Subject-Independent Continuous Locomotion Mode Classification for Robotic Hip Exoskeleton Applications

Inseung Kang, *Member, IEEE*, Dean D. Molinaro, *Graduate Student Member, IEEE*, Gayeon Choi, Jonathan Camargo, and Aaron J. Young, *Member, IEEE*

***Abstract*—Autonomous lower-limb exoskeletons must modulate assistance based on locomotion mode (e.g., ramp or stair ascent) to adapt to the corresponding changes in human biological joint dynamics. However, current mode classification strategies for exoskeletons often require user-specific tuning, have a slow update rate, and rely on additional sensors outside of the exoskeleton sensor suite. In this study, we introduce a deep convolutional neural network-based locomotion mode classifier for hip exoskeleton applications using an open-source gait biomechanics dataset with various wearable sensors. Our approach removed the limitations of previous systems as it is 1) subject-independent (i.e., no user-specific data), 2) capable of continuously classifying for smooth and seamless mode transitions, and 3) only utilizes minimal wearable sensors native to a conventional hip exoskeleton. We optimized our model, based on several important factors contributing to overall performance, such as transition label timing, model architecture, and sensor placement, which provides a holistic understanding of mode classifier design. Our optimized DL model showed a 3.13% classification error (steady-state: 0.80 ± 0.38% and transitional: 6.49 ± 1.42%), outperforming other machine learning-based benchmarks commonly practiced in the field ($p<0.05$). Furthermore, our multi-modal analysis indicated that our model can maintain high performance in different settings such as unseen slopes on stairs or ramps. Thus, our study presents a novel locomotion mode framework, capable of advancing robotic exoskeleton applications toward assisting community ambulation.**

Manuscript received 8 July 2021; revised 7 December 2021 and 17 February 2022; accepted 3 April 2022. Date of publication 7 April 2022; date of current version 20 September 2022. This work was supported in part by the NSF NRI under Award 1830215, in part by NSF GRFP under Award DGE-1650044, in part by the NSF NRT: Accessibility, Rehabilitation, and Movement Science (ARMS) under Award 1545287, and in part by the Fulbright Foreign Student Fellowship. *(Corresponding author: Inseung Kang.)*

Inseung Kang is with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and also with the Department of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: inseung@mit.edu).

Gayeon Choi is with the Department of Mechanical Engineering, Georgia Institute of Technology, Massachusetts Institute of Technology, USA.

Dean D. Molinaro, Jonathan Camargo, and Aaron J. Young are with the Department of Mechanical Engineering, Georgia Institute of Technology, USA, and also with the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2022.3165547, provided by the authors.

Digital Object Identifier 10.1109/TBME.2022.3165547

***Index Terms*—Locomotion mode classification, robotic exoskeleton, user intent recognition, deep convolutional neural network.**

I. Introduction

THE field of robotic exoskeletons has grown to a great extent and has shown the potential to improve people's quality of life through enhanced human mobility [1], [2]. Exoskeletons augment the user with external assistance and improve the user's gait functions, such as reducing the energetic demand [3], [4] and increasing preferred walking speed [5]. In particular, hip exoskeletons show great promise due to the hip joint's contributions to a wide range of locomotor tasks [6]. During community ambulation, humans change their hip dynamics with changes in locomotion mode (e.g., level-ground, ramp ascent, etc.) [7]–[9]. To account for these changes in biological joint torque demand, previous research studies have optimized torque assistance profiles for specific locomotion modes to improve human ambulation [10]–[12]; however, this drives the need for a system able to detect the user's locomotion mode to trigger a change in exoskeleton assistance. Currently, commercially available systems use inconvenient or unintuitive methods to detect the user's locomotion mode online (e.g., clicking a manual button on a hand-held controller) [13], [14]. Instead, detecting this environmental information for these exoskeletons should be natural, automatic, and seamless. Thus, to provide effective and continuous assistance during various locomotion mode changes (e.g., ambulating over an outdoor terrain), an accurate understanding of the user's locomotion mode is important.

Several autonomous classification strategies have been introduced to detect the user's locomotion modes using different wearable robotic platforms [15]–[32]. Most of these studies have originated from robotic prosthesis applications [15]–[24], but recent studies have started to expand their approach to exoskeletons [27]–[32]. Often these methods leverage machine learning (ML) techniques, utilizing different types of sensors from both the user and the device. For robotic hip exoskeletons, Long *et al.* proposed an online SVM-based model to identify different locomotion modes in real-time for exoskeletons [27]. However, the study was limited as it required additional distal sensors in the user's shoes, which may not be feasible for a hip exoskeleton. Additionally, the model had up to 30% of gait phase

identification delays during level-ground to descent mode transitions, which can lead to suboptimal exoskeleton assistance. Kim *et al.* presented a kinematic-based mode classifier using two-layer decision trees for exoskeletons [26]. However, this classifier only updated locomotion mode during the mid-stance of the gait cycle, which can also delay the overall prediction during a mode transition period. Wang *et al.* developed an LSTM-based neural network model to recognize five locomotion modes, including sit/stand, level-ground, and stair ambulations [28]. However, the study was limited as the model was not trained to classify ramp setting, which is one of the key locomotion modes. Zhong *et al.* utilized a combination of a shank-mounted inertial measurement unit (IMU) and a camera to detect different environmental contexts [33]. Similarly, Laschowski *et al.* presented a vision-based environmental mode recognition system including level-ground and stair locomotion using a deep neural network achieving a 5% overall classification error [30]. However, similar to other studies, it required external sensors and only classified a limited set of locomotion modes.

