IEEE Robotics and Automation Letters (RAL) paper presented at the
2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Machine Learning Model Comparisons of User Independent & Dependent Intent Recognition Systems for Powered Prostheses

Krishan Bhakta[1], Jonathan Camargo[2], *Student Member, IEEE,* Luke Donovan[3], Kinsey Herrin[4], and Aaron Young[5], *Member, IEEE*

***Abstract*— Developing intelligent prosthetic controllers to recognize user intent across users is a challenge. Machine learning algorithms present an opportunity to develop methods for predicting user's locomotion mode. Currently, linear discriminant analysis (LDA) offers the standard solution in the state-of-the-art for subject dependent models and has been used in the development of subject independent applications. However, the performance of subject independent models differ radically from their dependent counterpart. Furthermore, most of the studies limit the evaluation to a fixed terrain with individual stair height and ramp inclination. In this study, we investigated the use of the XGBoost algorithm for developing a subject independent model across 8 individuals with transfemoral amputation. We evaluated the performance of XGBoost across different stair heights and inclination angles and found that it generalizes well across preset conditions. Our findings suggest that XGBoost offers a potential benefit for both subject independent and subject dependent algorithms outperforming LDA and NN (DEP SS Error: 2.93% $\pm$ 0.49%, DEP TS Error: 7.03% $\pm$ 0.74%, IND SS Error: 10.12% $\pm$ 3.16%, and IND TS Error: 15.78% $\pm$ 2.39%)($p<0.05$). We were also able to show that with the inclusion of extra sensors the model performance could continually be improved in both user dependent and independent models ($p<0.05$). Our study provides valuable information for future intent recognition systems to make them more reliable across different users and common community ambulation modes.**

***Index Terms* — Prosthetics and exoskeletons, wearable robots, human performance augmentation, robotic prosthesis, mode classification, transfemoral amputation**

I. Introduction

Over the last two decades there have been many advancements in powered prosthetic technology that can aid users with lower-limb amputation and restore their locomotive abilities [1]–[3]. However, best practices for effectively coupling powered prostheses to individual users remain elusive. A recent challenge in creating smarter controllers is understanding how to recognize and adapt to user intent. Controllers which seamlessly decipher user intent and provide appropriate assistance will have greater viability in clinical scenarios.

Recent projections indicate that the number of individuals with lower-limb loss will increase significantly over the next several decades [4]. The steady increase of lower-limb amputations warrants the need to develop more advanced technology to allow users to ambulate more naturally and over terrains that they would often encounter in the community, such as stairs and ramps. Current solutions are mainly passive, which lack the ability to generate net positive work over a gait cycle. Hence, users develop compensatory strategies to walk which include having higher intact limb-joint moments that may lead to joint degradation, pain, and osteoarthritis [5]–[8]. Powered prostheses may help to reduce some of these compensatory strategies, but still require more advanced and reliable controller designs [9].

Recent advances on the integration of microprocessors, microcontrollers, sensors, and actuators coupled with innovative mechanical design have paved the way toward further advancing smarter prosthetic technology. These powered devices show promise in being able to help lower-limb amputees function at higher levels in their daily lives because of their ability to accommodate and provide appropriate assistance on different terrain which may in turn improve overall quality of life [10]. Robust and reliable implementation of controllers capable of accurate intent recognition (i.e. recognizing the desire to change between ambulation modes) is a non-trivial requirement given the high variability that frequently presents itself within given clinical populations.

Prosthetic control strategies have been explored to understand what techniques can be utilized to develop smarter algorithms [11]. Many research groups have focused on single lower-limb joint devices, and the lower limb prosthetic market to date only includes single joint (knee or ankle) powered technology [12], [13]. However, when more than one biological joint is missing, such as in a transfemoral amputation, an additional challenge is to ensure that two independently powered prosthetic joints can be controlled in a synchronous and stable manner. The most common prosthetic control strategies typically employ a three-tier controller paradigm: high-level, mid-level and low-level control

Manuscript received: February, 24, 2020; Revised May, 22, 2020; Accepted June, 16, 2020.

