Transfer Learning for Efficient Intent Prediction in Lower-Limb Prosthetics: A Strategy for Limited Datasets

Duong Le, Shihao Cheng, Robert D. Gregg, and Maani Ghaffari

Abstract—This paper presents a transfer learning method to enhance locomotion intent prediction in novel transfemoral amputee subjects, particularly in data-sparse scenarios. Transfer learning is done with three pre-trained models trained on separate datasets: transfemoral amputees, able-bodied individuals, and a mixed dataset of both groups. Each model is subsequently fine-tuned using data from a new transfemoral amputee subject. While subject-dependent models, trained and tested using individual user data, can achieve the least error rate, they require extensive training datasets. In contrast, our transfer learning approach yields comparable error rates while requiring significantly less data. This highlights the benefit of using pre-existing, pre-trained features when data is scarce. As anticipated, the performance of transfer learning improves as more data from the subject is made available. We also explore the performance of the intent prediction system under various sensor configurations. We identify that a combination of a thigh inertial measurement unit and load cell offers a practical and efficient choice for sensor setup. These findings underscore the potential of transfer learning as a powerful tool for enhancing intent prediction accuracy for new transfemoral amputee subjects, even under data-limited conditions.
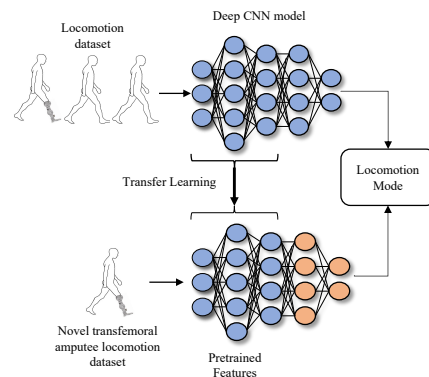
Fig. 1: Conceptual framework illustrating the application of transfer learning in locomotion intent prediction. The baseline model is initially trained on a comprehensive dataset of prosthesis users and/or non-disabled individuals. The learned patterns are then transferred and fine-tuned on smaller, user-specific data from a novel transfemoral amputee subject. This approach aims to enhance the model's prediction accuracy, reducing the need for extensive user-specific data collection.

I. INTRODUCTION

Powered lower-limb prosthetic devices can significantly enhance the quality of life for individuals with lower-limb amputations by providing net positive work [1], [2]. These advanced devices employ actuators, sensors, and microprocessors to actively control lower-limb joint(s) for various activities, including ramp/stair ascent, which requires high power input [3], [4]. The net positive work provided by the lower-limb joints minimizes the compensations from the residual limb or the intact side of the amputee users, thus resulting in a more natural gait [5], [6], improved stability [7], and reduced energy expenditure [8].

The controller plays a crucial role in the overall performance, safety, and functionality of the powered prosthetic device by regulating the interactions between the user and the prosthesis, such as generating appropriate joint torques, maintaining stability during various activities, and adapting to changes in terrain and walking conditions. User intent prediction is a vital aspect of controller development for powered prosthetic legs because it allows the prosthesis to switch between different activity modes accordingly in real-time [9]–[12]. Intent prediction methods can be broadly classi-

fied into heuristic rule-based or learning-based methods [11], [13]. Heuristic rule-based methods rely on predefined rules and thresholds to determine the user's intended actions based on the sensor signals (e.g., thigh angle, inertial measurements unit (IMU), and ground reaction forces) [14]–[19]. These methods are intuitive and can achieve good performance with hand-selected features, but they are not flexible enough to adapt to individual users or account for large variations in gait patterns [11]. Rule-based methods also sometimes involve a one-step delay [20], [21].

Learning-based methods provide a solution to the limitations of heuristic rule-based intent prediction, offering greater flexibility and adaptability to individual users and variations in gait patterns [11]. These techniques classify the user's intended movements based on the features extracted from sensory data, such as electromyography (EMG) signals [22], encoder, IMU [23], and ground reaction forces [24], and can create models tailored to the user's specific movement patterns, offering personalized predictions. Many machine learning algorithms like artificial neural networks (ANN) [25], linear discriminant analysis (LDA) [26], Support Vector Machines (SVMs) [27], dynamic Bayesian network (DBN) [26], and XGBoost [28] have achieved satisfactory recognition rates in the Vanderbilt series prostheses. Deep learning techniques, such as convolutional neural networks (CNN), have recently gained attention in locomotion intent prediction for powered prostheses or exoskeleton devices, achieving high prediction rates across a wide range of motion patterns [23], [29]–[33].

However, a primary constraint associated with these methods is the necessity for comprehensive data collection. These methods require a substantial amount of data from each user to

The authors are with the College of Engineering, University of Michigan, Ann Arbor, MI 48109, USA. {duongqle,chengsh,rdgregg,maanigj}@umich.edu

train and validate the models. Previous research suggests that a user-independent model, trained on pre-collected datasets from multiple subjects, could serve as a generalizable model for new transfemoral amputee participants [25], [26], [28]. Nonetheless, the performance of these models tends to be less accurate for new users. This decrease in accuracy is primarily due to the inherent variability among individuals, which includes distinct movement patterns, muscle activation signals, and varying physical conditions [26].

Previous research has explored adaptation techniques to improve user dependency, demonstrating promising results in personalizing models for individual users. For example, Woodward et al. [25] used a pseudo-real-time adaptation method for an artificial neural network (ANN) with transfemoral amputee participants. Their method relied on backward estimated labels, not real-time data, for updates. While this allowed for simulated real-time adaptation, the backward estimation might not accurately reflect the user's true intent. This could lead to training inaccuracies, affecting the model's real-world reliability and performance. Moreover, gathering diverse and comprehensive datasets from transfemoral amputee participants for pre-trained model development poses a significant challenge, given the time-intensive nature of data collection and the limited availability of these participants [7]. In contrast, able-bodied individuals are far more accessible for study participation, with fewer constraints or complications. In fact, numerous human locomotion datasets have been published featuring able-bodied individuals participating in various modes of locomotion captured through diverse sensor configurations. [34]–[37].

