCENARIOS DURING INFERENCE.

	Normal	Scanrio 1	Scenario 2
Number of repetitions	20	20	20
Prediction loss	0.97 ± 0.12	1.33 ± 0.24	1.54 ± 0.19
MSE (Reference trajectory)	0.79 ± 0.1	1.24 ± 0.1	1.93 ± 0.1
RMS EMG	0.9 ± 0.12	0.9 ± 0.34	1.4 ± 0.45
RMS torque (PSI)	64 ± 11	70 ± 19	60 ± 8

the gravitational forces counteracted by gravity compensation, the predicted joint positions deliver precise assistance, matching the exact needs of the user. Notably, at the peak of the movement, the model predicts a higher requirement for torque when the end effector approaches that point, indicating an understanding of the assistance required. Furthermore, the robot-assisted average final path across all repetitions remains within the predefined boundary.

The trajectory's mean square error (MSE) in the first scenario, with one sensor disabled, is 1.24 ± 0.1 cm, increasing to 1.93 ± 0.1 cm in the second scenario with two sensors disabled. This variance highlights a superior performance of the robotic exoskeleton when fewer sensors are missing, as detailed in II, which compares these scenarios against normal operations with all sensors active.

Moreover, the elevated RMS value of the sEMG signals in the second scenario suggests an increased effort exerted by the user, implying a diminished level of assistance from the exoskeleton. This RMS value, calculated for both EMG signals and generated torque, was averaged over the entirety of the movement, encompassing all available sensors and joints during the task. The comparison shows that despite having one or three sensors disabled during inference the model has a relatively good performance. As the number of disabled sensors increases, the system's performance deteriorates, leading to a scenario where assistance predominantly depends on gravity compensation, effectively circumventing the model's predictive capabilities.

While this article has assessed the model's performance under sensor fault conditions, there is potential to further investigate the effects of various types and degrees of sensor faults on the model's performance that do not necessarily involve zero inputs. This additional research could enhance the system's robustness and reliability. Exoskeleton technology demands high real-time performance; therefore, further research could focus on enhancing the real-time control and prediction capabilities of models to reduce system latency and increase response speed. In the future, we will delve deeper into these limitations.

VI. CONCLUSIONS

This paper presents the development of an attention-based encoder-decoder architecture for decoding user movement intention while wearing an upper-limb exoskeleton to provide assistance during rehabilitation procedures and industrial applications. The attention-based model takes sequential data of sEMG and IMU as input and then predicts the future joint position of the robotic pneumatic cable-driven upper limb exoskeleton. The authors identified the relevant muscles that would provide relevant information to the neural network. The attention mechanism was able to detect which inputs contained more relevant information than others as the human user moved in different directions. In addition to better performance and higher accuracy compared with other models, it was demonstrated that the addition of attention to the model would make it more modular while handling longer sequences. Furthermore, the model demonstrates remarkable robustness during inference, even with one or two sensors being unavailable, underscoring its potential for user-adaptive exoskeleton control.

ACKNOWLEDGMENT

The authors would like to thank all the participants who collaborated in this study.

REFERENCES

[1] A. Golabchi, N. Riahi, M. Fix, L. Miller, H. Rouhani, and M. Tavakoli, "A framework for evaluation and adoption of industrial exoskeletons", *Applied Ergonomics*, vol. 113, p. 104103, 2023.

[2] N. Singh, M. Saini, N. Kumar, M. V. P. Srivastava, S. S. Kumaran, and A. Mehndiratta, "A Case Report: Effect of Robotic Exoskeleton Based Therapy on Neurological and Functional Recovery of a Patient With Chronic Stroke", *Front Neurol*, vol. 12, p. 680733, Jul. 2021.

[3] J. Tu, Z. Dai, X. Zhao, and Z. Huang, "Lower limb motion recognition based on surface electromyography", *Biomedical Signal Processing and Control*, vol. 81, p. 104443, 2023.

[4] M. Oghogho, M. Sharifi, M. Vukadin, C. Chin, V. K. Mushahwar and M. Tavakoli, "Deep Reinforcement Learning for EMG-based Control of Assistance Level in Upper-limb Exoskeletons," in *2022 International Symposium on Medical Robotics (ISMR)*, GA, USA, 2022, pp. 1-7, doi: 10.1109.