This paper was recommended for publication by Editor P. Valdastri upon evaluation of the Associate Editor and Reviewers' comments. The work was supported, in part, by a Fulbright fellowship awarded to Jonathan Camargo-Levya. This work was funded by Department of Defense Congressionally Directed Medical Research Programs (DoD CDMRP) Award No. W81XWH-17-1-0031.

[1]The authors are all part of the Exoskeleton and Intelligent Controls (EPIC) Lab. Krishan Bhakta, Jonathan Camargo, Kinsey Herrin, and Aaron Young are with the Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA, kbhakta3@gatech.edu; jon-cama@gatech.edu; kinsey.herrin@me.gatech.edu; aaron.young@me.gatech.edu

[2]Luke Donovan is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA, ldonovan7@gatech.edu

[3]Jonathan Camargo, Kinsey Herrin, and Aaron Young are with the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA, 30332 USA

Digital Object Identifier (DOI): 10.1109/LRA.2020.3007480

[11], [14].

The high-level controller is responsible for detecting and deciphering user intent (i.e. determining locomotion mode or estimating environmental variables). The mid-level controller generates a desired profile at each joint throughout the gait cycle using either torque or position laws. The low-level controller's responsibility is to ensure that the actual torque output from the motor and transmission matches the desired torque. The focus of this study was to enhance the high-level controller as it is critical that these predictions have high accuracy given their direct impact on the behavior of the other two tiers, triggering actions of the prosthesis that depend on ambulation mode. Hence, the complete response of the powered prosthesis heavily depends on the determination of the user's ambulation mode. Manual triggers to transition between different locomotion modes are non-intuitive and presents a cognitive burden to the user while walking.

Machine learning or pattern recognition techniques have been shown to classify the ambulation mode in powered prostheses from mechanical sensors [15]–[19] and neuromuscular signals [20]. These techniques have shown levels of accuracy that demonstrate potential for the application in intelligent control of detecting ambulation modes. Simon et al. implemented a mode-specific classifier that utilized a delayed mode transition decision of 90 ms while achieving an error of less than 0.5% [21]. However, as has been demonstrated in stability and error recovery studies, the response of such a system is highly sensitive to classification errors [22]. Traditionally, these implementations use methods that are relatively simple such as linear discriminant analysis (LDA) based on Bayesian theory [16], [17], [23], [24]. This offers the advantage of ease of use and fast training but are limited in capturing complex data dependencies. In addition, most of the methods have been used in a subject dependent setting, where they require training for everyone that wears the device, failing to capture patterns that are generalized across different users. Overcoming this challenge could reduce the burden of gathering lots of training data and may facilitate the adoption of smarter prostheses. In a previous study we proposed one of the first attempts of a method to produce subject independent classification that reduced the error levels to a range that allowed its use to control a prosthetic device [24]. However, we consider that additional development is needed to improve such systems, in particular with respect to a limitation that is consistently found in the literature: which is that all the training data is collected on a single ramp grade/stair height and tested on the same conditions. Furthermore, in most prior studies, classification accuracy have only been reported for a single height/incline in the training set. For real-world applications, these methods do not adequately represent community ambulation which has a larger variation in terrain, and as such, an intent recognition system must have the capability of adapting to different stair heights and inclination angles. However, this is a much more complex problem for a machine learning algorithm and is still an under-explored area of research.

Recent results in machine learning literature show the practical advantage of the gradient tree boosting method in classification problems with tabular data in complex classification scenarios [25]–[29]. Amongst different implementations, the open source package XGBoost [30] has been established as a robust solution in a wide range of problems, dominating competitions such as Kaggle [31]. XGBoost is a supervised machine learning algorithm based on gradient boosting and ensemble learning techniques. This method allows representation of the learning problem as gradient descent on an arbitrary differentiable loss function. This technique uses clever penalization of individual trees by including an additional regularization term in the loss function to combat overfitting and to improve the classification or regression output compared to its predecessors. This algorithm was selected for multiple reasons which include: 1) additive tree models that can be seen to adaptively determine the size of local neighborhoods (i.e. improves the flexibility of the fit to the data), 2) weight functions are updated at each subsequent iteration of creating the tree while taking the bias-variance tradeoff into consideration during fitting, and 3) approximating complex functional relationships using additive tree models [25], [30], [32] . In addition, this model is easily usable and efficient when training on different tasks. This decision tree boosting algorithm also allows us not to be constrained with the assumptions of Bayesian classifiers in which a certain covariance structure is specified. In the realm of exoskeletons and prostheses, several groups have implemented ensemble algorithms in gait classification tasks [33]–[36]. To the authors' knowledge, this is a new algorithm that has not been implemented in the field of lower-limb powered prostheses.