In this study, we introduce a novel approach that leverages transfer learning capabilities to enhance intent prediction accuracy for prosthesis users, especially targeting new transfemoral amputee subjects with limited labeled data available, as illustrated in Fig. 1. Transfer learning has become increasingly vital in machine learning, especially useful when data is limited [38], [39]. It adapts a model from one task to enhance performance on a related task. Transfer learning has shown promising results for hand prosthesis applications, as demonstrated in notable studies. For example, Lehmler et al. [40] showed its effectiveness in sEMG decoders with varying subject data, while Ameri et al. [41] used it to improve EMG pattern recognition robustness. These examples highlight its potential to advance intent recognition with scarce data. Our novel contribution incorporates human locomotion data from able-bodied individuals [34] for conducting pre-trained models, which supplements the limited availability of transfemoral amputee datasets. These pre-trained models are fine-tuned to cater specifically to amputee subjects, using only a small amount of user-specific data. By leveraging pre-existing knowledge from able-bodied datasets, transfer learning diminishes the requirement for large labeled datasets from transfemoral amputee participants, a task which is notably challenging and time-consuming to gather [7].

In particular, this paper has the following contributions.

1) We develop pre-trained CNN models for intent prediction, utilizing either transfemoral amputee, able-bodied, or a combination of these datasets aiming to capture typical locomotion features present in both populations.

2) We study transfer learning for intent prediction in lower-limb prosthetics by leveraging the developed pre-trained CNN models. Our results indicate enhanced intent prediction accuracy with minimal prosthesis user-specific data.

3) We evaluate and contrast the performance of the developed models against traditional subject-independent and subject-dependent models, emphasizing the advantage of using able-bodied data, especially when transfemoral amputee data is scarce.

4) We comprehensively analyze various sensor configurations applicable to powered knee-ankle prosthetic legs and investigate the impact of dataset size on the accuracy of fine-tuning transfer learning models.

The remainder of this paper is organized as follows. Section II details the processing steps for data from both able-bodied individuals and prosthesis users. Section III details the formulation of the CNN architectures, as well as the training and evaluation methods used. Sections IV and V present the results and discussions, respectively. The article is brought to a conclusion in Section VI.

II. DATA PREPARATION

A. Dataset

The primary goal of this research is to utilize transfer learning techniques to enhance the intent prediction accuracy for amputee users with limited data by leveraging a more extensive dataset from able-bodied subjects. Two datasets have been selected to achieve this objective: the Transfemoral Amputee (TF) dataset from [25] and the Able-Bodied (AB) dataset from ENABL3S [34]. These datasets have been chosen due to their structure and data type similarities, including a focus on various ambulation modes and sensor data.

1) Able-bodied Dataset: We use a comprehensive, publicly available AB dataset called the ENcyclopedia of Able-bodied Bilateral Lower Limb Locomotor Signals (ENABL3S) [34], which includes data from ten AB subjects. The dataset consists of various sensor signals, including wearable electrogoniometers (GONIO), surface electromyography (EMG), and inertial measurement unit (IMU) sensors mounted on the thigh and shank. Each subject in this dataset completed approximately 25 trials of a circuit comprising sitting, standing, level walking (LW), ramp ascent (RA), ramp descent (RD), stair ascent (SA), and stair descent (SD).

2) Transfemoral Amputee Dataset: The TF Dataset, collected at the Rehabilitation Institute of Chicago [25], comprises data from four individuals with unilateral transfemoral amputations (two males and two females), aged 32 to 69 years. Each participant was fitted with a second-generation powered knee and ankle prosthesis developed at Vanderbilt University. Data capture involved multiple mechanical sensors, including a one-degree-of-freedom (DOF) load cell, a shank-mounted inertial measurement unit (IMU), and joint encoders to record lower-limb joint kinematics and kinetics across various locomotion modes in a controlled laboratory setting. The experimental process included a series of locomotion circuits, such as level-ground walking, ramps, stairs, and

necessary transitions between these modes. Mode transitions were manually triggered. Thigh IMU data were derived from the shank IMU and knee encoder readings (angular position and velocity).

B. Data Preprocessing

We derive classification features with four different sensor selections: (i) only thigh IMU, (ii) thigh IMU and load cell, (iii) thigh IMU, load cell, and shank IMU, and (iv) thigh IMU, load cell, shank IMU, and encoders (ENC). In the AB dataset, the GONIO sensor provides knee/ankle angle and velocity signals corresponding to the TF dataset's knee/ankle ENC sensor readings. Each IMU consists of a 3-axis gyroscope and a 3-axis accelerometer, providing three signals for angular velocity and three for linear acceleration, respectively. ENC or GONIO signals yield four signals, including knee/ankle angles and velocities. The one-degree of freedom (DOF) load cell is only available in the TF dataset.

Initially, both datasets were processed and saved at 1000 Hz. Then, we resample the datasets at 5 ms intervals to reduce the number of data points. Following this step, we segment the data into sliding windows with a duration of $W_s = 900$ ms and a 50 ms frame increment between consecutive windows. This segmentation is designed to create data segments that are paired with labels that correspond to the upcoming window's activity, effectively setting our model to make predictions 50 ms ahead of actual movements. The data preprocessing pipeline, incorporating these steps, is illustrated in Fig. 2. Once processed, we standardize each channel's data by normalizing each signal to a zero mean and unit variance based on the subject-specific mean and standard deviation values obtained from the entire sample.