While these studies have provided foundational information toward developing a mode classifier for robotic hip exoskeletons, several limitations are halting the field to move forward. First, there is a lack of understanding in developing a continuous mode classifier (e.g., prediction of mode based on a sliding window of sensor input at a fast frame rate). Discrete classification (often done to yield high classification accuracy) inherently delays the mode prediction time (typically by half a gait cycle), inducing a lag in the assistance profile, which can lead to suboptimal human-exoskeleton performance (especially during mode transitions). Second, there is inadequate information about developing a user-independent classification strategy. User-dependent models require cumbersome manual tuning on a user-specific basis which may not be feasible for real-world applications. Finally, there is a lack of exploration in applying the state-of-the-art deep learning (DL) algorithms for locomotion mode classification. The majority of these studies have only exploited a certain algorithm (often tailored to the researcher's preference) that fits specifically to its targeted application. This can potentially induce a bias in evaluating the algorithm's overall effectiveness. Thus, a more in-depth analysis of DL algorithms in multi-perspective evaluation, that resembles real-world settings, is vital to help researchers develop a classifier that is translatable to exoskeletons available on the market.

Our central hypothesis was that a DL-based model would have a lower classification error than the models using conventional ML learning techniques when developing a subject-independent and continuous mode classifier. The underlying rationale for improved performance was that the deep neural network leverages an end-to-end approach. The model architecture includes a feature engineering stage during the training process, enabling the model to further extract relevant representation for mode classification, which is inherently limited by conventional manual feature extraction. Moreover, these models can take advantage of a large dataset during the training process to generalize across different subjects. We designed a deep convolutional neural network (CNN) utilizing a dataset from human gait biomechanics with wearable mechanical sensors (sensors providing kinematic information) while mimicking different locomotion settings [9].

The key contribution of this study is that it provides a holistic design guideline to develop a DL-based classifier that can detect the user's locomotion mode for robotic hip exoskeleton applications. A fundamental limitation in the current literature is that studies are often specifically tuned based on the study design, and key information such as type of algorithm, optimal sensor suite, model architecture, and data processing techniques and their effect on classification error were not explored formally. This greatly limits the capability of understanding the feasibility of deploying a classifier to an exoskeleton system in the real-world. Furthermore, our study is novel as we mainly focused on developing a framework for the hip joint where designing a classifier using only proximal sensors is a challenging task. Using our multi-perspective analysis that is translatable to other hip exoskeleton platforms, future exoskeleton designers can efficiently plan and develop systems that can robustly classify the user's locomotion mode for a fully autonomous system in a more realistic setting, such as outdoor locomotion.

II. Data Preparation

A. Human Subject Locomotion Dataset

Our study utilized a publicly available open-source dataset collected by our group [9]. The study was approved by the Georgia Institute of Technology Institutional Review Board and informed written consent was obtained for all subjects. Twenty able-bodied participants (11 males/9 females) with the age of 21.2 ± 3.1 years, the height of 1.70 ± 0.08 m, and the body mass of 67.64 ± 11.56 kg were asked to walk on an in-lab height-adjustable terrain park (Fig. 1(a)). The terrain park consisted of 5 different locomotion settings, including level-ground (LG), ramp ascent (RA), ramp descent (RD), stair ascent (SA), and stair descent (SD). The terrain park setting was adjusted and set to four different ramp inclines of 7.8°, 9.2°, 11°, and 12.4° and four different stair heights of 10.2 cm, 12.7 cm, 15.2 cm, and 17.8 cm (covering the ADA accessibility standards).

For all locomotion modes, subjects were instructed to walk at their self-selected walking speed. For LG, subjects were asked to walk over a straight path 6 times, where accel/deceleration phases were excluded to only use the steady-state walking conditions. For a given ramp incline, subjects were instructed to complete a set of 3 trials starting with their right leg and a set of 3 trials starting with their left leg for a total of 24 RA trials and 24 RD trials. For all ramp and stair conditions (for both the ascent and descent mode), subjects ambulated with the following locomotion mode order: LG, mode transition, ramp/stair mode, mode transition, and LG. Similar to the ramp trials, subjects ambulated over a 6-step staircase completing a set of 3 trials for each starting leg at a given stair height for a total of 24 SA trials and 24 SD trials. A total of 102 trials were conducted per subject across 5 locomotion modes.

B. Simulated Sensor Data Generation

Simulated IMUs from the motion capture data were created to simulate different locations on the thigh for nine potential placements of wearable sensors (Fig. 1(c)). To generate simulated sensor data, we utilized the motion capture marker

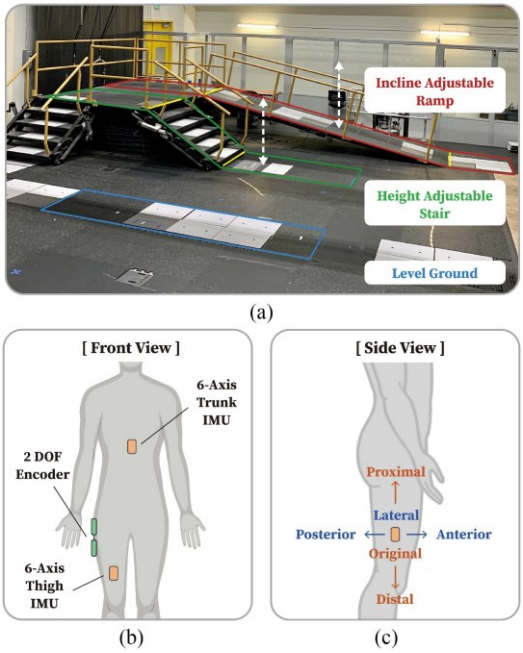

Fig. 1. Human subject experiment setup for sensor data collection. (a) In-lab height-adjustable terrain park was utilized to simulate different ramp and stair settings. (b) Simulated wearable sensors (encoder and IMUs) were located on the subject's body to represent the user's kinematic information. (c) Using the motion capture marker clusters, simulated IMU data were generated at nine different locations on the user's thigh segment.

clusters (3 makers per limb segment) that were placed on the user's body during the data collection. Using the scaled OpenSim model, gyroscopic data was computed as the rotational velocity of each relevant segment (i.e., pelvis, left thigh, and right thigh) resulting from the model's joint kinematics during each trial. The accelerometer data was computed as the linear acceleration at the specified IMU locations on each of these segments for each trial. For both trunk and thigh IMUs, the sensor orientation was specified such that the x, y, and z axes referred to the axis in vertical, lateral, and longitudinal directions, respectively. Using this method, we have replicated other sensor data (hip joint encoder and trunk IMU) because we wanted to ensure that the sensor noise was consistent across different wearable sensors and mitigate any potential confounding factors that impose inconsistency in the overall model performance. Our final dataset includes 3 wearable sensors: a 2-DOF hip joint encoder (sagittal and frontal) and a 6-axis trunk and thigh IMUs.