One novel aspect of this study compared to prior literature is having an expanded mechanical sensor set embedded on the prosthesis which includes 6-axis IMUs on the foot, shank and thigh, a 6-axis load cell, and joint encoders at the knee and ankle. Furthermore, our experimental paradigm allows for a unique data set to test generalizability across multiple stair heights and ramp inclination angles, and lastly exploring the use of a new algorithm for wearable robotics (XGBoost) on classifier performance. Thus, our first hypothesis is that more complex algorithms like XGBoost will improve the accuracy of the classification and scale across different grades of terrain by learning patterns that are not described by more simple methods, especially when using inputs from multiple sensors. Secondly, we hypothesized that adding extra mechanical sensors can help to improve the performance of the user-independent classifiers compared to its dependent counterparts. This study will help provide meaningful information for future development of user-intent recognition systems that can be clinically relevant.

II. METHODS

A. Powered Knee & Ankle Device

Our study utilized a powered knee and ankle prosthesis that features two independently controlled joints at the knee and ankle, providing powered assistance in the sagittal plane; a more detailed presentation of the prosthesis can be found

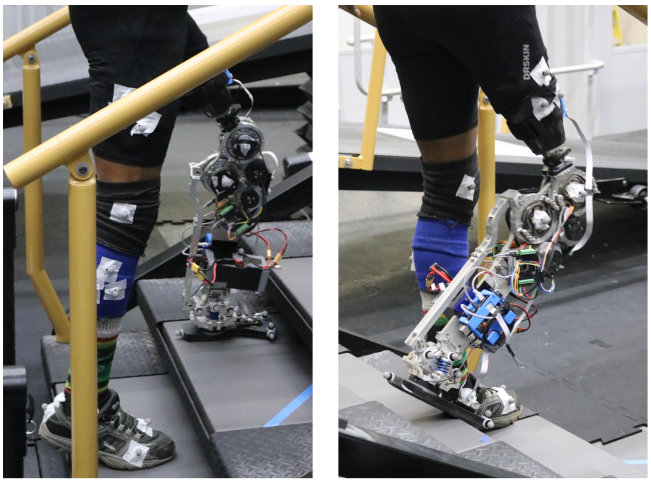

Fig. 1. Experimental setting in which one individual with transfemoral amputation is completing a stair ambulation trial across our custom-built terrain park. The terrain park is adjustable and can be modified between different stair heights and inclination angles. Embedded sensors on the prosthesis which include two joint encoders, one 6-DOF load cell, and three 6-axis inertial measurement units are useful in deciphering user intent.

in a previous paper [37], [38]. Briefly, the prosthesis includes six embedded mechanical sensors: two joint incremental encoders (US Digital E5) to measure knee and ankle kinematics, a 6-DOF (degree of freedom) load cell (SRI M3714C2) to measure ground reaction forces and moments, and three 6-axis (accelerometer & gyroscope) inertial measurement units (YOST 3-Space LX embedded) (IMUs). All sensors were collected at 100 Hz except for IMUs, which were sampled at 250 Hz. A three-tier control paradigm was implemented in this study. The low-level controller was responsible for minimizing error between the desired and actual torque profiles. The mid-level controller was an impedance controller paired with a finite state machine. Furthermore, the gait cycle was discretized into four states (early stance, late stance, swing flexion, and swing extension) for each ambulation mode [14]. Detailed approaches of how impedance parameters were tuned were based on previous literature [14], [39]. The high-level controller is responsible for predicting transitions between different ambulation modes (i.e. user intent) and estimating features of different terrain. The focus of this paper was to demonstrate a method of developing a user intent recognition system that could be utilized on wearable robotic devices.