In this research, we aim to create a classifier that predicts LW, RA, RD, SA, and SD, as well as the transitions between them. We label an event as steady-state (SS) when the previous gait event (heel contact or toe-off) remains the same event, and we identify transitional steps (TS) when the step changes from one mode to another. This paper primarily concerns the classification of rhythmic locomotion modes. In contrast, non-rhythmic modes such as sitting and standing have been separately addressed in our prior research [42].

C. Data Splitting

This study evaluates three models to predict user intent: subject-dependent, subject-independent, and transfer learning.

1) **Subject-dependent (DEP):** We train and test a deep CNN model on each participant in the amputee dataset [25]. In this approach, we allocate 75% of the total number of circuits for training, 15% for validation, and the remaining 10% for testing for each subject. Moreover, we also evaluate the subject-dependent model using a limited number of circuits (DEP-C) to compare its performance with the transfer learning method.

2) **Subject-independent (IND):** We employ leave-one-subject-out cross-validation to train a deep CNN model on data from multiple participants. We evaluate the model's performance on an untrained participant, using training data from two subjects and validation data from one subject. We reserve the testing data for the remaining subject, the same as DEP, for testing. We repeat this process four times to comprehensively assess all subjects.

3) **Transfer learning:** We apply transfer learning to various pre-trained models that incorporate a portion of the test subject's training data, the same as the DEP model, during the training process. We evaluate three transfer learning models for three distinct pre-trained models:

   a) Prosthesis user pre-trained model (TF-1): A CNN model trained on the TF dataset (i.e., IND model).

   b) Able-Bodied pre-trained model (TF-2): A CNN model trained on the AB dataset [34].

   c) Combined pre-trained model (TF-3): A CNN model is trained using the TF and AB dataset. TF-3's base model uses leave-one-subject-out cross-validation on training with the TF dataset. To give more prominence to our target TF data, we replicate the training data multiple times to effectively create a dataset equivalent to 18 subjects.

We evaluate the effect of data size on transfer learning models by utilizing a fixed number of data circuits from the test subjects' training dataset while holding the testing and validation data constant throughout all model evaluations.

III. METHODS

A. Deep Neural Networks Model Architectures

The convolutional neural network (CNN) model has shown better effectiveness over traditional machine learning algorithms [43]. Utilizing deep, complex architectures in neural networks facilitates the creation of a comprehensive feature representation that is not achievable through conventional hand-engineered feature extraction techniques [44]. In addition, deep architectures can facilitate the acquisition of intricate inter-subject representations, which may not be attainable through conventional machine learning architectures [29]. This section introduces the CNN structure for the dependent, independent, and pre-trained models. Then, we will introduce the transfer learning method using a small number of circuits for model fine-tuning with novel prosthesis subject data.

1) Deep Convolutional Neural Network-Based Classifier: The baseline architecture of our CNN models is derived from the Fully Convolutional Network (FCN) introduced by Wang et al. [45]. The basic structure utilized in this study is composed of three CNN blocks. Fig. 3 illustrates the baseline CNN model employed in this research. The number of CNN blocks for the DEP, DEP-C, and IND models presented in Section II-C will be found through an optimization process using the Tree-Structured Parzen Estimator (TPE) Sampler. The final DEP and DEP-C models include four convolutional blocks, while the IND model implements six convolutional blocks. Each block has a convolutional layer, a batch normalization layer, and a Rectified Linear Unit (ReLU) activation layer. The initial and final blocks of the architecture have kernels of size 3 and 8, respectively, while all of the intermediate blocks have kernels of size 5. The initial and final blocks employ a filter size of 128, while all the intermediate blocks are 256. After the
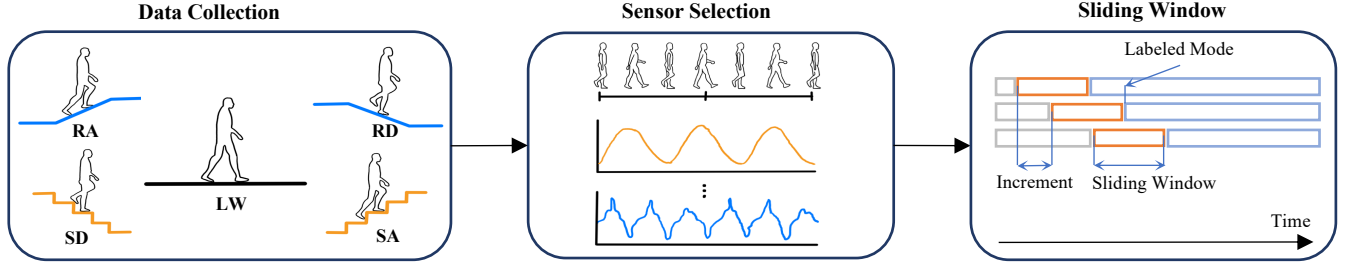
Fig. 2: Data Preprocessing Pipeline Overview for CNN Training. The pipeline starts with gathering locomotion data during various activities alongside recording the user's kinematic data from different sensors. The data segments are then used as the input for the deep CNN models, aiming to predict the user's next locomotion mode.

convolutional blocks, the features are sent to a global average pooling layer, which replaces a fully connected layer and reduces the number of weights by a significant amount. The ultimate classification is generated through the implementation of a softmax layer.

2) Transfer learning: Transfer learning is a technique that leverages the knowledge gained from a pre-trained model to improve the performance of a new model on a different but related task [39]. In this study, we refine the three models trained on different datasets using transfer learning. This process involves freezing the learned parameters in several convolutional blocks, which means their weights and biases are maintained without further updating and relearning the parameters in the remaining layers. By adopting this method, we can effectively adapt the model to novel subject data while retaining the valuable feature representations learned from the previous training. The number of transfer layers shown in Fig. 3 is found through an optimization process using a TPE Sampler. The search results reveal that freezing the first two CNN blocks for all three pre-trained models yields the best validation accuracy, which we use as the final configuration for transfer learning.