C. Data Processing

Unilateral mechanical sensor information from the experiment and additional thigh IMU data from the simulation were used to process the data. The user's gait phase was calculated using linear interpolation between heel strikes from 0% to 100% gait cycle, and heel strike was determined from the motion capture data as the point of zero linear velocity of the heel marker. For ramp and stair ambulations which include transition modes, the mode transition point was characterized as the user's initial heel contact with the following mode. Steady-state walking conditions at the beginning and end of each ramp and stair trial were segmented out for the data to only contain the transition gait cycles and desired mode's gait cycles. Since the instrumented leg was right-side for all subjects, each ramp trial included 4 gait cycles when starting with the left leg and 5 gait cycles when starting with the right leg with each transition gait cycle before and after the desired mode. Similar to the ramp, each stair trial contained 3 to 4 gait cycles depending on the type of starting leg and included two transition gait cycles (transition onto and off the desired mode). As a result, each processed data contained sagittal and frontal encoder data, 6-axis trunk and thigh IMU data, gait phase data, and mode labels during the time range of interest. After all the data were processed, we have standardized each channel data by normalizing each signal to a zero mean and unit variance where each subject's mean and standard deviation values were obtained from the entire trial dataset.

III. Model Implementation and Optimization

A. Benchmark Locomotion Mode Classifiers

To evaluate our DL approach, we also implemented multiple benchmark mode classifiers found in the literature [22], [25], [27], [31], [34]–[38]: 1) linear discriminant analysis (LDA), a standard practice for mode classification that finds a linear combination of features that separates different classes [39], 2) the support vector machine (SVM), a hyperplane that maximizes the margin between classes in high dimensional feature space [40], 3) the multilayer perceptron (MLP), a fully connected feedforward artificial neural network with activation functions [41], and 4) extreme gradient boosting (XGB), a decision tree-based ensemble algorithm with gradient boosting [42] (Fig. 2(a)). To ensure a fair comparison between our approach and the benchmark methods, the hyperparameters of each approach were optimized using a grid search (See Supplemental Document).

B. Deep Convolutional Neural Network-Based Classifier

DL techniques have several advantages over the conventional ML algorithms [43]. Mainly, deep and complex architectures in the neural network can have a rich feature representation that is not feasible in the traditional hand-engineered feature extraction. Additionally, deep architecture is beneficial for learning complex, intersubject representations that may not be possible under standard ML architectures. For CNN, we have explored different hyperparameters optimization routines. The baseline architecture of our CNN model was adopted from our previous study that utilized a CNN-based model for a robotic exoskeleton for gait phase estimation [44]. For CNN, a model optimization process was performed where the input window size was swept across a range from 300 ms to 1000 ms with a 100 ms increment. The CNN model was trained with the same number of epochs and stopping criteria as the MLP. For hyperparameters, the number of convolution layers (1 to 10), kernel size (50 ms, 100 ms, and 150 ms), activation function (*tanh*, ReLU, and *sigmoid*), and the dense layer's optimizer (SGD, Adam, and RMSprop) were searched. Our finalized CNN model had 5 convolutional layers with an input window size of 500 ms and a kernel size

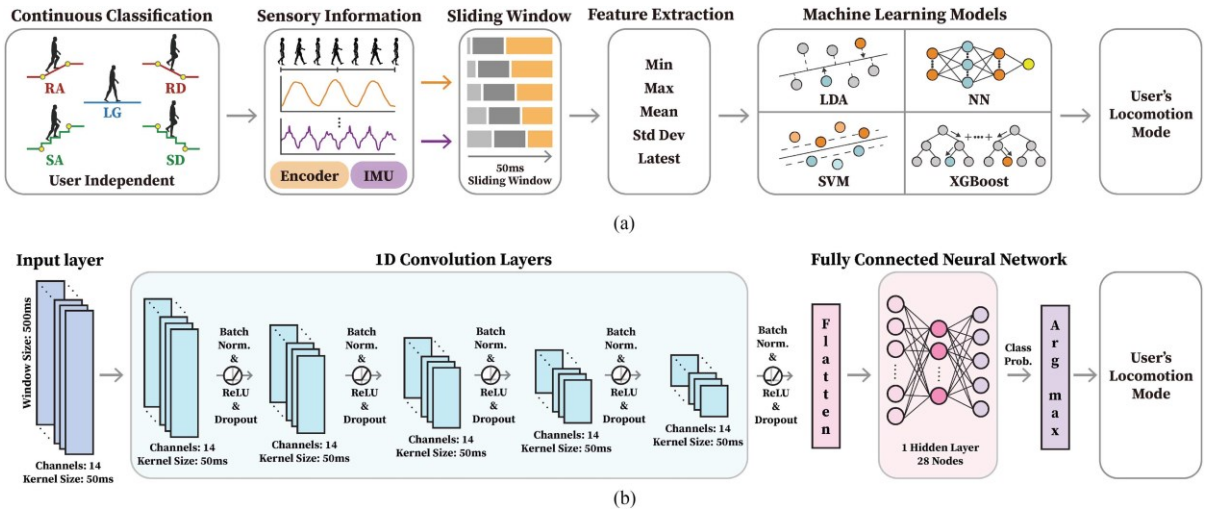

Fig. 2. (a) Conventional machine learning model pipeline for locomotion mode classification. During ambulation (level-ground, ramp ascent/descent, and stair ascent/descent), the user's kinematic information is logged from wearable sensors (joint encoder and IMUs). Five time-domain features are extracted from each sensor channel with a defined window size. Final feature sets are utilized as input to relevant machine learning models to classify the user's current locomotion mode. (b) Subject-independent continuous locomotion mode classification using a deep convolutional neural network. Information from time-series data from different wearable sensors is captured continuously with a defined input window. Five convolutional layers with a fully connected dense layer extract relevant features and map the raw sensor data to five locomotion modes.

of 50 ms (Fig. 2(b)). Each convolutional block consisted of a convolutional layer, batch normalization layer, a 20% dropout, and an activation layer (final activation being ReLU). Following the convolutional block, a flatten layer and a fully connected dense layer (1 hidden layer with an Adam optimizer). For the hidden layer at the end, the number of hidden nodes was half of the output size from the last convolutional layer (28 nodes).

