RESEARCH ARTICLE

LLMT: A Transformer-Based Multi-Modal Lower Limb Human Motion Prediction Model for Assistive Robotics Applications

S. HOSSEIN SADAT HOSSEINI[1], **NADER N. JOOJILI**[2],
AND MOJTABA AHMADI[1], (Senior Member, IEEE)

[1]Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada
[2]Department of Information Technology, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: S. Hossein Sadat Hosseini (sayyedhosseinsadatho@cmail.carleton.ca)

ABSTRACT Recognition of human intended motion is key to developing intelligent human-robot interaction (HRI) controllers in assistive devices. This study aims to develop a human motion recognition architecture tailored explicitly for real-time assistive robotics, such as exoskeletons and robot-assisted walking systems. We introduced a multi-modal lower limb modified transformer (LLMT), an architecture that bridges the gap in existing HRI technologies by defining a comprehensive set of relevant motions that generalize well for unseen subjects, ensuring adaptability and precision in diverse interaction scenarios. LLMT uses sparse multi-channel surface electromyography (sEMG) and Inertial Measurement Unit (IMU) signals to classify different motion patterns. The accuracy of the proposed method was compared with that of the classical machine learning (cML) models and a convolutional neural network (CNN). This comparison uses experimental data from seven human participants in two motion scenarios and a benchmark dataset. The validation methods included inter-subject, leave-one-subject-out, and intra-subject approaches. The proposed method demonstrated excellent accuracy, achieving $99.42 \pm 0.25\%$, $99.07 \pm 0.32\%$, and $97.08 \pm 1.16\%$ in inter-subject, leave-one-subject-out, and intra-subject validation methods on the collected and benchmark datasets, respectively. Additionally, it exhibited an average online prediction time of 84.09 ms within the recording loop.

INDEX TERMS Surface electromyography, inertial measurement unit, modified transformer, human–robot interaction, lower limb motion recognition, assistive robotics.

I. INTRODUCTION

As the population ages, the need for assistive and rehabilitation robots is increasing. These robots are practical tools that can significantly increase motor capabilities, independence, and quality of life for disabled individuals [1], [2]. For optimal benefits of rehabilitation devices, it is essential to customize them to meet the unique needs of people with disabilities [3]. Moreover, accurate predictions of human motion ensure safety and foster a more intuitive

The associate editor coordinating the review of this manuscript and approving it for publication was Siddhartha Bhattacharyya.

and cooperative interaction between humans and robots. Kinematic, kinetic, and digital data are frequently used in human motion recognition systems [4], [5]. However, a drawback of these signals is that they can only be measured once the human starts the motion. This adds an inherent delay to the kinetic, kinematic, and image-based human motion recognition [6]. Therefore, new technologies are required that can predict the motions intended by humans for robots to react in real-time.

Surface electromyography (sEMG) is an alternative non-intrusive signal for detecting a human's intended motion. This signal is produced 30-150 ms before a human initiates a

motion [7]. Early detection may help decrease the inherent delay in motion recognition systems and improve the response time of robots in human-robot interaction (HRI), leading to a more natural cooperation. However, the accuracy of the sEMG-based motion recognition system may not be sufficient due to the limb position effect [8]. Mechanical sensors, such as Inertial Measurement Units (IMUs), can supply extra information regarding the position of the limb, thus offsetting the limitations of sEMG. Therefore, to improve the performance of motion intent recognition systems, neuromuscular-mechanical sensor fusion has been developed [9]. The fusion of two modalities helps to offset the limitations of each modality, ensuring early detection and precise limb position information, which enhances the overall robustness and consistency of the motion recognition system.

classical machine learning (cML) and deep learning (DL) are the two main methods for multi-channel sEMG and mechanical-based recognition of gestures and motion in the upper and lower limbs. Various cML models have been applied to hand gesture recognition, including multi-layer perceptron (MLP) [10], k-nearest neighbor (KNN), support vector machine (SVM) [11], and random forest (RF) [12]. These ML models use time and frequency features such as mean absolute value (MAV), root mean square (RMS), slope sign changes (SSC), waveform length (WL), integrated sEMG (iEMG), auto-regressive coefficients (AR), and other statistical measures [13].

DL has been extensively used for motion and gesture recognition. In [14], five cML and DL algorithms were employed to detect and classify muscular activities through sEMG signals, where Ninapro DB1 [15] was utilized to validate the model. An autonomous learning framework combining depth vision and sEMG signals was introduced in [16], enabling the automatic labelling of sEMG data classes using depth insights. The system then employs a multi-layer neural network (MNN) to detect hand gestures in real-time. A convolutional neural network (CNN) approach for recognizing different wrist and finger movements was presented in [17]. Performance evaluations were conducted using the Ninapro DB2.

Despite abundant research on hand movement and gesture recognition, few studies have been conducted on the lower limbs. cML techniques such as SVM, LDA, and MLP were used in [18] to predict gait phases during walking on different terrains. Their method predicted the gait phases for leg prosthesis control by fusing multi-channel sEMG, IMU, and goniometer signals. The study found that including contralateral signals significantly reduced the errors in gait phase prediction. A neural network algorithm was developed in [19] to detect eight lower limb movements. The algorithm extracts time-frequency features from the sEMG signal and uses a genetic algorithm to optimize and speed up the convergence. They achieved an average recognition accuracy of 94.89% and an average recognition time of 109.67 ms. To mitigate the impact of subjects' slow response and lack of concentration, the original multi-channel sEMG data undergoes a preprocessing step called ''breaking off both ends,'' which involves discarding the initial and final 200 ms segments of each action data to eliminate potential bias. The authors of [20] proposed a vision transformer-based recognition algorithm using mechanomyography (MMG) and kinematic signals for recognizing eight lower limb motions. Testing on six subjects, achieved 94.62% and 80.13% accuracy in the inter-subject and leave-one-subject-out validation methods, respectively.

Prior research on lower limb motion recognition in exoskeletons and robot-assisted walking systems often lacks a broad spectrum of motions, including transitions between sitting and standing, directional changes, and walking. Moreover, for lower-limb robotics applications, it is critical to develop a highly accurate and robust model that performs well in various validation scenarios, such as inter-subject, intra-subject, and leave-one-subject-out validations. Misclassification in this context can lead to falls, underlining the criticality of a model that can precisely predict human motion. High-accuracy motion/intention detection remains an open and vital challenge in lower-limb motion recognition, particularly when working in online applications. In recent years, numerous variants of the transformer model [21] have emerged, known for their effectiveness in various domains. In the biosignal processing field, a transformer-based architecture was presented in [22] for electroen-cephalography (EEG) signal decoding. Another transformer architecture has been presented in [23] for hand gesture recognition. The recent transformer-based architecture for biosignal processing had many model parameters, making it challenging in real-time and online applications in assistive robotics. Therefore, based on [21], this study proposes a modified transformer architecture tailored for lower limb motion recognition for real-time applications. Our approach aims to fill the gaps in the existing research by providing a solution that enhances the accuracy and robustness of motion recognition in lower-limb assistive robotics. The main contributions of this study are as follows.

- To the best of our knowledge, this study is the first to propose a transformer structure for recognizing various lower limb motions customized for exoskeletons and assistive walking systems through sparse multi-channel sEMG and IMU signal fusion.
- This study shows the superior accuracy of the proposed model under several rigorous validation methods in which the model has not seen the test subject in the training phase.
- In this study, we evaluated the online performance of the proposed model for real-time applications.
- This study adopted the average weighting method for the attention score to achieve balanced information aggregation.