In the AB dataset, vertical load information (from load cells) is unavailable. To effectively apply the pre-trained AB model to the prosthesis user dataset, an additional CNN block must be integrated into the pre-trained model as illustrated in Fig. 3. This added CNN block has an input corresponding to the vertical load signal, while its output is concatenated with the output of the frozen layers from the pre-trained model. This additional network also features three stacked CNN blocks with filter sizes of {128, 256, 128} and kernel sizes of {8, 5, 3}, using a stride of 1. This modification enables the seamless application of the pre-trained model to the amputee dataset despite the absence of vertical load data in the original AB dataset.

B. Deep Neural Networks Model Training

This section covers how to train deep convolutional neural networks. We use a PC with NVIDIA TITAN RTX GPUs to conduct the training, and we construct and train the model using the Pytorch Lightning Library.

1) Loss function: We set up a classification problem by minimizing the cross-entropy loss, which is defined as $L = -\sum_{c=1}^{C} w_c \log \frac{\exp(x_c)}{\sum_{i=1}^{C} \exp(x_i)} y_c$. In this equation, y_c is the actual class label, corresponding to a locomotion mode, encoded
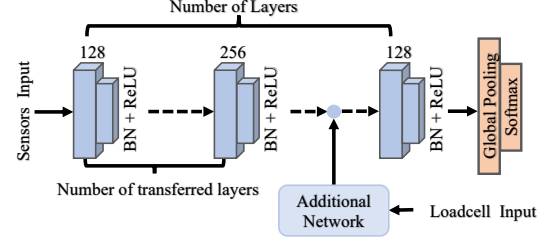

Fig. 3: The FCN network presented in the figure is composed of an array of convolutional blocks accompanied by an auxiliary network for load cell input in pre-trained models lacking load cell signals from the AB dataset.

as integers 0-4 (0 for LW, 1 for RA, etc.). x_c is the predicted logit (or score) for class c, w_c is the weight for imbalance, and C represents the total class count. The number of data samples from different classes (i.e., locomotion modes) is unbalanced. Therefore, we assign different weights to each class based on the percentage of strides for each activity and account for that in the loss function.

2) Hyperparameters: The TPE Sampler, which is a Bayesian optimization methodology from the Optuna library, is utilized to optimize the hyperparameters of the CNN model and adeptly manages the trade-off between exploration and exploitation in the quest for the most favorable hyperparameters. The process aids in identifying optimal hyperparameters, including sliding window size (W_s), number of convolutional layers, batch size, learning rate (lr), and number of transfer layers. This study also determines the optimal number of transfer layers through a systematic search that involves freezing CNN blocks sequentially until the final layer. For each iteration, the learning rate and batch sizes are optimized using the TPE Sampler to achieve the highest validation accuracy.

3) Optimizer: We use the Adaptive Moment Estimation [46] optimizer to train CNN models with the learning rate tuned to be $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1\mathrm{e}{-8}$ [45].

C. Evaluation

Similar to previous studies on locomotion intent classifiers [17], [28], [30], [34], we evaluate our model based on three types of error rates: Overall (Ovr), Steady State (SS), and Transitional State (TS). The formula for each error type is Error $= 1 - \frac{\text{Correct}}{\text{Total}}$, with "Correct" being the count of correct predictions and "Total" the total instances for each category. In our study, the computational efficiency of the model is gauged using a measure known as "megaFLOPs." The term FLOPs stands for Floating Point Operations Per Second, which
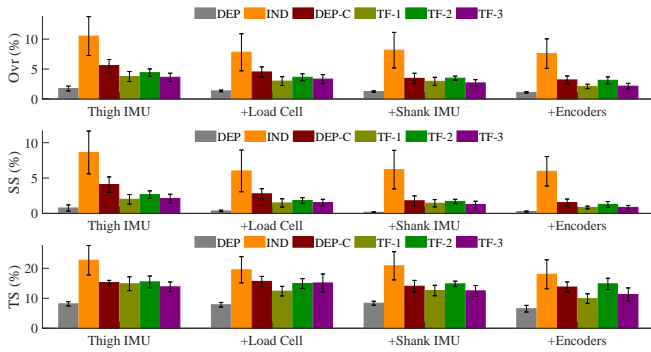
Fig. 4: Comparison of various models——subject-dependent (DEP, DEP-C), subject-independent (IND), and transfer learning (TF-1, TF-2, TF-3)——across different sensor setups and error types. DEP-C and all transfer learning models are fine-tuned using three circuits from the test subject dataset.

measures the number of floating-point calculations a computer can perform each second [29].

In addition, we perform statistical analysis to compare the model performance over two factors: learning condition (DEP, IND, DEP-C, TF-1, TF-2, TF-3) and sensor setup (thigh IMU, +load cell, +shank IMU, and +ENCs). To that end, we conduct a two-way repeated measures analysis of variance (ANOVA) with classification error as the dependent variable and learning condition and sensor setup as independent variables. Then, we followed up with a Bonferroni posthoc analysis for the significant factors to determine the statistically significant differences (p < 0.05) between each level within each factor. Finally, we ran multiple pairwise comparisons to calculate the significant difference in error rate between each specific pair of learning conditions within each sensor setup.

IV. RESULTS

A. Model Comparison

Fig. 4 compare the error rates between subject-dependent (DEP, DEP-C), subject-independent (IND), and transfer learning models (TF-1, TF-2, TF-3). The results are categorized into Ovr, SS, and TS errors for four different sensor setups. A consistent pattern across all models and sensor configurations is the higher incidence of TS errors than SS errors. The DEP model consistently yields the lowest error across all sensor choices and error categories, whereas the IND model reports the highest error. In the evaluation of models trained or fine-tuned with three circuits (DEP-C, TF-1, TF-2, and TF-3), DEP-C tends to have the highest error rates across most sensors and error types. An exception is found in the transitional state (TS) errors with Thigh IMU, +Shank IMU, and +Encoders, where TF-2 marginally exceeds DEP-C. TF-2 often incurs the highest error rates among the transfer learning models, with TF-3 only surpassing it in TS errors with the +Load Cell sensor setup. When comparing TF-1 and TF-3, their error rates are generally comparable, although TF-3 does show a notably higher TS error with the +Load Cell configuration than TF-1.