B. Experimental Design

Eight individuals (7 males/1 female, age: 49.63 $\pm$ 13.68 years, height: 1.77 $\pm$ 0.07 m, mass: 87.31 $\pm$ 16.47 kg) with unilateral transfemoral amputation (4R/4L) were recruited and provided informed consent in accordance with the Georgia Institute of Technology Institutional Review Board. The prosthetic device was configured to each user by a certified prosthetist for appropriate comfort and alignment. The prosthetist guided the subjects in adjusting their gait to overcome any exaggerated or over-compensatory movements. When the prosthetist was satisfied with the tuning process, we conducted our collection of ambulation circuits. Users were asked to complete 2 types of ambulation circuits (ramp circuit: level walking (LW), ramp ascent (RA), & ramp descent (RD), stair circuit: level walking (LW), stair ascent (SA), & stair descent (SD)) of each preset condition in our in-lab terrain park area using our powered prosthesis. Our custom built terrain park was adjusted and set to 4 different presets for which 4 ramp trials occurred at 7.8°, 9.2°, 11.0°, and 12.4°and 4 stair trials at 10.2 cm, 12.7 cm, 15.2 cm, and 17.8 cm. Hence a total of 32 trials were collected from each subject across all the different modes. Ambulation mode labels were generated using our finite state machine. Steady state steps (SS) were identified if the previous gait event (heel contact or toe-off) remained in the same event. While transitional steps (TS) were identified if the previous gait event on the previous mode was different on the next mode (e.g. LW_LateStance to SA_SwingFlexion – was labeled as SA).

C. Data Processing and Feature Extraction

To ensure an appropriate input of data to train our mode classifiers, a general workflow was implemented to investigate and compare multiple machine learning algorithms for predicting locomotion mode. We had 6 embedded sensors on the prosthesis (2 encoders, 3 IMUs, and 1 load cell). We had 2 channels from each encoder and 6 channels from each IMU and the loadcell. We extracted 5 feature types (minimum, maximum, mean, standard deviation, and ending value) resulting in a total of 140 features (28 channels x 5 feature types = 140 total features) for a given window [15], [17], [20]. We ran two sequential forward simulations to see if certain feature types and channels were needed. It was found that removing any given feature type did not result in a decrease in classification error. Similar results were seen in the channel simulation. Hence, to be conservative, all the features were kept, since there was no evidence of overfitting from these simulations. The experimental data was neither based on timing or % gait cycle, but rather the 6-DOF loadcell to transition between phases of each ambulation mode as seen in previous studies [14]. Gait decisions were made at toe off (weight under a threshold) and heel contact (weight exceeding a threshold) to create reliable time points for transitioning the device between ambulation modes. A normalization scheme was applied first at the sensor level by dividing the load cell signal by each subject's respective weight. Furthermore, a z-score normalization was explored across all sensors on the feature level, with the load cell showing the best improvement compared to non-normalized data.

D. Locomotion Mode Classifiers

Initially, five algorithms were chosen for performance comparison in mode classification. These algorithms were linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Naive Bayes (NB), neural networks (NN), and XGBoost. The Bayesian classifiers were selected as being the current standard in the field of low-error classifiers.