Section II of this paper discusses the experimental protocols, tools, and motion scenarios. Section III elucidates the cML framework and introduces the proposed lower limb modified transformer (LLMT) algorithm. Section IV

showcases the classification results obtained using the collected and benchmark public dataset. Finally, Section V presents the paper's discussion, conclusions, and future works.

II. TOOLS AND PROTOCOLS

A. EXPERIMENTAL PROTOCOL

This study was approved by the Research Ethics Board (REB) of Carleton University under ID 114417. All participants provided informed consent to participate in the study. The study was conducted with seven healthy participants, three males and four females. The age range was 33.14 ± 7.75 years. Participants' heights averaged 166.85 ± 8.31 cm, and their weights were approximately 69.28 ± 18.34 kg.

Wearable sensors were attached to the dominant leg of the subjects to monitor lower limb muscle activity and joint and limb movements. Multi-channel sEMG and IMU signals were collected using EMG-IMU Delsys Wireless® Trigno Avanti™ sensors from seven specific muscles in the dominant leg. These muscles included the gastrocnemius lateralis (GL), tibialis anterior (TA), biceps femoris (BF), rectus femoris (RF), gluteus maximus (GMA), gluteus medius (GME), and sartorius (SA). Figure 1 shows the lower limb muscles included in this study and the placement of the IMU sensors used [24], [25]. These particular muscles were selected for data collection as they exhibit low correlation [24], are suitable for detecting gait initiation intentions [25], and are crucial for movements of the hip, knee, and ankle, which are often supported by wearable lower limb devices [18]. While we have not specifically studied the impact of fewer muscles on the performance of the models, the best selection of fewer or more muscles can be optimized in real-time applications based on the accuracy and runtime of the models. The placement of electrodes followed the standards set by the sEMG for the non-invasive assessment of muscles [26]. To determine the ideal position for electrode placement, we physically examined the muscle belly and aligned the electrode in the direction of the primary muscle fibres [27]. Before sensor attachment, any excess hair at the muscle sites was removed, and the skin was gently cleansed using an alcohol wipe. Double-sided adhesives were then used to secure the sensors to the skin.

sEMG signals were sampled at 1778 Hz and subjected to hardware band-pass filtering within 20 to 450 Hz range. 6-degree of freedom (DOF) IMUs, comprising tri-axial accelerometers and gyroscopes, were positioned on the subjects' thigh–below the RF muscle sEMG electrode–and on the shank adjacent to the TA muscle sEMG electrode (Figure 1). IMUs were sampled at 374 Hz, resampled to 1778 Hz, and fused with sEMG signals at the feature level [28]. All of the data were streamed wireless into Python for further analysis. To diminish motion artifacts, sEMG signals underwent high-pass filtration using a 6th−order Butterworth filter at 20 Hz. Subsequently, the sEMG signals underwent notch filtering with a 6th−order Butterworth filter

having a bandwidth of 6 Hz at frequencies 60 Hz, 180 Hz, and 300 Hz. These frequencies were chosen based on spectral analysis to effectively eliminate surrounding electromagnetic interference. The filtering steps help ensure that the output of the models is noise-free and remains consistent.
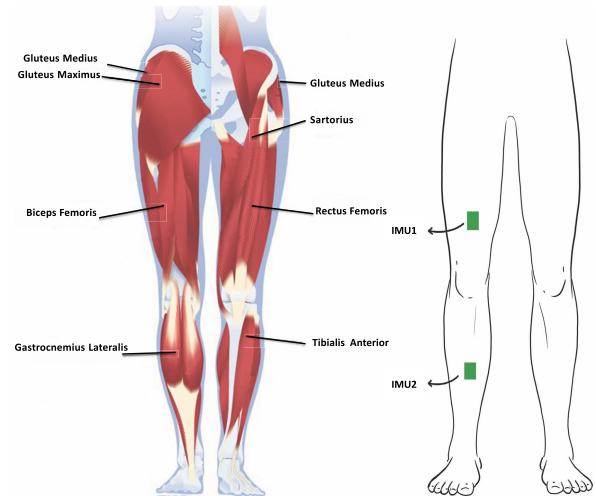

FIGURE 1. Lower limb muscles of the human [29] and the location of the IMU sensors.

B. MOTION SCENARIOS

Two distinct motion scenarios were defined to comprehensively represent the range of daily activities that an individual typically engages in. These scenarios were designed for application to both exoskeletons and robot-assisted walking systems.

- **Scenario 1 (distinct motions):** This scenario comprises ten distinct motions, which include sitting, standing, transitioning between standing and sitting and vice versa, moving forward and backward, turning left and right, and sidestepping (Figure 2a). At the start of each recording trial, the subjects were instructed to perform one of the tasks, and then the recording started for the entire duration of motion. Each movement was repeated 30 times. At the end of the movement, a sound prompt was provided by the recording software was given to them, and the subsequent movement started. The recording software randomly chose the movements. A time progression bar prompted the participant to perform activity onset and offset.
- **Scenario 2 (walking motions):** This scenario was segmented into six distinct phases: (1) standing, (2) level walking (LW) on a flat surface, (3) stairs ascending (SA), (4) stairs descending (SD), (5) ramp ascending (RA), and (6) ramp descending (RD). Participants were instructed to navigate various terrains, as depicted in Figure 2b. Each terrain was traversed for 2 seconds, from direction 1 to direction 2 and then in the reverse order. Upon receiving a cue from the software, participants commenced walking. Notably, the software recorded only 2 seconds of each motion, and each activity was repeated 30 times.

In both scenarios, to diminish the fatigue effects, the participants were given a 5-minute rest period every five repetitions. In these motion scenarios, the gait did not get disturbed and the participants walked on the ground at a self-selected, comfortable speed.

After data collection, to reduce the fluctuation and nonstationarity of the sEMG and IMU signals, the Z-score method was employed as follows:

$$X = \frac{x - \mu}{\sqrt{\sigma^2}} \tag{1}$$

where $X \in \mathbb{R}^{W \times S}$, $x \in \mathbb{R}^{W \times S}$, μ, σ^2, S, and W represent the standardized input, input signal, mean value, variance of the data, number of data channels, and sample points in each trial, respectively.
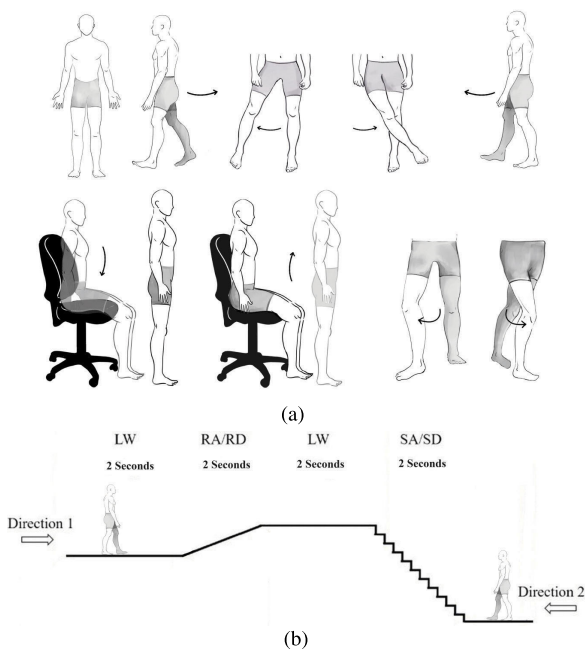

FIGURE 2. Motion scenarios. (a) Scenario 1. (b) Scenario 2.