B. Sensor Selections Comparison

Fig. 4 also shows that adding more sensors resulted in improved classification accuracy across all models. However,

no improvements are significant according to the ANOVA ($p > 0.05$). The Thigh IMU sensor configuration demonstrated the highest overall error for all models, while incorporating additional sensors like Load Cell and Shank IMU led to lower error scores. Finally, adding encoders for knee and ankle angles and velocities (+ENCs) to the Thigh IMU, Load Cell, and Shank IMU configuration resulted in the lowest overall error for all models.

C. Confusion Matrix Analysis

Table I presents confusion matrices that demonstrate the classification error rates for different locomotion modes, utilizing the fourth sensor option. This sensor option for the confusion matrix is consistent with others, as no statistical differences exist between sensor options. Additionally, this choice allows for a fair comparison with other research on the same dataset [25], where all sensors are utilized. Across all models, the LW yields the lowest classification errors along the diagonal, except for DEP, where the RD has the lowest error (0.46%). In contrast, RA has the highest error rate along the diagonal for all models, except for TF-3, where the SD has the highest error (5.09%). Comparing transfer learning models (TF-1, TF-2, and TF-3) with the DEP-C model—each trained on three circuits of the dataset—it is evident that TF-1 exhibits the lowest error rate across LW (1.19%), RD (2.38%), and SD (2.94%) locomotion modes, while TF-3 shows the lowest error rate for RA (4.26%) and SA (1.85%). Conversely, the DEP-C model generally experiences the highest error across all modes, except for LW (1.36%) and RD (3.64%), marginally below TF-3's LW (1.60%) and RD (3.84%).

D. Impact of Training Data Quantity on Model Performance

This study uses the second sensor option to examine the impact of training data volume on the performance of transfer learning models (TF-1, TF-2, and TF-3) and the subject-dependent model (DEP-C, DEP). We train all models with different numbers of ambulation circuits, the details of which are outlined in Fig. 5. The DEP model, on the other hand, consistently trained on roughly 35 circuits per subject, serves as a benchmark to ascertain the data sufficiency for achieving comparable model performance. It is generally observed that prediction errors for transfer learning models and DEP-C declined as the volume of test subjects' training data increased. Notably, TF-3, TF-2, and TF-1 generally surpass DEP-C in accuracy, especially when trained with fewer circuits (2 and 3). TF-3 outperforms in minimizing Ovr and SS errors across all volumes of training data, except when trained with three circuits. Regarding TS errors, TF-1 displayed the lowest error rates in scenarios with fewer circuits (2 and 3). Conversely, TF-3 demonstrated the best performance in TS error at higher circuit counts (7 and 10). Meanwhile, DEP-C consistently had the highest error rates in almost all examples, except the seven-circuit scenario, where TF-2's transitional error marginally surpassed it. Notably, with ten circuits, the overall (Ovr) error rate of TF-3 closely approached that of the DEP model.

TABLE I: Distribution of error percentages across locomotion modes for different models utilizing the fourth sensor option. The table contains six confusion matrices that indicate the percentage of classification errors for all models. The bolded diagonal elements serve as indicators of the correct classification rates.

|		LW	RA	RD	SA	SD	LW	RA	RD	SA	SD		
DEP	LW	**1.07**	99.57	99.87	99.69	99.80	**1.19**	99.62	99.71	99.63	99.85	LW	
	RA	97.75	**2.25**	100.00	100.00	100.00	90.60	**9.40**	100.00	100.00	100.00	RA	
	RD	99.54	100.00	**0.46**	100.00	100.00	97.66	100.00	**2.38**	100.00	99.97	RD	TF-1
	SA	99.49	100.00	100.00	**0.51**	100.00	97.71	100.00	99.97	**2.32**	100.00	SA	
	SD	97.69	100.00	100.00	99.76	**2.55**	97.25	100.00	99.91	99.91	**2.94**	SD	
IND	LW	**1.94**	98.89	99.81	99.60	99.75	**1.60**	99.43	99.83	99.40	99.74	LW	
	RA	68.96	**31.31**	99.94	99.78	100.00	91.00	**9.00**	100.00	100.00	100.00	RA	
	RD	93.38	99.85	**7.16**	100.00	99.60	96.77	99.83	**3.84**	99.90	99.66	RD	TF-2
	SA	98.22	100.00	100.00	**1.93**	99.86	97.21	100.00	99.90	**2.89**	100.00	SA	
	SD	97.66	100.00	99.68	99.49	**3.00**	94.96	100.00	99.95	100.00	**5.09**	SD	
DEP-C	LW	**1.36**	99.59	99.78	99.73	99.54	**1.34**	99.33	99.79	99.78	99.77	LW	
	RA	89.20	**11.13**	99.67	100.00	100.00	95.74	**4.26**	100.00	100.00	100.00	RA	
	RD	96.60	100.00	**3.64**	99.86	99.90	96.50	99.90	**4.28**	99.93	99.39	RD	TF-3
	SA	96.06	100.00	100.00	**4.14**	99.80	98.15	100.00	100.00	**1.85**	100.00	SA	
	SD	95.47	100.00	99.58	99.81	**5.10**	95.24	100.00	99.91	99.77	**5.09**	SD	
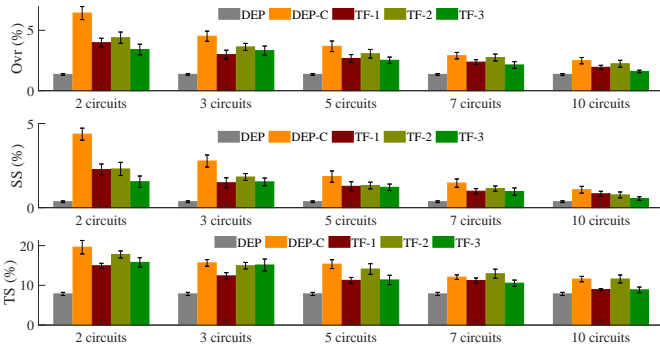
Fig. 5: Comparison of the performance of transfer learning models (TF-1, TF-2, and TF-3) and subject-dependent models (DEP, DEP-C) across varying numbers of training circuits.