IV. Model Performance Comparison

During the validation phase, we wanted to evaluate our model in four scenarios that represent the model's capacity for real-time implementation: 1) model's definition of mode transitions, 2) model's generalizability to novel environments, 3) exoskeleton's sensor placement, and 4) model's response to sensor signal disconnection. We focused on analyzing these cases because they are practical considerations for implementing a mode classifier in real-time, but often are never discussed or evaluated in prior studies in the literature. With our fully optimized hyperparameters, we have trained our subject-independent locomotion mode classifier for all five algorithms. LDA and SVM were trained using a Scikit-learn library, MLP and CNN were trained using Keras with a TensorFlow backend, and XGB was trained using the XGBoost library all based on Python 3.8.

A. Effect of Locomotion Mode Transition Label

A transitional gait cycle is a single gait cycle (with heel contact denoted as 0%) that includes a transition period between two different locomotion modes. Typically, since the intent recognition system focuses on the prediction of mode changes, ground truth labeling of two modes within this gait cycle is defined sometime during the swing phase of the gait cycle prior to heel contact on the new surface. This labeling process is critical because, depending on how the mode transition label is defined (e.g., transition happening in early swing vs late swing), the overall classification error can change substantially. Ideally, the transition label is set as early as possible (i.e., close to toe-off) without sacrificing the overall model performance. For our initial model optimization sweeps, we have fixed the transition label to occur during mid-swing (80% of the gait cycle) to ensure our model comparison procedure is consistent.

An important analysis in our study was to evaluate the effect of shifting the mode transition label on the overall classifier's performance. To achieve this, we have systematically swept the entire gait cycle during the locomotion mode transition. We have defined the transition stride to perform this sweep as the previous mode's toe-off (0%) to the next mode's toe-off (100%) with a 10% phase increment. We have further investigated the model's classification error by separating it into steady-state (error within a single steady locomotion mode) and transitional error.

B. Generalization to Different Terrain Settings

Generalizability of the model's performance across other terrain conditions such as different slope inclines is a critical factor in designing a robust mode classifier. We have evaluated our model's performance on the data with a withheld terrain condition that was not included in the training dataset. During this analysis, we have validated the model's classification error with leave-one-terrain-setting cross-validation for both ramps and stairs across all subjects.

C. Effect of Thigh IMU Sensor Location

A physical location of a thigh IMU on a robotic exoskeleton varies heavily depending on the developer's design principles.

TABLE I
CONTINUOUS LOCOMOTION MODE CLASSIFICATION RESULTS

	Our Approach (CNN)	MLP	XGB	SVM	LDA
Stead-State Error (%)	**1.16±** **0.47%*†**	6.58± 3.09%†	5.99± 2.60%†	6.48± 3.54%†	10.69± 6.00%
Transitional Error (%)	**8.58±** **1.61%*†**	14.28± 3.63%†	14.23± 2.76%†	14.15± 2.40%†	31.43± 4.49%

* represents a statistical difference between the CNN and MLP, XGB, and SVM, respectively ($p < 0.05$). † represents a statistical difference between the LDA and other models ($p < 0.05$). All results are represented as mean ± 1 standard deviation across all subjects.

As the effect of an IMU location on mode classification is unknown, we have explored this by evaluating our classifier's performance using different thigh IMU placements. As noted above, we have simulated the thigh IMU sensor data at nine possible locations (proximal/original/distal and anterior/lateral/posterior) on the subject's leg. We have validated the model's steady-state and transitional error using this simulated thigh IMU data along with original trunk IMU and encoder data.

D. Effect of Sensor Signal Drop

Sensor or channel signal drop is a common phenomenon that can happen in a real-time system potentially due to the controller's data communication failure. We have simulated this scenario to evaluate our model's robustness in maintaining the overall classification performance even with a certain sensor channel being disconnected. To simulate a sensor signal drop scenario, we changed the column (which represents a single channel in the sensor) value to 0. This analysis can also be representative of a feature importance analysis as the resulting performance can relate to which channel drop will cause a greater impact on the overall model accuracy.

E. Statistical Analysis

To test our model performance in different conditions, we conducted a two-way repeated measures analysis of variance (ANOVA) (α set to 0.05). After ANOVA, we conducted a *post hoc* analysis with a Bonferroni correction for multivariate analysis to compute the statistical difference across the 5 models (MATLAB R2020b, MathWorks, USA).

V. RESULTS

A. Optimized Locomotion Mode Classifier Performance

As shown in Table I, the optimized model had an overall classification error across all subjects of 4.20 ± 0.61%, 9.96 ± 2.32%, 9.69 ± 2.01%, 9.92 ± 2.20%, and 18.93 ± 3.75% for the CNN, MLP, XGB, SVM, and LDA, respectively.

B. Locomotion Mode Transition Label

The best performing transition label was 30%, 50%, 30%, 30%, and 60% of the gait cycle for the CNN, MLP, XGB, SVM, and LDA which corresponded to an overall classification error

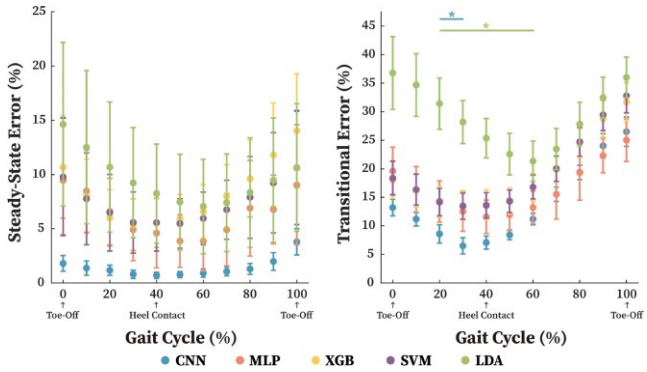

Fig. 3. Effect of locomotion mode transition label. Each model's classification error is presented for both the (left) steady-state and (right) transitional gait cycle where the gait cycle is represented from 0% (toe-off in the previous mode) to 100% (toe-off in the next mode). The error bars represent ± 1 standard deviation and asterisks indicate statistical differences ($p < 0.05$). Only statistical comparisons between the optimized transition point and baseline condition are shown.