III. METHODOLOGY

In section III-A, we discuss various cML algorithms and our implementation methodology for the models. Section III-C describes the detailed structure and design of the LLMT architecture.

A. CLASSICAL MACHINE LEARNING (CML)

Time-domain (TD) features tend to outperform frequency-domain (FD) features in classification tasks, demanding less computational power. This efficiency reduces latency in real-time applications [13]. Therefore, TD features were extracted in this study. Nine TD features are extracted per channel, including RMS, iEMG, MAV, WL, SSC, variance, logarithm, skewness, and kurtosis from the sEMG signals [13]. For the IMU signals, nine features per channel were extracted, including the mean, standard deviation, minimum, maximum, energy, median absolute deviation, mean absolute deviation,

kurtosis, and skewness [30]. The time window length was W=250, with an overlap of 64 between two consecutive windows. Given the sEMG and IMU signals sampling rates (1778 Hz), 250 samples were equivalent to approximately 140 ms. The sEMG had a feature space dimensionality of 63, whereas the IMU had 108.

B. FEATURE SELECTION

In this study, we employed the sklearn library [31] to compute the mutual information (MI) and quantify the relationship between two random variables. MI is characterized by a value of zero when the variables are independent, with higher values indicating a greater degree of dependency between them [32]. As the entire feature space has a dimensionality of 171, we selected the top 100 features based on their MI values and fed them into the RF, MLP, KNN, and SVM classifiers to predict human motion.

C. LOWER LIMB MODIFIED TRANSFORMER (LLMT) MODEL

Sequential learning traditionally uses recurrent neural networks (RNNs), which have challenges such as gradient instability and difficulty in capturing long-range dependencies [33]. Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks have been introduced to solve these problems. However, they have limitations in terms of training duration and scalability for long sequences [34].

With their self-attention (SA) mechanism, transformers have revolutionized the natural language processing (NLP) field [35]. They can effectively capture long-range dependencies in sequences of various lengths. The model employs an encoder-decoder architecture that processes sequential data in parallel, in contrast to the local receptive fields of CNNs and sequential order of RNNs. The Vanilla Transformer extends standard transformer applications to time series classification [36].

As illustrated in Fig. 3, the proposed architecture encompasses a preprocessing block in Section III-C1, embedding layer in Section III-C2, transformer encoder in Section III-C3, and classifier in Section III-C4.

1) PREPROCESSING OF DATA

Transforming a raw time series dataset into a format suitable for supervised learning [37] involves segmenting the dataset into parts, denoted as X_i, where i represents the segment index. These segments consist of S channels and W samples, capturing distinct data subsets. The transformation of each segment begins by creating a lagged input sequence, $X_{i,\mathrm{L}}$. This sequence is formed by shifting the data segment backward using a lag parameter $n_{\mathrm{in}} = 10$. A prediction sequence, $X_{i,\mathrm{P}}$, is then generated by shifting the data forward, as determined by the prediction parameter $n_{\mathrm{out}} = 1$. Finally, the matrices of the lagged and predicted sequences are concatenated horizontally to form $X_{i,\mathrm{T}} \in \mathbb{R}^{W \times (M.S)}$, where
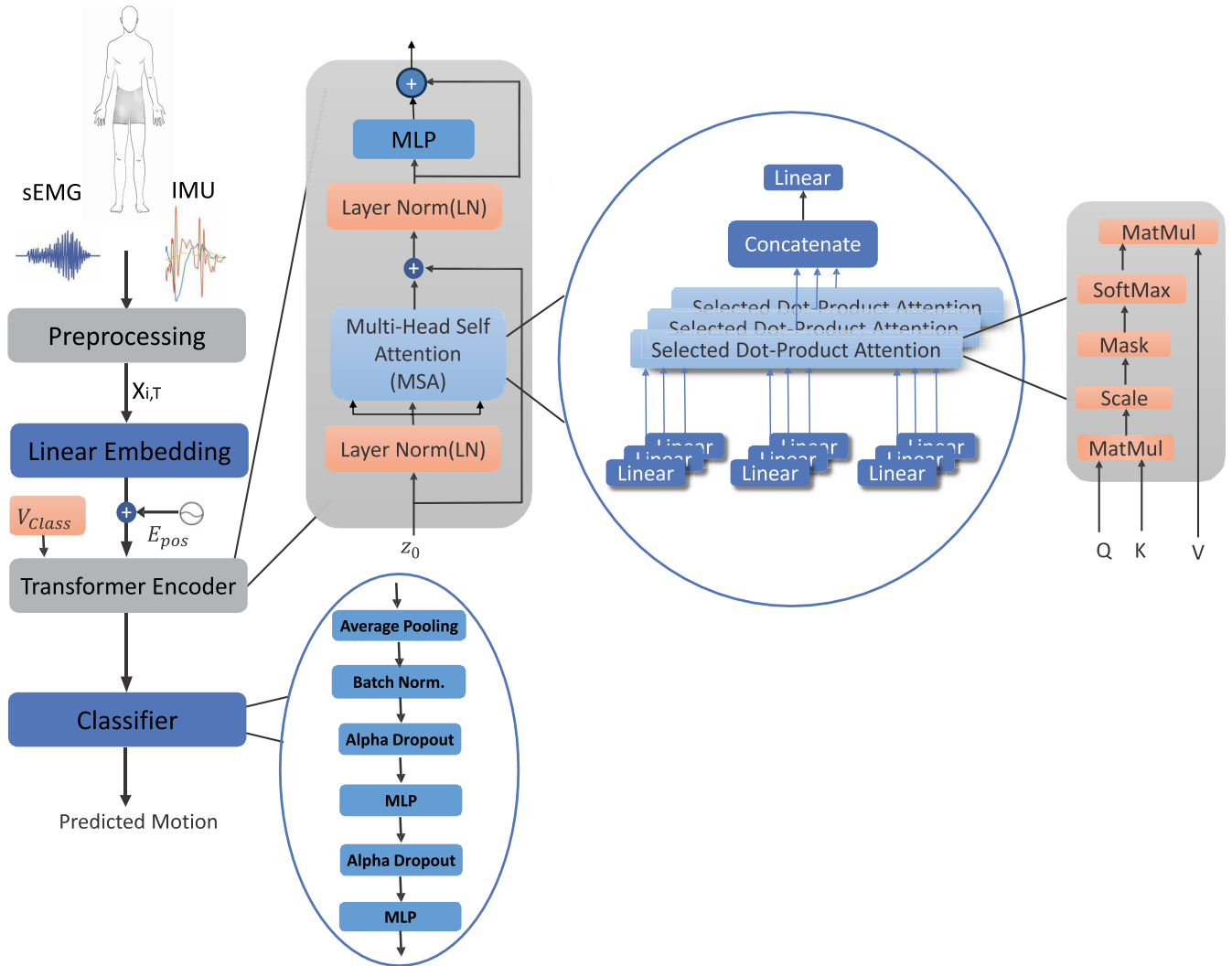
FIGURE 3. The architecture of the proposed LLMT.

$M = (n_{in} + n_{out})$. Applied across all segments, this method yields a transformed dataset D.