V. DISCUSSIONS

A. Comparison to Prior Studies

This study uses DEP and DEP-C as benchmarks to assess our transfer learning methods, with the IND model foundational to TF-1. We compare our DEP and IND models' performance to those in studies [25], [28], setting the stage for discussing our transfer learning models' benefits over traditional methods.

Bhakata et al. [28] explored the XGBoost algorithm and reported overall errors (DEP: 3.81%, IND: 10.12%). Our study, although employing a distinct TF dataset, demonstrates improved performance with lower Ovr errors (DEP: 1.1%[0.10], IND: 7.6%[2.5].

Woodward et al. [25] utilized a Scaled Conjugate Gradient ANN on the same TF dataset, reporting lower overall errors (DEP: 0.97[0.13], IND: 2.93[0.73]) than our study (DEP: 1.1%[0.10], IND: 7.6[2.5]) with the fourth sensor option. Our analysis revealed that RA errors were the most significant contributors to classification errors in our models (DEP: 2.25%, IND: 31.31%). This contrasts with Woodward et al.'s approach, where RA was not distinctly categorized but rather merged with LW in their analysis. In addition, their standing error is notably minimal (DEP: 0.06%, IND: 1.35%), contributing to the low overall error rate. Conversely, their RD error rates (DEP: 13.35%, IND: 24.43%) are higher compared to ours (DEP: 0.71%, IND: 7.16%), while the other locomotion modes are comparable to our results.

Our CNN model's effectiveness can be attributed to its deep learning framework, which is adept at extracting detailed features for accurate locomotion classification. This approach outperforms simpler neural networks and XGBoost algorithms [25], [28] by generating advanced feature representations and capturing nuanced inter-subject variations [29], [43], [44]. This capability is key to our model's enhanced accuracy in intent prediction.

B. Comparison of Sensor Options

The results of our study suggest that adding sensor signals generally results in enhanced classification accuracy for all models. However, the ANOVA found no statistically significant differences between the sensor configurations ($p > 0.05$).

Integrating joint encoder signals and shank IMU into the classifier can enhance the intent prediction accuracy [28]–[30]. However, there are potential downsides to using encoder signals as inputs. Most significantly, the controller's output directly affects these signals. Thus, any alterations or adjustments to the controller can directly impact the classifier's performance, potentially leading to prediction instability. Furthermore, misclassification can initiate a detrimental feedback loop. If the classifier incorrectly interprets locomotion mode, this can influence the controller's output, introducing errors into the encoder and shank IMU signals.

In contrast, signals that are not primarily determined by the controller's output, such as the Thigh IMU and Load Cell, can more precisely depict the user's intentions. The mentioned signals offer a more robust and generalizable input for models that predict intent, given their reduced susceptibility to potential misclassification and alterations in the low-level controller. Most studies on the prediction of intent for prosthetic lower limbs typically incorporate the use of all available sensors, including encoders. Nevertheless, some research endeavors have attempted to employ only the thigh IMU. For instance, Bruinsma et al. [32] and Marcos et al. [23] utilized recurrent neural networks (RNN) for training the DEP model for single subjects, achieving 13.5% and 8.0% classification errors, respectively. These results are higher than our finding of DEP model using only thigh IMU (1.7%[0.4]). It is important to note, however, that the approaches in [32] and [23] involved separate labeling for each transition mode, resulting in a total of 26 classes. This difference in classification approach could partly explain these discrepancies in error rates.

C. Comparison of Transfer Learning Models

We employ three transfer learning models (TF-1, TF-2, TF-3) to enhance the accuracy of intent prediction by leveraging prior knowledge from different data sources. TF-1 and TF-3 surpass DEP-C in performance, with TF-3 showing superior results over TF-1, particularly when trained with more data circuits. This underscores the benefit of integrating AB data to enrich the model. Notably, TF-3 has comparable overall (Ovr) and steady-state (SS) performance to the DEP model, even with limited subject data (10 circuits). This suggests that TF-3 can achieve similar accuracy to a DEP model trained on a more extensive dataset, offering a promising approach for scenarios with restricted subject data availability.

TF-2 surpasses DEP-C in steady-state and overall errors, highlighting its utility of pre-existing knowledge about able-bodied locomotion patterns when extensive TF subject data collection is difficult. Moreover, when the training data is restricted to only 2 or 3 circuits, TF-2 has a significantly lower error rate, demonstrating the potential of transfer learning in situations where data from the test subject is limited.

It is important to highlight that TF-2 does not outperform DEP-C in terms of transitional state (TS) error, and similarly, the TS error of TF-3 is comparable to that of TF-1. A plausible explanation for this could be the intrinsic differences in transition steps between AB and TF subjects. In the TF dataset, transitions are often carefully adjusted by prosthetists, leading to more controlled and less natural movement patterns than in the AB dataset.