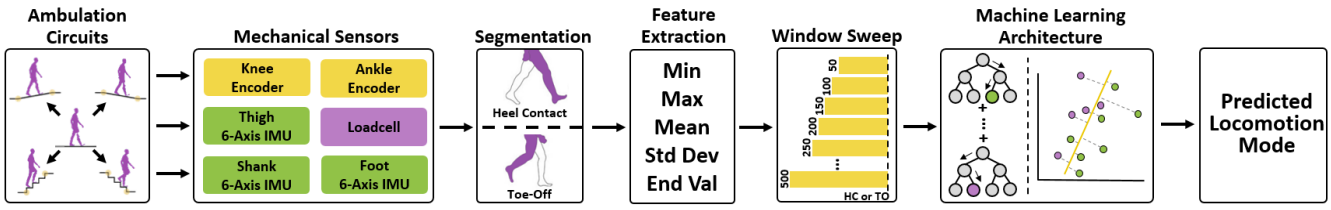

Fig. 2. The machine learning workflow used to predict user locomotion mode is shown. Users were asked to perform ambulation circuits which involved walking in 5 locomotion modes - LW, RA, RD, SA, and SD. Joint encoders that measured angular position and velocity, 6-DOF loadcell that measured ground reaction forces & moments, and IMUs that measured acceleration and rotational velocities were first segmented into two phases - heel contact and toe-off. The next algorithm consisted of transforming the data into several features (minimum, maximum, mean, standard deviation, and ending value) for a fixed window size. Next, a window size sweep was performed to find optimal window length to predict locomotion mode based on transitional error. The features were then passed through each machine learning algorithm (LDA, NN, and XGBoost) to predict locomotion mode. NN and XGBoost had to undergo an extra step of hyper-parameter tuning. In NN, layers, nodes, optimizer, learning rate, activation function, and batch size were swept. In XGBoost, learning rate, maximum tree depth, regularization term, and minimum gain were swept. After an optimized set of parameters were chosen, models were trained for our 6 different case studies to predict locomotion mode.

We ran a performance comparison across these Bayesian techniques in which LDA showed the lowest error. LDA has also been seen in prior work to be the gold standard to compare against [40]. Hence, the model comparison was reduced to 3 models: LDA, NN, and XGBoost. For each phase, a specific classifier was trained to capture the optimal transition point. Depending on the gait mode transition, the time during the gait cycle in which the classifier must make its decision is inherently gait phase specific. This strategy of transitioning is not unique and many groups have used a similar phase dependent scheme to change between gait modes based on gait events, and we adopt a similar approach here [2], [14], [17], [21].

E. Algorithm Optimization

Hyperparameter optimization of NN and XGBoost was completed to ensure model architectures were appropriate for generalizing our mechanical sensor information for the task of mode classification; LDA did not require any additional tuning. Scripts were written for all models and case studies (dependent, independent, remove-one-height/incline) and an initial window size of 250 ms was selected. We started with a directed search of unique hyperparameters for each algorithm. The subset of parameters resulting in the lowest average error between steady state and transitional error was selected until all of the parameters were swept. In XGBoost, the set of hyperparameters explored included: the learning rate used to influence the convergence to a solution, the maximum allowable tree depth, the regularization term to control the sensitivity, and minimum gain required to further split on a node in the tree. In the NN, the set of hyperparameters included: layers and nodes to determine adequate size of the network needed, optimizer and learning rate to influence the convergence to a solution, and activation function and batch size to limit model complexity.

After model optimization, a window size sweep was performed from 50 ms to 500 ms in increments of 50 ms, with the evaluation metric of taking the average of the steady state and transitional error together. The optimal window size was found to be 250 ms. Specifically, the transitional error had a minimum at 250 ms, with larger error associated with smaller or larger windows; while steady-state error was reduced the most at 250 ms and held approximately constant with larger window sizes.

TABLE I
FINAL OPTIMIZED PARAMETERS

	NN	XGBoost
DEP	Layers: 1 Nodes: 50 Optimizer: *Adam* Learning rate: 0.001 Activation function: *relu* Batch size: 32	Max depth: 1 Lambda: 0.5 Min split loss: 0 Learning rate: 0.3
IND	Layers: 3 Nodes: 10 Optimizer: *Adam* Learning rate: 0.001 Activation function: *tanh* Batch size: 128	Max depth: 3 Lambda: 1 Min split loss: 0.1 Learning rate: 0.3