2) LINEAR EMBEDDING LAYER

Before the processed sequence is input to the transformer encoder, we split $X_{i,T}$ into M non-overlapping patches with a size of $(S \times S)$. Each patch was then flattened into a vector $x_j^p \in \mathbb{R}^{S^2}$, $1 \leq j \leq M$. A linear projection is then applied within the dimensional vector space of the model, denoted as d, utilizing the embedding matrix $E \in \mathbb{R}^{S^2 \times d}$. This step transformed the input into an embedded representation. Subsequently, these representations were concatenated with a learnable classification token, v_{class} [35]. If the spatial arrangement of these patches is not preserved, the transformer treats the inputs as an unordered patch collection. However, because the sEMG and IMU signals are time series, it is crucial to maintain the spatial arrangement, mirroring the order of the original sequence. To achieve this, the positional

information $E_{pos} \in \mathbb{R}^{(M+1) \times d}$ is encoded and integrated with the patches, ensuring the sequential integrity of the data. It has been claimed that $1 - D$ and $2 - D$ positional encodings produce almost identical results [38], [39]. Therefore, simple $1 - D$ positional encoding is used in the proposed model to maintain positional information. Equation 2 describes the final embedded sequence.

$$z_0 = [v_{class}; x_1^p E; x_2^p E; \ldots; x_M^p E] + E_{pos} \quad (2)$$

3) TRANSFORMER ENCODER

The embedded sequence, z_0, is sent to the transformer encoder (Figure 3). The encoder consists of L identical layers (in the proposed model, it consists of one layer). Each layer has two main components: a multihead self-attention block (MSA), and a fully connected MLP block consisting of two linear layers and a Gaussian Error Linear Unit (GELU) activation function. MSA and MLP are described as follows:

$$z_\iota' = MSA(LN(z_{\iota-1})) + z_{\iota-1}, \iota = 1, .., L \quad (3)$$

$$z_t = MLP(LN(z_t')) + z_t', \iota = 1, .., L \qquad (4)$$

where LN represents the normalization layer. The final output of the transformer is then described as

$$Z_L = [z_{L0}; z_{L1}; \ldots ; z_{LN}] \qquad (5)$$

we take the first element of matrix Z_L in Equation 5, z_0^L, and send it to a layer normalization function as

$$y = LN(z_0^L) \qquad (6)$$

The output from Equation 6 is sent to the classifier block, as described in Section III-C4.

The MSA block (Figure 3), as the central component of the transformer, has two linear layers: SA and a concatenation layer. Next, we focus on SA and MSA.

1) Self-attention (SA): SA was first introduced in [21], intended to capture the interaction between different vectors in the input sequence of $Z \in \mathbb{R}^{N \times d}$, where d is the embedding dimension and N is the number of vectors in the input sequence. The attention mechanism can be mathematically represented as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (7)$$

where Q, K, V, and d_k are the query, key, and value matrices, and the size of each vector in these matrices, respectively. These matrices can be represented by the following linear transformation of input sequence of Z:

$$[Q, K, V] = ZW^{QKV} \qquad (8)$$

where $W^{QKV} \in \mathbb{R}^{d \times 3d_k}$ denotes the learnable weight matrix. The attention score between query element q_i and key element k_j is calculated as:

$$score(q_i, k_j) = q_i \cdot k_j^T \qquad (9)$$

To improve the weighting, attention scores are often scaled by the square root of the dimension of keys d_k:

$$scaled\_score(q_i, k_j) = \frac{score(q_i, k_j)}{\sqrt{d_k}} \qquad (10)$$

The scaled scores were then passed through a softmax function to obtain the attention weights w_{ij}:

$$w_{ij} = \frac{e^{scaled\_score(q_i, k_j)}}{\sum_{j=1}^{n} e^{scaled\_score(q_i, k_j)}} \qquad (11)$$

These attention weights represent the importance of each key element k_j with respect to the query element q_i. They are then used to compute the weighted sum of the value elements v_j to obtain the final output.

$$attention(q_i, K, V) = \sum_{j=1}^{n} w_{ij} \cdot v_j \qquad (12)$$

Calculating the attention scores using the average weighting method for every input sequence ensures

a balanced information aggregation without reducing overall results.

2) Multihead Self-Attention (MSA): We applied the SA mechanism (Eq. 7- Eq. 12) h times (number of heads in Table 1) to the input sequence Z and concatenate the outputs in matrix $[Head_1; Head_2; \ldots.; Head_h] \in \mathbb{R}^{N \times h.d_k}$. We then apply a linear transformation to obtain the following results:

$$MSA(Q, K, V) = [Head_1; Head_2; \ldots ; Head_h]W^O \qquad (13)$$

where $W^O \in \mathbb{R}^{h.d_k \times d}$ and $d_k = d/h$. This completes the description of LLMT architecture.

4) CLASSIFIER

After the above steps, we applied average pooling to Equation 6, the output of the transformer encoder, batch normalization followed by an alpha dropout, an MLP, and an alpha dropout layer, and then an MLP with an output number of neurons equal to the number of motion classes. Then, the softmax function was used to predict the probability of each motion (classifier block in Figure 3). The cross-entropy objective function is employed to calculate the classification loss as follows:

$$L = -\sum_{n=1}^{N} y_n \log(\hat{y}_n) \qquad (14)$$

where N, y_n, and $\hat{y}_n$ are the number of motions, real label of the motion, and predicted probability of the motion, respectively.

IV. RESULTS

The proposed method was implemented using Python 3.9 and the TensorFlow library on an Nvidia A100 GPU. It was trained using the Lazy Adam optimizer with β values of 0.9 and 0.999, and weight decay of 0.001. The training process used a batch size of 250 for each model. An early stopping function was employed to avoid overfitting and, monitor the validation loss. To demonstrate the effectiveness of LLMT, we compared its performance with other models, such as MLP, SVM, KNN, RF, and CNN proposed in [40]. The parameters for both the LLMT and classical models were chosen heuristically to optimize their performance and ensure satisfactory results. This includes a KNN model with four nearest neighbors, an SVM using a radial basis function (RBF) kernel, an RF model with 120 trees, an MLP configured with hidden layers of sizes 200 and 80, and a maximum of 1000 iterations.

A. VALIDATION METHODS

To validate the results comprehensively, distinct approaches were considered as:

- Inter-subject validation: In this validation approach, the dataset pf each subject was divided into 80% for training and 20% for testing.

- Intra-subject validation: In this method, we trained the model with one specific subject and tested it with all other subjects. This process was repeated for all subjects.
- Leave-one-subject-out: In this strategy, one subject is excluded from the training dataset for each iteration. The model was then tested using excluded subjects. This process was repeated for all subjects, ensuring that each subject was used once as the test set.

The last two approaches allow us to thoroughly assess the model's ability to generalize and make accurate predictions for subjects that they have not encountered during training [41], [42].

B. CLASSIFICATION RESULTS

Section IV-B1 presents the classification results of cML, CNN, and LLMT for the collected dataset, followed by a comparative analysis of these models on a public benchmark dataset in Section IV-B2.

1) COLLECTED DATASET

Initially, different variants of LLMT were assessed using our dataset (section II). To evaluate the accuracy of models, we focused on their mean and standard deviation (STD) for the classification of the two motion scenarios involving all subjects, as detailed in Table 1. Notably, Table 1 reveals that the model's accuracy improved and the STD reduced as we increased both the embedding size and number of model parameters. As the number of parameters in a model increases, there is a corresponding increase in training and testing durations and computational costs. This trend presents a challenge for real-time control of assistive robotics, which is the main objective of our project. Consequently, we consciously chose not to increase the model's parameters to mitigate these challenges and to align with our goals.