This observation suggests a key consideration for combining AB and TF datasets to improve transitional task performance. To make the most of both datasets, it may be beneficial to have transition tasks in TF subjects that mimic natural movements, similar to those observed in AB subjects [4]. This approach could potentially align the two datasets more closely, enhancing the model's ability to accurately predict transitions in a manner that reflects natural human locomotion.

D. Application to Prosthesis Control

Our proposed model employs continuous-time prediction, similar to the one described in Huang et al. [27], to substantially mitigate misclassification issues by facilitating corrections in the subsequent frame. This shows advantages over other approaches that rely on discrete prediction (a discrete window right before critical events, such as toe-off or heel contact) [28], where the accurate identification of transition mode for each stride is crucial. This continuous framework is particularly advantageous during transitions between different gait modes. While discrete methods risk prolonged misclassification effects, our model can rapidly adjust to the changing gait pattern, significantly reducing the impact of any initial misclassification [36]. Furthermore, the significantly lower steady-state error compared to transition error indicates that the model performs more reliably during steady-state phases of the gait cycle. This is particularly important for real-time control, as maintaining stability and smooth control during these phases is crucial for user comfort and safety.

Our CNN models have a model complexity of approximately 318 megaFLOPs. Although the CNN model requires substantial processing time, its deployment in a real-time system is achievable since the FLOPs can be converted to less than 50 ms on a conventional GPU-supported microprocessor (e.g., Nvidia Jetson Nano with a performance of 472 gigaFLOPs). This demonstrates the potential of applying these techniques to real-time control of prostheses.

E. Limitations and Future Directions

The data used in this study were collected in a controlled environment, and subjects were instructed to follow a pre-defined circuit for each activity. This may not fully capture the complexity and unpredictability of real-world scenarios. Additionally, our models are validated on a limited dataset, which may restrict the generalizability of our findings. Future research should collect and incorporate more diverse real-world data to enhance the model's robustness and applicability. Furthermore, implementing and testing our models in real-time control of a powered prosthetic leg is crucial for practical application. This preliminary study emphasizes model development and offline validation, motivating further development of suitable hardware for on-board testing. Future work will extend to online validation with prosthetic device data to assess performance and inference time in real-world scenarios.

VI. Conclusions

This study underscores the potential of transfer learning to enhance the accuracy of prosthesis user intent predictions, particularly with limited data from amputee subjects. By developing pre-trained CNN models using TF and/or AB datasets, we achieved enhanced model performance with substantially less amputee subject-specific data needed for fine-tuning using transfer learning, outperforming the accuracy of retraining the CNN models with the same dataset. Moreover, we identify optimal sensor configurations, endorsing a combination of a thigh inertial measurement unit and a load cell as the most effective setup. Consequently, this work is a cornerstone in leveraging AB datasets through transfer learning to facilitate more precise and individualized solutions for individuals with lower limb amputations.

References

[1] R. R. Torrealba and E. D. Fonseca-Rojas, "Toward the development of knee prostheses: Review of current active devices," *Appl. Mech. Rev.*, vol. 71, no. 3, p. 030801, 2019.

[2] C. Jayaraman, S. Hoppe-Ludwig, S. Deems-Dluhy, M. McGuire, C. Mummidisetty, R. Siegal, A. Naef, B. E. Lawson, M. Goldfarb, K. E. Gordon, *et al.*, "Impact of powered knee-ankle prosthesis on low back muscle mechanics in transfemoral amputees: A case series," *Front. Neurosci.*, vol. 12, p. 134, 2018.

[3] T. Elery, S. Rezazadeh, C. Nesler, and R. D. Gregg, "Design and validation of a powered knee–ankle prosthesis with high-torque, low-impedance actuators," *IEEE Trans. Robot.*, vol. 36, no. 6, pp. 1649–1668, 2020.

[4] S. Cheng, C. Laubscher, and R. D. Gregg, "Controlling Powered Prosthesis Kinematics over Continuous Transitions Between Walk and Stair Ascent," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2023.

[5] K. R. Embry, D. J. Villarreal, R. L. Macaluso, and R. D. Gregg, "Modeling the kinematics of human locomotion over continuously varying speeds and inclines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2342–2350, 2018.

[6] T. Lenzi, L. Hargrove, and J. Sensinger, "Speed-adaptation mechanism: Robotic prostheses can actively regulate joint torque," *IEEE Robot. Autom. Mag.*, vol. 21, no. 4, pp. 94–107, 2014.

[7] S. B. Finucane, L. J. Hargrove, and A. M. Simon, "Functional mobility training with a powered knee and ankle prosthesis," *Frontiers in rehabilitation sciences*, p. 41, 2022.

[8] H. M. Herr and A. M. Grabowski, "Bionic ankle–foot prosthesis normalizes walking gait for persons with leg amputation," *Proc. R. Soc. B: Biol. Sci.*, vol. 279, no. 1728, pp. 457–464, 2012.

[9] M. R. Tucker, J. Olivier, A. Pagel, H. Bleuler, M. Bouri, O. Lambercy, J. d. R. Millán, R. Riener, H. Vallery, and R. Gassert, "Control strategies for active lower extremity prosthetics and orthotics: a review," *J. NeuroEng. Rehabil.*, vol. 12, no. 1, pp. 1–30, 2015.

[10] A. M. Simon, K. A. Ingraham, N. P. Fey, S. B. Finucane, R. D. Lipschutz, A. J. Young, and L. J. Hargrove, "Configuring a powered knee and ankle prosthesis for transfemoral amputees within five specific ambulation modes," *PloS one*, vol. 9, no. 6, p. e99387, 2014.

[11] D. Xu and Q. Wang, "Noninvasive human-prosthesis interfaces for locomotion intent recognition: A review," *Cyborg and Bionic Systems*, vol. 2021, 2021.

[12] R. Gehlhar, M. Tucker, A. J. Young, and A. D. Ames, "A review of current state-of-the-art control methods for lower-limb powered prostheses," *Annu. Rev. Control*, 2023.