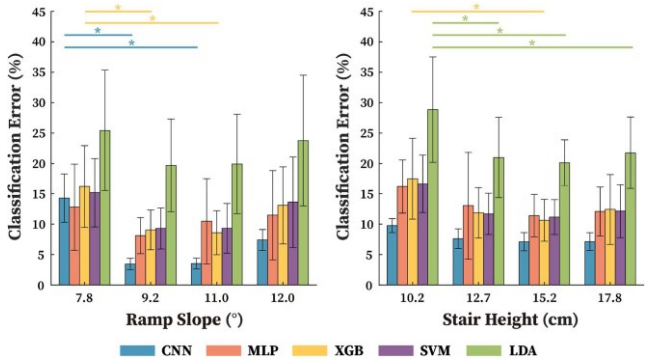

Fig. 4. Leave-one-terrain-setting evaluation for different (left) slope inclines and (right) stair heights. Results are presented with the overall classification errors. The error bars represent ± 1 standard deviation and the asterisks indicate statistical differences ($p < 0.05$). Statistical comparisons across different settings are not shown. Only statistical comparisons between settings within each model are shown.

across subjects for each algorithm of 3.13 ± 0.67%, 7.38 ± 2.10%, 9.04 ± 2.06%, 9.09 ± 1.99%, and 12.73 ± 3.27%, respectively (Fig. 3). For the CNN, the overall classification error was reduced by 24.95 ± 13.57% when the transition label was optimized compared to the baseline ($p < 0.05$), with similar relative error reductions across the other algorithms.

C. Model Generalization

For the leave-one-incline-out setting, compared to the slope incline inside the range (9.2° and 11.0°), 7.8° hold-out condition increased the overall classification error by 10.80 ± 4.23% and 7.38 ± 6.79% for the CNN and XGB, respectively ($p < 0.05$) (Fig. 4). However, no statistical differences were observed when compared to the 12° hold-out condition. For the leave-one-height-out setting, no statistical differences were observed for the CNN across all 4 hold-out conditions. For the 10.2 cm hold-out condition, the LDA increased the overall classification error by 7.91 ± 7.08% compared to other hold-out conditions,

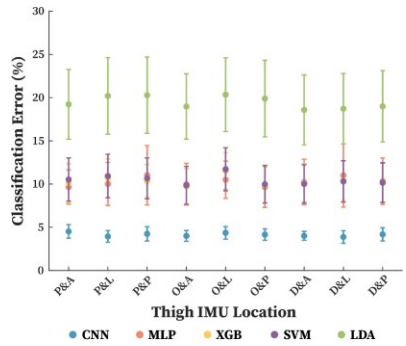

Fig. 5. Locomotion mode classifier's performance with different thigh IMU locations (P: Proximal, O: Original, D: Distal & A: Anterior, L: Lateral, P: Posterior). The error bars represent ± 1 standard deviation. The statistical comparisons across different models are not presented.

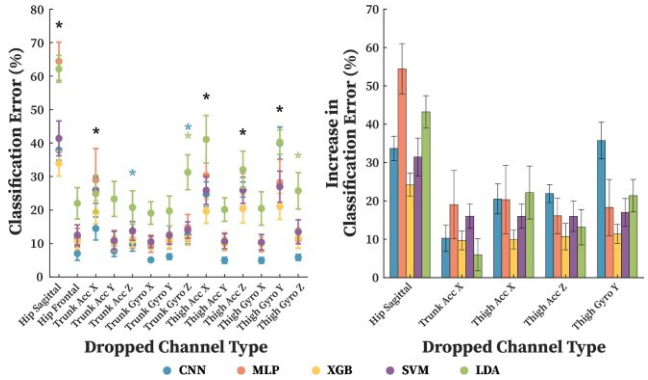

Fig. 6. Effect of different sensor signal drop on the classifier's performance. (Left) Each channel represents a scenario where the corresponding sensor signal was dropped during evaluation. (Right) Five common channels significantly influenced the classifier's performance compared to the model with no channel drop. The error bars represent ± 1 standard deviation and asterisks indicate statistical differences compared to the model with no channel drop ($p < 0.05$). The black asterisks indicate a channel where all five models showed significant differences. Statistical comparisons across models are not presented.

while the XGB only increased the classification error by 6.85 ± 6.58% compared to the 15.2 cm hold-out condition ($p < 0.05$).

D. Thigh IMU Location

For all algorithms, there were no significant differences observed across different thigh IMU locations (Fig. 5).

E. Sensor Signal Drop

Across 14 different available sensor channels, 5 channels (sagittal hip position, x-axis trunk, and thigh acceleration, z-axis thigh acceleration, and y-axis thigh gyroscope) significantly increased the overall classification error when dropped, across all algorithms ($p < 0.05$) (Fig. 6). Additionally, the CNN increased the overall classification error when either the z-axis trunk acceleration or gyroscope channel was dropped ($p < 0.05$). On average, when a single channel was dropped (for the ones that exhibited significant changes), the overall classification error was increased by 19.45 ± 11.69%, 25.65 ± 16.27%, 13.17 ± 6.20%, 19.30 ± 7.18%, and 17.93 ± 12.75% for the CNN, MLP, XGB, SVM, and LDA, respectively ($p < 0.05$).