TABLE 1. Description, classification accuracy, and STD of LLMT variants for two motion scenarios.

Model	Heads	Embedding size	Parameters	Accuracy (%)	STD (%)
Scenario 1					
LLMT 1	8	8	6,986	83.16	9.59
LLMT 2	8	16	12,426	98.99	0.29
LLMT 3	8	24	17,607	99.38	0.12
Scenario 2					
Model	Heads	Embedding size	Parameters	Accuracy (%)	STD (%)
LLMT 1	4	4	3,798	68.44	8.97
LLMT 2	8	8	6,470	98.80	0.52
LLMT 3	8	16	11,910	99.57	0.29

Using the inter-subject validation method, we comprehensively evaluated various cML models, a CNN model alongside the LLMT variants, namely LLMT 1-3. Table 2 shows the results of these assessments for both motion scenarios. LLMT 3 demonstrated an outstanding performance and, emerged as the most effective model. It achieved accuracies of 99.38% in scenario 1 and 99.57% in scenario 2, as listed in Table 2. LLMT 2 was the second-best performer, exhibiting robust accuracies of 98.99% and 98.80% for scenarios 1 and 2, respectively. LLMT 1, however, achieved a lower accuracy of 83.16% in scenario 1 and 68.44% in scenario 2. CNN achieved accuracy of 94.26% and 96.37% in scenarios 1 and 2, respectively. The RF classifier was notable in the realm of the cML models, with significant accuracies of 92.99% and 93.44% in the two scenarios. The MLP model exhibited accuracies of 90.91% in scenario 1 and 94.24% in scenario 2. The SVM classifier attained 80.34% and 91.99% accuracy in the respective scenarios. Finally, the KNN algorithm demonstrated accuracies of 80.07% in scenario 1 and 93.73% in scenario 2.

In the rigorous assessment of the proposed architecture, we focused on LLMT 2 and LLMT 3, employing a leave-one-subject-out validation method. LLMT 1 was excluded from further analysis because of its insufficient classification accuracy in inter-subject validation. This validation approach involved training the models on six subjects and testing the remaining one, which was iteratively applied to each subject in turn. The detailed results of this process are compiled in Table 3. As detailed in the Table, LLMT 2 demonstrated an accuracy of $98.47 \pm 0.72\%$ in scenario 1 and $95.19 \pm 5.10\%$ in scenario 2, across all subjects. LLMT 3, on the other hand, achieved accuracies of $99.25 \pm 0.16\%$ in scenario 1 and $99.30 \pm 0.60\%$ in scenario 2. $S1$ in the table denotes that the model was trained on all subjects except $S1$, and then tested exclusively on $S1$. This methodology was replicated for other subjects, as listed in table.

The superior variant of the proposed model, LLMT 3, underwent intra-subject validation to substantiate its efficacy, in which each subject's data were used exclusively for training and then tested against all other subjects. The detailed results of this validation approach are illustrated in Figures 4. LLMT3 demonstrated an impressive overall accuracy of $97.41 \pm 0.91\%$ in scenario 1 and $98.64 \pm 0.79\%$ in scenario 2. These results underscore the robustness and ability of the model to generalize well, even when trained on data from a single subject and tested on entirely distinct data from other subjects. In Figure 4, $'S1'$ denotes where the model was trained on Subject 1 and tested on the remaining subjects, with analogous interpretations for other legends in the Figure.

2) PUBLIC BENCHMARK DATASET

The two best variants of the proposed model in terms of accuracy are LLMT 2 and LLMT 3 (Table 1). To further assess these models, we utilized the "Encyclopedia of Able-bodied Bilateral Lower Limb Locomotor Signals (ENABL3S)" benchmark dataset [43]. This dataset includes bilateral sEMG data from seven muscles, IMU, and goniometer data collected from 10 non-disabled participants performing various activities. These activities include sitting, standing, walking, ramp ascending and descending, step ascending, and descending. To make it comparable with our dataset, we used

TABLE 2. Average accuracy and F1-score of various models using the inter-subject validation method in different motion scenarios.

				Scenario 1			
Model	KNN	SVM	MLP	RF	CNN [40]	LLMT2	LLMT3
Accuracy	80.07	80.34	90.91	92.99	94.26	98.99	99.38
F1-Score	80.17	77.80	90.87	92.97	94.22	98.95	99.35
				Scenario 2			
Model	KNN	SVM	MLP	RF	CNN [40]	LLMT2	LLMT3
Accuracy	93.73	91.99	94.24	93.44	96.37	98.80	99.57
F1-Score	93.72	91.98	94.23	93.43	96.34	98.79	99.59

TABLE 3. The accuracy and F1-score of LLMT 2 and LLMT 3 models for leave-one-subject-out validation on the collected dataset.

					Scenario 1			
Model	Metric	S1	S2	S3	S4	S5	S6	S7
LLMT2	Accuracy	96.83	98.62	98.54	98.88	98.82	99.24	98.33
	F1-Score	96.85	98.65	98.52	98.89	98.83	99.21	98.32
LLMT3	Accuracy	99.33	99.25	99.20	99.29	99.48	99.28	98.91
	F1-Score	99.31	99.27	99.18	99.27	99.46	99.26	98.90
					Scenario 2			
Model	Metric	S1	S2	S3	S4	S5	S6	S7
LLMT2	Accuracy	99.67	99.52	86.61	93.53	98.24	89.05	99.68
	F1-Score	99.63	99.48	86.60	93.50	98.27	89.02	99.65
LLMT3	Accuracy	99.69	99.74	99.77	99.42	98.53	98.23	99.73
	F1-Score	99.64	99.71	99.75	99.41	98.50	98.22	99.71

TABLE 4. Average classification performance metrics of various models on public benchmark dataset for inter-subject validation method.

Model	KNN	SVM	MLP	RF	CNN	LLMT2	LLMT3
Accuracy	89.97	85.94	90.76	89.93	95.02	97.70	99.28
F1-Score	90.00	85.86	90.71	89.89	95.01	97.68	99.37
Precision	89.98	86.02	90.80	90.06	95.06	97.77	99.36
Recall	89.46	85.93	90.75	89.92	95.01	97.69	99.37

only the sEMG and IMU data from the right leg, omitting the left leg and goniometer data. The resulting dataset comprised seven sEMG and twelve IMU channels from the shank and thigh of the participants.

Table 4 shows the average performance metrics for the cML, CNN, and LLMT models for inter-subject validation on the public dataset. LLMT 3, LLMT2, CNN, MLP, SVM, RF, and KNN achieved accuracies of 99.33%, 97.70%, 95.02%, 90.75%, 85.93%, 89.92%, and 89.97%, respectively, across all subjects.

The leave-one-subject-out validation method results for LLMT 2 and LLMT 3 for the public dataset are presented in Table 5, where LLMT 2 and LLMT 3 attained an accuracy of $89.82 \pm 3.72\%$ and $98.67 \pm 0.93\%$, respectively. In this table, $S156$ indicates that this subject was reserved for testing, which applies similarly to the other subjects in the dataset. Intra-subject validation of LLMT3, depicted in Figure 5, revealed an accuracy of $95.20 \pm 1.77\%$. In Figure 5, $'S156'$ shows that the model was trained with this subject and tested on other subjects. It can be applied similarly to all subjects in Figure.