[13] L. Li, X. Wang, Q. Meng, C. Chen, J. Sun, and H. Yu, "Intelligent knee prostheses: A systematic review of control strategies," *J. Bionic Eng.*, vol. 19, no. 5, pp. 1242–1260, 2022.

[14] F. Zhang, M. Liu, and H. Huang, "Effects of locomotion mode recognition errors on volitional control of powered above-knee prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 1, pp. 64–72, 2014.

[15] H. Hobara, Y. Kobayashi, T. Nakamura, N. Yamasaki, and T. Ogata, "Foot clearance strategy for step-over-step stair climbing in transfemoral amputees," *Prosthet. Orthot. Int.*, vol. 38, no. 4, pp. 332–335, 2014.

[16] D. J. Villarreal, H. A. Poonawala, and R. D. Gregg, "A robust parameterization of human gait patterns across phase-shifting perturbations," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 3, pp. 265–278, 2016.

[17] S. Cheng, E. Bolívar-Nieto, and R. D. Gregg, "Real-time activity recognition with instantaneous characteristic features of thigh kinematics," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1827–1837, 2021.

[18] L. Ambrozic, M. Gorsic, J. Geeroms, L. Flynn, R. M. Lova, R. Kamnik, M. Munih, and N. Vitiello, "Cyberlegs: A user-oriented robotic transfemoral prosthesis with whole-body awareness control," *IEEE Robot. Autom. Mag.*, vol. 21, no. 4, pp. 82–93, 2014.

[19] A. Parri, E. Martini, J. Geeroms, L. Flynn, G. Pasquini, S. Crea, R. Molino Lova, D. Lefeber, R. Kamnik, M. Munih, *et al.*, "Whole body awareness for controlling a robotic transfemoral prosthesis," *Front. Neurorobotics*, vol. 11, p. 25, 2017.

[20] R. Stolyarov, M. Carney, and H. Herr, "Accurate heuristic terrain prediction in powered lower-limb prostheses using onboard sensors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 384–392, 2020.

[21] F. Gao, G. Liu, F. Liang, and W.-H. Liao, "Imu-based locomotion mode identification for transtibial prostheses, orthoses, and exoskeletons," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1334–1343, 2020.

[22] A. Cimolato, J. J. Driessen, L. S. Mattos, E. De Momi, M. Laffranchi, and L. De Michieli, "Emg-driven control in lower limb prostheses: A topic-based systematic review," *J. NeuroEng. Rehabil.*, vol. 19, no. 1, pp. 1–26, 2022.

[23] D. Marcos Mazon, M. Groefsema, L. R. Schomaker, and R. Carloni, "Imu-based classification of locomotion modes, transitions, and gait phases with convolutional recurrent neural networks," *Sensors*, vol. 22, no. 22, p. 8871, 2022.

[24] 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, 2009.

[25] R. B. Woodward, A. M. Simon, E. A. Seyforth, and L. J. Hargrove, "Real-time adaptation of an artificial neural network for transfemoral amputees using a powered prosthesis," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1202–1211, 2021.

[26] 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, 2015.

[27] 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, 2011.

[28] K. Bhakta, J. Camargo, L. Donovan, K. Herrin, and A. Young, "Machine learning model comparisons of user independent & dependent intent recognition systems for powered prostheses," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5393–5400, 2020.

[29] I. Kang, D. D. Molinaro, G. Choi, J. Camargo, and A. J. Young, "Subject-independent continuous locomotion mode classification for robotic hip exoskeleton applications," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 10, pp. 3234–3242, 2022.

[30] U. H. Lee, J. Bi, R. Patel, D. Fouhey, and E. Rouse, "Image transformation and cnns: A strategy for encoding human locomotor intent for autonomous wearable robots," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5440–5447, 2020.

[31] 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, 2019.

[32] J. Bruinsma and R. Carloni, "Imu-based deep neural networks: Prediction of locomotor and transition intentions of an osseointegrated transfemoral amputee," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1079–1088, 2021.

[33] J. Liu, X. Zhou, B. He, P. Li, C. Wang, and X. Wu, "A novel method for detecting misclassifications of the locomotion mode in lower-limb exoskeleton robot control," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 7779–7785, 2022.

[34] 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," *Front. Robot. AI*, vol. 5, p. 14, 2018.

[35] C. Schreiber and F. Moissenet, "A multimodal dataset of human gait at different walking speeds established on injury-free adult participants," *Scientific data*, vol. 6, no. 1, p. 111, 2019.

[36] J. Camargo, A. Ramanathan, W. Flanagan, and A. Young, "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, p. 110320, 2021.

[37] E. Reznick, K. R. Embry, R. Neuman, E. Bolívar-Nieto, N. P. Fey, and R. D. Gregg, "Lower-limb kinematics and kinetics during continuously varying human locomotion," *Scientific Data*, vol. 8, no. 1, p. 282, 2021.

[38] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[39] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)*. IEEE, 2019, pp. 47–57.

[40] S. J. Lehmler, M. Saif-ur Rehman, G. Tobias, and I. Iossifidis, "Deep transfer learning compared to subject-specific models for semg decoders," *J. Neural Eng.*, vol. 19, no. 5, p. 056039, 2022.

[41] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in emg pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, 2019.

[42] C. Welker, T. K. Best, and R. Gregg, "Improving Sit/Stand Loading Symmetry and Timing Through Unified Variable Impedance Control of a Powered Knee-Ankle Prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2023.

[43] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, pp. 1–74, 2021.

[44] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern recognition*, vol. 71, pp. 158–172, 2017.

[45] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *Proc. Int. Jt. Conf. Neural Netw.* IEEE, 2017, pp. 1578–1585.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.