VI. Discussion

Our study introduced a novel approach for classifying the subject's locomotion mode using a deep CNN. While several previous literature studies have explored different classification methods [22], [27], [31], [37], [38], our proposed framework is unique as our model did not require subject-specific training/tuning, classified mode continuously through the gait cycle, and only utilized proximal sensor information relating to robotic hip exoskeleton applications. The final CNN architecture had a model complexity of approximately 1.49 megaFLOPs (floating-point operations). While the CNN required a greater amount of processing time to maintain the same frame rate compared to other algorithms (LDA: 2.7 kiloFLOPs, SVM: 55.3 kiloFLOPs, XGB: 25.7 kiloFLOPs, and MLP: 7.6 kiloFLOPs), deployment to a real-time system is feasible as these FLOPs would be converted to less than 1 ms on a conventional microprocessor used in a wearable robotics system (e.g., Nvidia Jetson Nano). Furthermore, with advanced parallel processing techniques, using GPUs on these microprocessors allows for high-performance DL inference (e.g., the CNN can be optimized to a more efficient inference structure called TensorRT).

In general, the CNN outperformed the four benchmark methods in both the steady-state and transitional periods ($p < 0.05$). Among the benchmarks, the optimized MLP, XGB, and SVM exhibited a similar classification performance, where all three models outperformed the LDA ($p < 0.05$), indicating that this subject-independent and continuous classification problem is not well-resolved by linear methods, unlike discrete dependent classification in which LDA has performed well [34]. This similarity in the overall performance is potentially due to the limitation of the manual feature engineering that the three models share.

Our CNN showed better performance in multiple criteria when compared to the current state-of-the-art models in the literature [21], [45], [46]. Hu *et al.* developed an LDA-based classifier using different wearable sensors with a 1.43% overall classification error [46]. However, the study was limited as the model was mode and phase-specific (multiple classifiers being trained) and discretely classified within the gait cycle (only on toe-off and heel contact). In contrast, our CNN is a single and unified model that can continuously classify the user's mode. Liu *et al.* used an SVM-based classifier for robotic knee exoskeleton control [32]. Using only two IMUs, the model showed a 3% overall classification error. However, the study was limited as the model was user-dependent and did not include ramp modes, which tends to have the highest error rates [47]. On the other hand, our CNN is able to generalize to a novel user for all locomotion modes with similar classification performance. Lee *et al.* designed a CNN-based classifier similar to our design with a 1.1% overall classification error [45]. However, the study was limited as the system was subject-dependent (the independent model had 7.7%) and required a full lower-limb sensor suite (e.g., encoder, IMU, and myoelectric signals), which was not

the case for our CNN as we focused on developing a subject-independent classifier for hip exoskeleton applications. In the perspective of using minimal wearable sensors, Laschowski *et al.* developed a deep CNN using a body-worn camera, achieving a 5% classification error [30]. The extent of this study showcased a greater number of locomotion modes (up to 12 classes) but was limited in the overall performance for a feasible real-time implementation [48]. These comparable results in the literature illustrate that, at least for hip exoskeleton applications, our proposed approach sets a new benchmark that is applicable for real-time implementation using sensor suites that are viable (i.e., only using sensors that are native to a typical hip exoskeleton device).

While previous studies have not validated the classifier's performance to unseen terrain conditions [45], [46], evaluating the model's generalizability is important as this would resemble the model's robustness in the real-world scenario. As expected, the CNN outperformed all benchmarks by maintaining a low classification error in the leave-one-terrain-setting validation for both the ramps and stairs ($p < 0.05$). The CNN showed a greater classification error for the lowest ramp incline compared to the inclines in the interpolation region ($p < 0.05$). This phenomenon is due to the hip kinematics between the lowest incline RA and LG showing a closer overall pattern (similarly, the highest incline RA exhibits a more similar pattern as the lowest incline SA) [9]. Likewise, similar trends can be found in descent modes (e.g., the lowest incline RD resembling the LG). These similarities in the data shape resulted in a misclassification where a high incline RA and RD were predicted as SA and SD at a higher rate, while a low incline RA and RD were predicted as LG at a higher rate. However, a similar classification error increase in the extrapolation region was not shown in the leave-one-height-out validation on stairs. This is due to the SA and SD kinematic patterns being distinctively different compared to the other modes.

The effect of ground truth labeling during the mode transitions is often a feature that has not been systematically studied in past literature. Our results indicate that the model can further improve the classification performance by optimally choosing the timing point to shift the transition label during the transition stride. Moreover, it was shown that the shift of transition labels had an effect on changing the steady-state classification performance. This is because this 'shift' in transition (towards either direction) forces the model to understand the current mode as the next/previous mode even though there's no resemblance in the kinematic patterns (e.g., delayed transition shift forces the model to label the next mode into LG even though data shape is different). For our CNN, just shifting the transition label by 10% of the gait cycle (mode transition occurring at late swing or 90% of the gait cycle) vastly improved the overall performance to 3.13% classification error ($0.80 \pm 0.38\%$ steady-state error and a $6.49 \pm 1.42\%$ transitional error) compared to the baseline error of 4.2% ($p < 0.05$).

The effect of a thigh IMU location on classification accuracy for placement in a wearable device was a novel analysis of this study. However, we did not see any statistical differences in the classification error when using a sensor in a certain location. We suspect that the minor discrepancies in the classification performance across various locations are possibly due to the sensitivity difference in the signal range. For example, during the swing phase, a distally attached IMU would read a greater linear acceleration in the sagittal plane to the wider displacement of the limb segment compared to the proximally attached IMU and vice-versa during the stance phase. However, this effect was simply not strong enough to influence the model significantly, and our results indicate that placement location on the thigh is relatively inconsequential in designing devices.