C. ASSESSMENT OF MODELS FOR REAL-TIME SCENARIOS

This study aimed to develop a model capable of accurately predicting human motion with strong generalization for unseen data that is light and quick for real-time assistive robotics applications. We trained various cML, CNN, and LLMT models and saved them for online analysis. The test times of these models were evaluated by incorporating them into a recording loop, as shown in Figure 6, where the fused sEMG and IMU signals were segmented into 250-sample windows (equivalent to 140 ms) and input into the models to recognize human motion. These experiments were performed on a Windows 11 laptop with an Intel Core^{TM} $i7$ CPU and 16 GB RAM. A custom-written Python program managed data streaming from the sensors to the laptop, and we measured the runtime of the models for each signal segment. As Table 6 indicates, the RF model achieved a 92.99% accuracy with a test time of 11.80 ms, while the MLP model reached 90.13% accuracy in 10.54 ms. The KNN, SVM, and CNN models recorded accuracies of 80.07%, 80.34%, and 94.6% with test times of 7.07 ms, 8.09 ms, and 45.16 ms, respectively. In contrast, the LLMT 2 and LLMT 3 models delivered the highest accuracies of 98.99% and 99.32%, but with longer test times of 75.50 ms and 84.09 ms, and parameter counts of 12,426 and 17,559, respectively. These findings suggest that an increase in the signal's test time per input batch and the number of parameters for the models correlates with improved accuracy in motion prediction.

TABLE 5. The accuracy and F1-score of the LLMT2 and LLMT3 for leave-one-subject-out validation on the public dataset.

Model	Metric	S156	S185	S186	S188	S189	S190	S191	S192	S193	S194
LLMT2	Accuracy	85.96	88.23	93.63	93.24	90.57	90.97	80.84	91.67	92.87	90.23
	F1-Score	94.12	95.92	93.62	92.73	90.66	91.46	80.76	91.81	95.77	97.59
LLMT3	Accuracy	98.67	99.51	97.29	96.98	98.54	99.27	99.73	97.87	99.55	99.33
	F1-Score	98.68	99.53	97.31	96.96	98.52	99.29	99.75	97.86	99.54	99.35

Accuracy of LLMT3 in scenario 1 for intra-subject validation method.

(a)

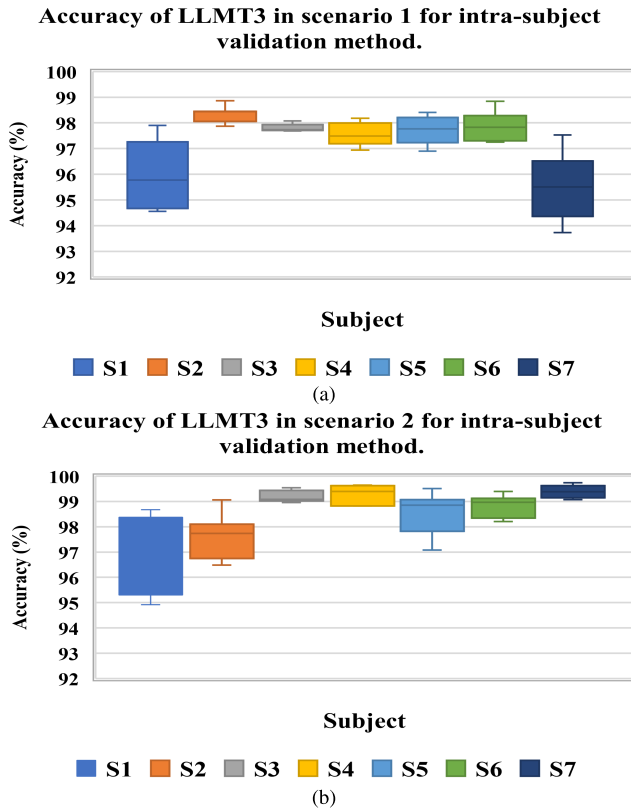**Accuracy of LLMT3 in scenario 2 for intra-subject validation method.**

(b)

FIGURE 4. The accuracy of LLMT 3 in intra-subject validation method. (a) Scenario 1. (b) Scenario2.

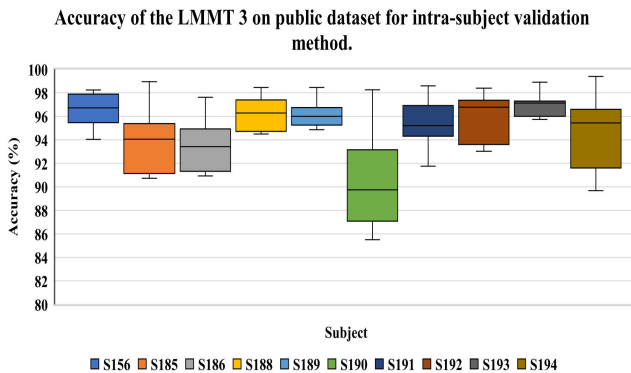**Accuracy of the LMMT 3 on public dataset for intra-subject validation method.**

FIGURE 5. The accuracy of the LLMT3 on the public dataset. Intra-subject validation.

V. DISCUSSION AND FUTURE WORK

As one of the popular DL architectures in human gesture and motion recognition using biosignals, CNN can only extract spatial features from the input signals and cannot capture
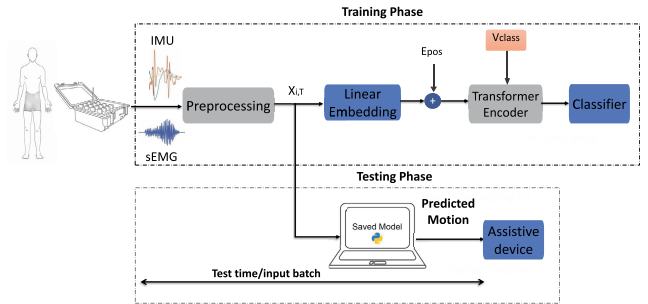

FIGURE 6. Recording loop for the on-line model evaluation.

TABLE 6. Average online prediction time of different models for scenario 1.

Model name	Test time/input batch	Parameters	Accuracy (%)
RF	11.80 ms	—	92.99
MLP	10.54 ms	—	90.13
KNN	7.07 ms	—	80.07
SVM	8.09 ms	—	80.34
CNN [41]	45.16 ms	72,586	94.26
LLMT 2	75.50 ms	12,426	98.99
LLMT 3	84.09 ms	17,559	99.32

the temporal features [40]. LSTM and GRU networks were introduced to solve this drawback; however, they suffer from long training duration and scalability for long sequences [33]. Recently, transformers with temporal and spatial feature learning have drawn the attention of researchers in biosignal processing. However, much of the existing research in this area typically has too many model parameters and restricted sets of motions [20], [23], making real-time and online applications challenging. Unlike previous works, by defining a broader spectrum of lower-limb motions, this work introduces LLMT, which leverages the fusion of sparse multi-channel sEMG and IMU signals and is uniquely designed for online applications, offering high accuracy and swift response times essential for assistive walking devices.