F. Model Evaluation

Several steps were taken to prevent overfitting of each model. First, six different case studies were explored: dependent (DEP), independent (IND), remove-one-height dependent (RM-1-H-DEP), remove-one-height independent (RM-1-H-IND), remove-one-incline dependent (RM-1-I-DEP), and remove-one-incline independent (RM-1-I-IND). The DEP case study was similar to previous literature where data was trained on each individual subject and evaluated using a remove-one-trial cross validation. The IND case study was also taken as a traditional method of training on all users except for one which in turn became the test set. Thus, in these first two cases, both the train and test set had examples from the same stair height and ramp incline conditions. The RM-1-H-DEP and RM-1-I-DEP conditions were trained with all of the ramps and stairs conditions except for the unknown height or incline that served as the test set. This procedure was repeated until each height or incline was included in the test set. Lastly, the RM-1-H-IND and RM-1-I-IND conditions were trained with all the data except for all of the subjects' data at a specific height or incline and

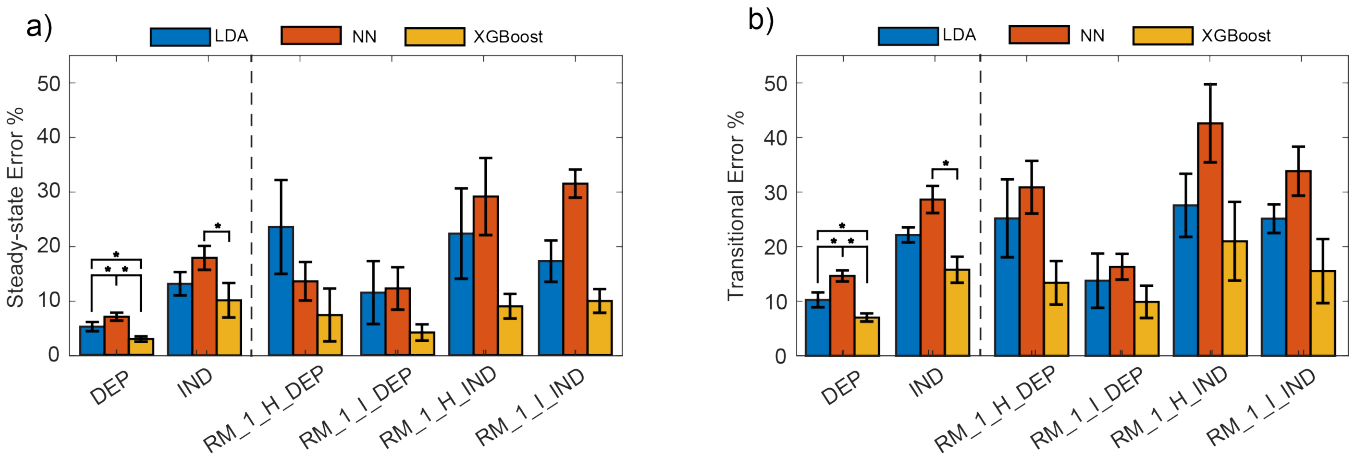

Fig. 3. Three different models (LDA/NN/XGBoost) were compared for the classification of ambulation mode in ramps, stairs and level-ground, resulting in XGBoost outperforming other models. The y-axis shows the performance metric, consisting of the error in classification during steady state walking within a mode and the error of transitioning between modes. The x-axis show the six conditions of evaluation: subject dependent (i.e. training and testing on the same subject), subject independent (i.e. training on all subjects but the testing subject), and remove-1 condition for the stair height and ramp inclination for both a) dependent and b) independent. Error bars represent $\pm$ standard error of the mean. Asterisks indicate statistical significance ($p<0.05$).

the removed subject's data. These were then tested on the removed subject's specific height or incline that was not in the train set. Similar to the RM-1-DEP cases, this procedure was repeated until each height or incline was swept. Our error criteria for evaluating the model's performance can be seen in the following equations, where SS is steady state steps, TS is transitional steps, HC is heel contact, and TO is toe off. This error was then averaged across subjects for both the steady state and transitional errors.

$$SS\_error = 1 - \frac{SS\_correct\_HC + SS\_correct\_TO}{SS\_total\_HC + SS\_total\_TO} \quad (1)$$

$$TS\_error = 1 - \frac{TS\_correct\_HC + TS\_correct\_TO}{TS\_total\_HC + TS\_total\_TO} \quad (2)$$

$$sensor\_error = 1 - \frac{SS\_error + TS\_error}{2} \quad (3)$$

G. Statistical Analysis

We conducted a one-way repeated measures analysis of variance (ANOVA) to compare the model performance across only the DEP and IND conditions (α = 0.05). The independent variable was the machine learning model (LDA/NN/XGBoost). A Dunn–Bonferroni post-hoc correction was used to compute the statistical differences between each condition (Minitab 19.0, USA).