During real-time inference, sensor signal drop can easily occur, likely due to communication or sensor failure. This can potentially be detrimental to the overall exoskeleton controller if the mode classifier is not robust to accommodate such abrupt changes in the input data structure. Generally, CNN maintained the overall performance even with channel data being dropped. However, 5 channels significantly increased the CNN's overall classification error when dropped ($p < 0.05$). This is because these channels represent the limb kinematics in the sagittal plane (largest variations in the data distribution), which makes it easier for the ML models to learn. One interesting aspect is that the CNN had 2 channels (z-axis trunk acceleration and gyroscope) in addition to the 5 channels that significantly impacted the overall performance when dropped ($p < 0.05$). While the trunk motion shows a minimal change in the overall movement during locomotion, the CNN was able to capture these differences to effectively leverage them. This is possible as our CNN has deep convolution layers that can extract additional features that a conventional method cannot. Another important finding is the XGB's robustness in this sensor signal drop. Among the 5 models, the XGB has the least amount of performance degradation on average when any given channel was dropped. This is due to the advantage of the XGB algorithm as it utilizes an ensemble tree boosting method, which typically tends to handle missing data well, as the penalization of the single-channel input is relatively minor in the perspective of a single subtree branch.

Similar to a sensor signal drop scenario, another common event that can induce performance degradation is inherent noise in mechanical sensors. While the main purpose of utilizing synthetic sensor data was to simulate different sensor locations, these artificial signals may not include relevant sensor noise that would be captured in a real-world scenario. To ensure our model's robustness to sensor noise, we've extended our sensor data analysis and evaluated our model performance on a dataset with a poor signal-to-noise ratio (SNR). We've synthetically induced a white Gaussian noise to our sensor channels with an SNR ranging from $0 \sim 30$ dB. For the CNN, the performance started to degrade with an SNR below 8 dB (overall classification error greater than 5%). Considering that a conventional IMU used in an exoskeleton (e.g., MPU-9250) has an SNR ranging from $30 \sim 50$ dB (which can vary depending on the sensor's clock frequency and sensitivity), our CNN was able to maintain its robustness to sensor signal noise.

There are several limitations to our study. The first is that our study only optimized and validated the model performance in an offline scenario. While the offline study allowed for a broader, in-depth analysis of our model, an online validation (including

real-time inference) is needed to fully evaluate the model performance on a hip exoskeleton in the real-world. Another limitation is that our study utilized an open-source dataset where the subject did not wear an exoskeleton. This limits understanding of the model's generalizability when assistance is provided since the user's kinematics can change depending on the assistance level. Also, a literature study showed a possible interaction effect due to misclassification (incorrect assistance to the user), potentially inducing a shift in the kinematic data causing additional misclassification, which was not accounted for in this study. However, our study is still valuable as it provides the upper bound for the locomotion mode classification and provides a meaningful guideline to researchers in initially developing a mode classifier. Moreover, the fact that our approach is subject-independent has a great value as it is the most applicable solution that can positively impact the exoskeletons currently available in the market as the model can be used as a stock controller without a need for any subject-specific fine-tuning. Future work from this study should focus on an online validation including data collected from a hip exoskeleton and evaluate the model performance with active exoskeleton control in real-time (including inference time). Additionally, as our study only utilized data from healthy young individuals (relatively small data distribution), the effect of model performance when translating our approach to other clinical populations, such as stroke survivors who exhibit different gait dynamics during locomotion, should be explored.

VII. Conclusion

We demonstrated a subject-independent and continuous locomotion mode classification strategy for a robotic hip exoskeleton application. Our framework established a new benchmark in the field with an overall accuracy of 3.13% and showcased robustness in multiple conditions that resembled real-world scenarios. Furthermore, our model only utilizes a sensor suite native to a conventional hip exoskeleton and can show a great performance to a novel user, indicating a direction for translatability of our technology to the exoskeletons available in the market. Future work from this study will focus on the implementation and online validation of our approach using a robotic hip exoskeleton.

References

[1] G. S. Sawicki *et al.*, "The exoskeleton expansion: Improving walking and running economy," *J. NeuroEngineering Rehabil.*, vol. 17, no. 1, pp. 1–9, 2020.

[2] A. J. Young and D. P. Ferris, "State of the art and future directions for lower limb robotic exoskeletons," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 2, pp. 171–182, Feb. 2017.

[3] J. Kim *et al.*, "Reducing the metabolic rate of walking and running with a versatile, portable exosuit," *Science*, vol. 365, no. 6454, pp. 668–672, 2019.

[4] J. Zhang *et al.*, "Human-in-the-loop optimization of exoskeleton assistance during walking," *Science*, vol. 356, no. 6344, pp. 1280–1284, 2017.

[5] S. Song and S. H. Collins, "Optimizing exoskeleton assistance for faster self-selected walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 786–795, 2021.

[6] G. S. Sawicki, C. L. Lewis, and D. P. Ferris, "It pays to have a spring in your step," *Exercise Sport Sci. Rev.*, vol. 37, no. 3, pp. 130–138, 2009.

[7] D. A. Winter, "Kinematic and kinetic patterns in human gait: Variability and compensating effects," *Hum. Movement Sci.*, vol. 3, no. 1/2, pp. 51–76, 1984.

[8] G. Bovi *et al.*, "A multiple-task gait analysis approach: Kinematic, kinetic and emg reference data for healthy young and adult subjects," *Gait Posture*, vol. 33, no. 1, pp. 6–13, 2011.

[9] J. Camargo *et al.*, "A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions," *J. Biomech.*, vol. 119, 2021, Art. no. 110320.

[10] Y. Ding *et al.*, "Human-in-the-loop optimization of hip assistance with a soft exosuit during walking," *Sci. Robot.*, vol. 3, no. 15, 2018, Art. no. eaar 5438.

[11] K. Seo, J. Lee, and Y. J. Park, "Autonomous hip exoskeleton saves metabolic cost of walking uphill," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, 2017, pp. 246–251.

[12] J. Jang *et al.*, "Assistance strategy for stair ascent with a robotic hip exoskeleton," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 5658–5663.

[13] G. Zeilig, H. Weingarden, M. Zwecker, I. Dudkiewicz, A. Bloch, and A. Esquenazi, "Safety and tolerance of the rewalk exoskeleton suit for ambulation by people with complete spinal cord injury: A pilot study," *J. Spinal Cord Med.*, vol. 35, no. 2, pp. 96–101, 2012.