In existing research on controlling assistive devices for the upper and lower limbs using sEMG signals, the common approach involves defining specific gestures. The participants were asked to perform and hold gestures for approximately five seconds. Based on these gestures, the pattern recognition system recognizes human activities and issues commands to the assistive device accordingly. However, this gesture recognition method requires a longer experimental period and hinders the recognition system's ability to swiftly identify the user's state. This delay is undesirable for assistive devices which aim to be intuitive and responsive. Some studies,

such as in [19], the initial and final 200 ms segment of the motion were discarded to reduce the effects of slow responses and lack of concentration in subjects. However, because sEMG is produced 30-150 ms before motion starts [7], it is one of the main reasons for choosing sEMG signals to control an assistive device, as it allows for early detection of intended movements and improves the response time of the device. Excluding the first and last 200 ms of the signal in real-time control is problematic because the model has not been trained on this signal segment. Consequently, this increases the risk of misclassification at the beginning and end of the motion cycle. To address these gaps, our research defines motions instead of gestures, each lasting between 900-1200 ms for scenario 1 and 2000 ms for scenario 2. Unlike previous studies, we did not exclude the initial part of the sEMG signals from our model's training. By including this segment, which carries the intended part of the motion, we aimed to develop a pattern recognition system that is more responsive and intuitive. The findings suggest that the LLMT model can accurately predict motion, even with these adjustments.

In Table 1, we show that an increase in the embedding size and number of heads in the model correlates with an increase in the number of parameters. This parameter escalation enhances the accuracy of motion recognition and reduces the STD, but also leads to longer training and testing durations. This creates conflict with the aim of our project for real-time applications. Consequently, we decided not to further increase the number of parameters after observing that LLMT 2 achieved accuracies of 98.99% and 98.80%, and LLMT 3 achieved 99.38% and 99.57% in scenarios 1 and 2, respectively. While models such as RF, MLP, SVM, KNN, and CNN achieved notable accuracy in the two motion scenarios, the LLMT 2 and LLMT 3 models demonstrated superior performance in terms of inter-subject validation accuracy. However, as shown in Table 6, it is essential to note that LLMT 3's average online prediction time in the recording loop is 84.09 ms. This duration is seven times longer than the best cML model in terms of accuracy, which is RF. This observation highlights a trade-off: achieving higher accuracies with LLMT models requires increased computational resources. Previously, we discussed how most sEMG-based frameworks for human motion recognition are not designed for real-time applications and often lack an online evaluation of their models. Despite its longer prediction time than the cML models and CNN, LLMT 3 outperforms some models such as the one reported in [19]. This comparison shows LLMT 3's balance between the accuracy and online prediction time.

To test the generalizability of LLMT models, we subject them to rigorous validation methods, such as leave-one-subject-out and intra-subject. In these scenarios, the model did not consider the data from the subjects during training. Because of the nonstationary nature and subject-to-subject variation of sEMG, generalization of the subject that the model has not seen during the training phase is very challenging. As shown in table 3, LLMT 2 demonstrated an accuracy of 98.47 ± 0.72% in scenario 1 and 95.19 ± 5.10% in scenario 2, across all subjects for the leave-one-subject-out method. LLMT 3, on the other hand, achieved accuracies of 99.25 ± 0.16% in scenario 1 and 99.30 ± 0.60% in scenario 2. The other rigorous validation method was intra-subject validation, in which LLMT is trained on one subject and tested against all other subjects. As shown in Figure 4, LLMT3 demonstrated an impressive overall accuracy of 97.41 ± 0.91% in scenario 1 and 98.64 ± 0.79% in scenario 2. LLMT3, a superior variant of the proposed model, was tested using these validation methods on a public benchmark dataset. As Figure 5 and Table 5 show, it attained the accuracy of 98.67 ± 0.93% and 95.20 ± 1.77% during leave-one-subject-out and intra-subject validation, respectively. These results indicate that the proposed method generalizes well and is robust even when the subject is not in the training phase.

Future research explores the effectiveness of the model in populations with high variability in sEMG signals, such as individuals with cerebral palsy (CP). This exploration, alongside long-term user studies in daily activities and testing of the model with advanced assistive walking robots, will offer valuable insights into practical challenges and real-life user experiences.

Our research presented the LLMT model for human motion recognition from the fusion of multi-channel sEMG and IMU signals. Unlike previous studies, we focused on a more comprehensive range of lower-limb motions and real-time applications of the proposed model. Our findings demonstrate that LLMT models, particularly LLMT 3, achieve high accuracy in motion prediction while balancing computational demands. Despite longer prediction times than cML models, LLMT models excelled in different validation methods even when the model had not seen the data of the subjects during the training phase, highlighting their robustness and potential for developing more intuitive and, adaptable to users and efficient assistive devices.

REFERENCES

[1] T. Yan, M. Cempini, C. M. Oddo, and N. Vitiello, "Review of assistive strategies in powered lower-limb orthoses and exoskeletons," *Robot. Auto. Syst.*, vol. 64, pp. 120–136, Feb. 2015.

[2] P. Beckerle, G. Salvietti, R. Unal, D. Prattichizzo, S. Rossi, C. Castellini, S. Hirche, S. Endo, H. B. Amor, and M. Ciocarlie, "A human–robot interaction perspective on assistive and rehabilitation robotics," *Frontiers Neurorobot.*, vol. 11, p. 24, Jan. 2017.

[3] M. Xiloyannis, D. Chiaradia, A. Frisoli, and L. Masia, "Physiological and kinematic effects of a soft exosuit on arm movements," *J. NeuroEng. Rehabil.*, vol. 16, no. 1, pp. 1–15, Dec. 2019.

[4] R. Ikeura, T. Moriguchi, and K. Mizutani, "Optimal variable impedance control for a robot and its application to lifting an object with a human," in *Proc. 11th IEEE Int. Workshop Robot Hum. Interact. Commun.*, Sep. 2002, pp. 500–505.

[5] H.-Y. Li, I. Paranawithana, L. Yang, T. S. K. Lim, S. Foong, F. C. Ng, and U.-X. Tan, "Stable and compliant motion of physical human–robot interaction coupled with a moving environment using variable admittance and adaptive control," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2493–2500, Jul. 2018.

[6] D. Sirintuna, I. Ozdamar, Y. Aydin, and C. Basdogan, "Detecting human motion intention during pHRI using artificial neural networks trained by EMG signals," in *Proc. 29th IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2020, pp. 1280–1287.

[7] L. Wang, J. Fu, H. Chen, and B. Zheng, "Hand gesture recognition using smooth wavelet packet transformation and hybrid CNN based on surface EMG and accelerometer signal," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105141.

[8] Y. Wang, X. Cheng, L. Jabban, X. Sui, and D. Zhang, "Motion intention prediction and joint trajectories generation toward lower limb prostheses using EMG and IMU signals," *IEEE Sensors J.*, vol. 22, no. 11, pp. 10719–10729, Jun. 2022.

[9] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2867–2875, Oct. 2011.

[10] J. G. Colli-Alfaro, A. Ibrahim, and A. L. Trejos, "Design of user-independent hand gesture recognition using multilayer perceptron networks and sensor fusion techniques," in *Proc. IEEE 16th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2019, pp. 1103–1108.

[11] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, "Classification of finger movements for the dexterous hand prosthesis control with surface electromyography," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 608–618, May 2013.

[12] P. Kaufmann, K. Englehart, and M. Platzner, "Fluctuating emg signals: Investigating long-term effects of pattern matching algorithms," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 6357–6360.