III. RESULTS

A. Model Comparison

In the DEP case for steady state error, XGBoost (2.93% $\pm$ 0.49%) was found to be the best model compared to LDA (5.20% $\pm$ 0.85%) and NN (7.01% $\pm$ 0.73%)($p<0.05$). Similar results were found in the transitional error case, where XGBoost (7.03% $\pm$ 0.74%) had the lowest error compared to LDA (10.26% $\pm$ 1.36%) and NN (14.66% $\pm$ 1.02%)($p<0.05$). In the IND case for steady state error, XGBoost (10.12% $\pm$ 3.16%) was found only to be statistically different than NN (17.89% $\pm$ 2.19%). Similar trend was seen in the transitional error, where XGBoost (15.78% $\pm$ 2.39%) was found to be only statistically different than NN (28.65% $\pm$ 2.48%).

B. Remove-One-Preset Comparison

From the model comparison above, XGBoost was selected as our best model, and the results for Fig. 4 are only displayed for this model. Across all of the remove-one-incline conditions, the transitional and steady state error rates were consistent across incline rates. However, for remove-one-stair height conditions, the transitional and steady-state error rate decreased with large stair heights. The average steady error for user-dependent classifiers across both preset conditions was 4.59% $\pm$ 2.05%, while the transitional error was 7.60% $\pm$ 2.57%. The average steady error for user-independent classifiers across both preset conditions was 6.54% $\pm$ 1.92%, while the transitional error was 17.26% $\pm$ 4.92%.

C. Sensor Contribution

A sequential forward sensor selection using the XGBoost algorithm revealed that in both DEP and IND cases, the inclusion of all sensors yielded the lowest error ($p<0.05$). In this analysis, each sensor's features were independently tested. In the first iteration, we took one of the sensors and its features (6 total sensors, 140 total features: 6-DOF loadcell – 30 features, foot IMU - 30 features, shank IMU – 30 features, thigh IMU – 30 features, and joint encoders – 20 features) and determined which sensor, when removed from the training set of the model, would yield the highest test error implying that this was the most useful sensor needed for the mode classifier. If selected, the sensor was kept in the feature space, while the remaining sensors were tested

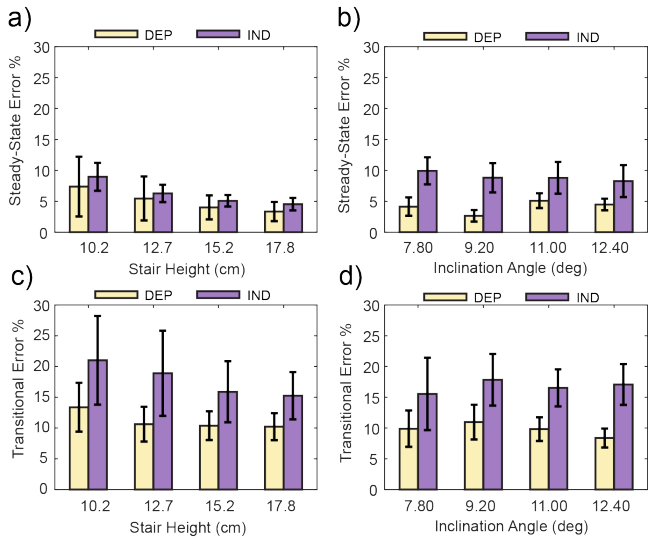

Fig. 4. XGBoost showed the best performance across all case studies. It can be seen that this model can generalize to different stair heights and inclination angles with relatively low error. The y-axes show the error for each condition - a) RM-1-H-DEP, b) RM-1-I-DEP, c) RM-1-H-IND, and d) RM-1-I-IND, while the x-axes show the 4 preset conditions for the 2 types of ambulation circuits. Results are presented for both DEP and IND cases to show how well the algorithm behaves under different validation strategies. Error bars represent the $\pm$ 1 standard error of the mean.