[14] S. A. Kolakowsky-Hayner *et al.*, "Safety and feasibility of using the eksotm bionic exoskeleton to aid ambulation after spinal cord injury," *J. Spine*, vol. 4, no. 3, pp. 1–8, 2013.

[15] 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.

[16] 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, Nov. 2011.

[17] A. J. Young *et al.*, "Intent recognition in a powered lower limb prosthesis using time history information," *Ann. Biomed. Eng.*, vol. 42, no. 3, pp. 631–641, 2014.

[18] A. J. Young and L. J. Hargrove, "A classification method for user-independent intent recognition for transfemoral amputees using powered lower limb prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 2, pp. 217–225, Feb. 2016.

[19] M. Liu, D. Wang, and H. Huang, "Development of an environment-aware locomotion mode recognition system for powered lower limb prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 434–443, Apr. 2016.

[20] A. M. Simon *et al.*, "Delaying ambulation mode transition decisions improves accuracy of a flexible control system for powered knee-ankle prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, pp. 1164–1171, Aug. 2017.

[21] J. Spanias *et al.*, "Online adaptive neural control of a robotic lower limb prosthesis," *J. Neural Eng.*, vol. 15, no. 1, 2018, Art. no. 0 16015.

[22] K. Bhakta *et al.*, "Machine learning model comparisons of user independent & dependent intent recognition systems for powered prostheses," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5393–5400, Oct. 2020.

[23] K. Zhang *et al.*, "A subvision system for enhancing the environmental adaptability of the powered transfemoral prosthesis," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3285–3297, Jun. 2021.

[24] N. E. Krausz and L. J. Hargrove, "Sensor fusion of vision, kinetics, and kinematics for forward prediction during walking with a transfemoral prosthesis," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 3, pp. 813–824, Aug. 2021.

[25] E. Ceseracciu *et al.*, "SVM classification of locomotion modes using surface electromyography for applications in rehabilitation robotics," in *Proc. IEEE 19th Int. Symp. Robot Hum. Interactive Commun.*, 2010, pp. 165–170.

[26] H. Kim, Y. J. Shin, and J. Kim, "Kinematic-based locomotion mode recognition for power augmentation exoskeleton," *Int. J. Adv. Robotic Syst.*, vol. 14, no. 5, 2017, Art. no. 1729881417730321.

[27] Y. Long *et al.*, "PSO-SVM-based online locomotion mode identification for rehabilitation robotic exoskeletons," *Sensors*, vol. 16, no. 9, 2016, Art. no. 1408.

[28] C. Wang *et al.*, "A flexible lower extremity exoskeleton robot with deep locomotion mode identification," *Complexity*, vol. 2018, 2018.

[29] Q. Guo and D. Jiang, "Method for walking gait identification in a lower extremity exoskeleton based on c4. 5 decision tree algorithm," *Int. J. Adv. Robotic Syst.*, vol. 12, no. 4, p. 30, 2015.

[30] B. Laschowski *et al.*, "Preliminary design of an environment recognition system for controlling robotic lower-limb prostheses and exoskeletons," in *Proc. IEEE 16th Int. Conf. Rehabil. Robot.*, 2019, pp. 868–873.

[31] I. Kang *et al.*, "Continuous locomotion mode classification using a robotic hip exoskeleton," in *Proc. 8th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics*, 2020, pp. 376–381.
[32] X. Liu and Q. Wang, "Real-time locomotion mode recognition and assistive torque control for unilateral knee exoskeleton on different terrains," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 6, pp. 2722–2732, Dec. 2020.
[33] B. Zhong *et al.*, "Environmental context prediction for lower limb prostheses with uncertainty quantification," *IEEE Trans. Automat. Sci. Eng.*, vol. 18, no. 2, pp. 458–470, Apr. 2021.
[34] H. Huang *et al.*, "A strategy for identifying locomotion modes using surface electromyography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 65–73, 2008.
[35] A. Young, T. Kuiken, and L. Hargrove, "Analysis of using emg and mechanical sensors to enhance intent recognition in powered lower limb prostheses," *J. Neural Eng.*, vol. 11, no. 5, 2014, Art. no. 0 56021.
[36] L. J. Hargrove *et al.*, "Intuitive control of a powered prosthetic leg during ambulation: A randomized clinical trial," *JAMA*, vol. 313, no. 22, pp. 2244–2252, 2015.
[37] M. Liu, F. Zhang, and H. H. Huang, "An adaptive classification strategy for reliable locomotion mode recognition," *Sensors*, vol. 17, no. 9, 2017, Art. no. 2020.
[38] J. Camargo *et al.*, "A machine learning strategy for locomotion classification and parameter estimation using fusion of wearable sensors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1569–1578, Mar. 2021.
[39] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tut," *Inst. Signal Inf. Process.*, vol. 18, no. 1998, pp. 1–8, 1998.
[40] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.
[41] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.
[42] T. Chen *et al.*, "Xgboost: Extreme Gradient Boosting," *R. Package Version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
[43] I. Goodfellow *et al.*, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
[44] I. Kang, D. D. Molinaro, S. Duggal, Y. Chen, P. Kunapuli, and A. J. Young, "Real-time gait phase estimation for robotic hip exoskeleton control during multimodal locomotion," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 3491–3497, Apr. 2021.
[45] U. H. Lee *et al.*, "Image transformation and cnns: A strategy for encoding human locomotor intent for autonomous wearable robots," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5440–5447, Oct. 2020.
[46] B. Hu, E. Rouse, and L. Hargrove, "Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities," *Front. Robot. AI*, vol. 5, p. 78, 2018.
[47] A. J. Young, A. Simon, and L. J. Hargrove, "An intent recognition strategy for transfemoral amputee ambulation across different locomotion modes," in *Proc. IEEE 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 1587–1590.
[48] B. Laschowski *et al.*, "Computer vision and deep learning for environment-adaptive control of robotic lower-limb exoskeletons," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Medicine Soc. (EMBC)*, pp. 4631–4635, 2021, doi: 10.1109/EMBC46164.2021.9630064.