[5] D. Farina, R. Merletti, and R. M. Enoka, "The extraction of neural strategies from the surface EMG", *Journal of Applied Physiology*, vol. 96, no. 4, pp. 1486–1495, 2004.

[6] C. M. P. Cases *et al.*, "An EMG-based Gesture Recognition for Active-assistive Rehabilitation," in *2020 IEEE 12th International Conference on Humanoid (HNICEM)*, Manila, Philippines, 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400132.

[7] A. Cisnal, J. Pérez-Turiel, J. -C. Fraile, D. Sierra and E. de la Fuente, "RobHand: A Hand Exoskeleton With Real-Time EMG-Driven Embedded Control. Quantifying Hand Gesture Recognition Delays for Bilateral Rehabilitation," *IEEE Access*, vol. 9, pp. 137809-137823, 2021.

[8] Fu J, Choudhury R, Hosseini SM, Simpson R, Park JH. "Myoelectric Control Systems for Upper Limb Wearable Robotic Exoskeletons and Exosuits-A Systematic Review". *Sensors (Basel)*. 2022 Oct 24;22(21):8134. doi: 10.3390/s22218134.

[9] George, J. A., Gunnell, A. J., Archangeli, D., Hunt, G., Ishmael, M., Foreman, K. B., *et al.*, "Robust Torque Predictions from Electromyography across Multiple Levels of Active Exoskeleton Assistance Despite Non-linear Reorganization of Locomotor Output". *Front. Neurorobot.*, 2021, 15, doi:10.3389/fnbot.2021.700823.

[10] C.-S. Son and W.-S. Kang, "Multivariate CNN model for human locomotion activity recognition with a wearable exoskeleton robot," *Bioengineering*, vol. 10, no. 9, Art. no. 1082, Sep. 2023. Available: https://doi.org/10.3390/bioengineering10091082

[11] Zhu M., Guan X., Li Z. *et al.* "sEMG-Based Lower Limb Motion Prediction Using CNN-LSTM with Improved PCA Optimization Algorithm". *J Bionic Eng* 20, 612–627 (2023).

[12] P. Sedighi, X. Li and M. Tavakoli, "EMG-Based Intention Detection Using Deep Learning for Shared Control in Upper-Limb Assistive Exoskeletons," in *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 41-48, Jan. 2024, doi: 10.1109/LRA.2023.3330678.

[13] Xu, K. *et al.* "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, 2048-2057.

[14] A. Zhang, Q. Li, Z. Li and J. Li, "Upper Limb Movement Decoding Scheme Based on Surface Electromyography Using Attention-Based Kalman Filter Scheme," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1878-1887, 2023, doi: 10.1109/TNSRE.2023.3262269.

[15] C. Zhu, Q. Liu, W. Meng, Q. Ai and S. Q. Xie, "An Attention-Based CNN-LSTM Model with Limb Synergy for Joint Angles Prediction," in *2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, Delft, Netherlands, 2021, pp. 747-752, doi: 10.1109/AIM46487.2021.9517544.

[16] Smriti. Bala and Deepak. Joshi, "An Attention-based Deep CNN-BiLSTM Model for Forecasting of Fatigue-induced Surface Electromyography Signals During Isotonic Contractions," in *Institute of Electrical and Electronics Engineers (IEEE)*, 2022.

[17] Coratella Gianpaolo Tornatore *et al.*, An Electromyographic Analysis of Lateral Raise Variations and Frontal Raise in Giuseppe, *Int. J. Environ. Res. Public Health*, 2020, 17, 6015.

[18] Yuri A. C. Campos *et al.*, Different Shoulder Exercises Affect the Activation of Deltoid Portions in Resistance-Trained Individuals, Journal of Human Kinetics volume 75/2020, 5-14 DOI: 10.2478/hukin-2020-0033.

[19] C. Ma *et al.*, "A Bi-Directional LSTM Network for Estimating Continuous Upper Limb Movement From Surface Electromyography," in *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7217-7224, Oct. 2021, doi: 10.1109/LRA.2021.3097272

[20] Naveen Kumar Karnam, Shiv Ram Dubey, Anish Chand Turlapaty, Balakrishna Gokaraju, "EMGHandNet: A hybrid CNN and Bi-LSTM architecture for hand activity classification using surface EMG signals," in *Biocybernetics and Biomedical Engineering*, Volume 42, Issue 1, 2022, Pages 325-340.