[13] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[14] N. M. Hye, U. Hany, S. Chakravarty, L. Akter, and I. Ahmed, "Artificial intelligence for sEMG-based muscular movement recognition for hand prosthesis," *IEEE Access*, vol. 11, pp. 38850–38863, 2023.

[15] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G.-M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 1, pp. 1–13, Dec. 2014.

[16] S. E. Ovur, X. Zhou, W. Qi, L. Zhang, Y. Hu, H. Su, G. Ferrigno, and E. De Momi, "A novel autonomous learning framework to enhance sEMG-based hand gesture recognition using depth information," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102444. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421000410

[17] S. Briouza, H. Gritli, N. Khraief, S. Belghith, and D. Singh, "A convolutional neural network-based architecture for EMG signal classification," in *Proc. Int. Conf. Data Anal. Bus. Ind. (ICDABI)*, Oct. 2021, pp. 107–112.

[18] B. Hu, E. Rouse, and L. Hargrove, "Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities," *Frontiers Robot. AI*, vol. 5, pp. 1–16, Jun. 2018.

[19] P. Zhang, J. Zhang, and A. Elsabbagh, "Lower limb motion intention recognition based on sEMG fusion features," *IEEE Sensors J.*, vol. 22, no. 7, pp. 7005–7014, Apr. 2022.

[20] H. Zhang, K. Yang, G. Cao, and C. Xia, "ViT-LLMR: Vision transformer-based lower limb motion recognition from fusion signals of MMG and IMU," *Biomed. Signal Process. Control*, vol. 82, Apr. 2023, Art. no. 104508.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Metadata.json

[22] Y. Song, X. Jia, L. Yang, and L. Xie, "Transformer-based spatial–temporal feature learning for EEG decoding," 2021, *arXiv:2106.11170*.

[23] S. Zabihi, E. Rahimian, A. Asif, and A. Mohammadi, "TraHGR: Transformer for hand gesture recognition via electromyography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 4211–4224, 2023.

[24] J. Wang, Y. Dai, and X. Si, "Analysis and recognition of human lower limb motions based on electromyography (EMG) signals," *Electronics*, vol. 10, no. 20, p. 2473, Oct. 2021.

[25] E. C. Wentink, S. I. Beijen, H. J. Hermens, J. S. Rietman, and P. H. Veltink, "Intention detection of gait initiation using EMG and kinematic data," *Gait Posture*, vol. 37, no. 2, pp. 223–228, Feb. 2013.

[26] SENIAM. *Surface Electromyography for the Non-Invasive Assessment of Muscles*. Accessed: Oct. 16, 2024. [Online]. Available: https://www.seniam.org

[27] F. P. Kendall, E. K. McCreary, P. G. Provance, M. M. Rodgers, and W. A. Romani, *Muscles: Testing and Function With Posture and Pain*, vol. 5. Baltimore, MD, USA: Lippincott Williams & Wilkins, 2005.

[28] D. Novak and R. Riener, "A survey of sensor fusion methods in wearable robotics," *Robot. Auto. Syst.*, vol. 73, pp. 155–170, Nov. 2015.

[29] T. K. Uchida and S. L. Delp, *Biomechanics of Movement: The Science of Sports, Robotics, and Rehabilitation*. Cambridge, MA, USA: MIT Press, 2021.

[30] H. A. Varol, F. Sup, and M. Goldfarb, "Multiclass real-time intent recognition of a powered lower limb prosthesis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 542–551, Mar. 2010.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[32] C. Magri, K. Whittingstall, V. Singh, N. K. Logothetis, and S. Panzeri, "A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings," *BMC Neurosci.*, vol. 10, no. 1, pp. 1–24, Dec. 2009.

[33] P.-A. Buestán-Andrade, M. Santos, J.-E. Sierra-García, and J.-P. Pazmiño-Piedra, "Comparison of LSTM, GRU and transformer neural network architecture for prediction of wind turbine variables," in *Proc. Int. Conf. Soft Comput. Models Ind. Environ. Appl.* Cham, Switzerland: Springer, 2023, pp. 334–343.

[34] A. Toro-Ossaba, J. Jaramillo-Tigreros, J. C. Tejada, A. Peña, A. López-González, and R. A. Castanho, "LSTM recurrent neural network for hand gesture recognition using EMG signals," *Appl. Sci.*, vol. 12, no. 19, p. 9700, Sep. 2022.

[35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[36] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," 2022, *arXiv:2202.07125*.

[37] S. Gaugel, B. Wu, A. Anand, and M. Reichert, "Supervised time series segmentation as enabler of multi-phased time series classification: A study on hydraulic end-of-line testing," in *Proc. IEEE 21st Int. Conf. Ind. Informat. (INDIN)*, Jul. 2023, pp. 1–8.

[38] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[39] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.*, vol. 13, no. 3, p. 516, Feb. 2021.

[40] B.-Y. Su, J. Wang, S.-Q. Liu, M. Sheng, J. Jiang, and K. Xiang, "A CNN-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1032–1042, May 2019.

[41] E. Yavuz and C. Eyupoglu, "A cepstrum analysis-based classification method for hand movement surface EMG signals," *Med. Biol. Eng. Comput.*, vol. 57, no. 10, pp. 2179–2201, Oct. 2019.

[42] F. Di Nardo, C. Morbidoni, G. Mascia, F. Verdini, and S. Fioretti, "Intra-subject approach for gait-event prediction by neural network interpretation of EMG signals," *Biomed. Eng. OnLine*, vol. 19, no. 1, pp. 1–20, Dec. 2020.

[43] B. Hu, E. Rouse, and L. Hargrove, "Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals," *Frontiers Robot. AI*, vol. 5, p. 14, Feb. 2018.

S. HOSSEIN SADAT HOSSEINI received the M.S. degree in biomedical engineering from Iran University of Science and Technology (IUST), Tehran, Iran, in 2015. He is currently pursuing the Ph.D. degree in biomedical engineering with the Department of Mechanical and Aerospace Engineering, Carleton University. Since 2015, he has been an Engineer in several medical imaging companies. His research interests include biomedical signal processing, control theory, machine learning, and rehabilitation robotics.

NADER N. JOOJILI received the M.S. degree in information technology from Carleton University, Ottawa, ON, Canada, where he is currently pursuing the Ph.D. degree in information technology. Since 2022, he has been a Data Scientist at Ericsson Canada and Sweden, National Research Council Canada, Ottawa. His research interests include machine learning, deep learning, applied big data, signal processing, and complex neural networks using complex numbers.

MOJTABA AHMADI (Senior Member, IEEE) received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 1989, the M.Sc. degree from the University of Tehran, Tehran, Iran, in 1992, and the Ph.D. degree in mechanical engineering from McGill University, Montreal, QC, Canada, in 1998. From 1998 to 2000, he worked on robotic teleoperation and autonomous operation as a Postdoctoral Fellow with École Polytechnique de Montréal. From 2000 to 2005, he held senior positions in different industries, such as Opal-RT Technologies Inc., Montreal; Quantum and Maxtor Corporations, San Jose, CA, USA; and the Institute for Aerospace Research, National Research Council Canada, Ottawa, ON, Canada. In 2005, he joined the Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON, Canada, where he is currently a Professor. He also founded the Advanced Biomechatronics and Locomotion Laboratory (ABL). His current interests include rehabilitation robotics, assistive devices, human–robot interaction, design and control of mechatronic systems, autonomous vehicles, and legged robots.

● ● ●