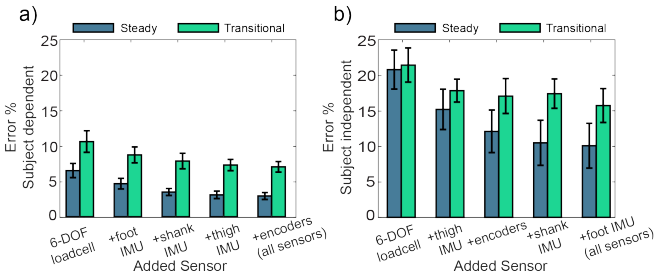

Fig. 5. The XGBoost algorithm was evaluated incrementally for each sensor that was selected on a sequential forward feature selection process. This was implemented for both a) subject dependent and b) subject independent models. In both cases, the loadcell was the most favorable sensor for locomotion mode classification. For example, in a) the first bars show the model trained with only the features of the loadcell (6 channels x 5 feature types = 30 features), the second bars show adding the foot IMU (30 previous features from loadcell + 30 new features, and so on until all the features were trained upon - 140 total features). The y-axes show the error for each condition, while the x-axes show the added sensor to the pool on each iteration. Error bars represent the $\pm$ standard error of the mean.

again in another iteration; this continued until all sensors were swept. The error metric used was the average of the steady state (SS) and transitional (TS) errors to determine what combination of sensors would yield the lowest error (Eq. 3). Across both steady state and transitional errors, the forward sensor selection algorithm chose the load cell as the best sensor that contributed the lowest error across both test (DEP and IND) cases. (Fig. 5).

D. XGBoost Confusion Matrices

We created several confusion matrices to show how the model performed across individual modes. Note LW data was present in both classifier types. We concatenated across

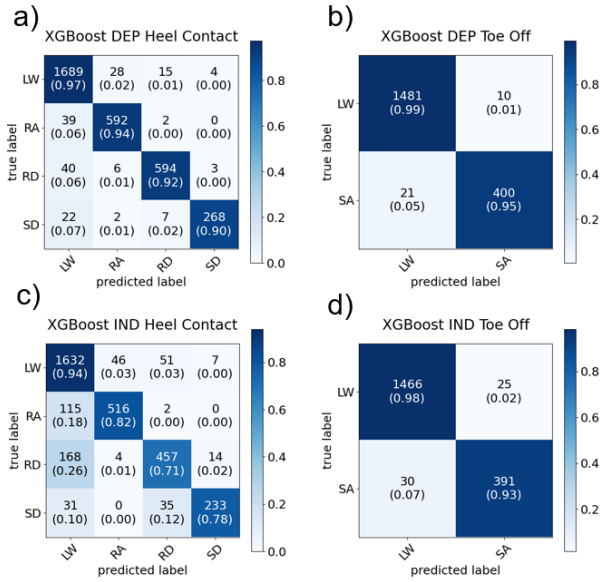

Fig. 6. Confusion matrices for our best model (XGBoost) to show individual classification accuracies for each mode and phase type - a) DEP HC classifier, b) DEP TO classifier, c) IND HC classifier, and d:) IND TO classifier). The results show that in the DEP case, XGBoost had a 96.19% classification accuracy and that in the IND case, XGBoost had a 89.89% classification accuracy. The y-axes show the true label while the x-axes show the predicted label.

subjects and combined steady state and transitional errors to show how XGBoost performed in classifying each mode. In the DEP case, XGBoost correctly classified 3170/3227 (98.23%) level walking (LW) steps, 592/633 (93.52%) ramp ascent (RA) steps, 594/643 (93.84%) ramp descent (RD) steps, 400/421 (95.01%) stair ascent (SA) steps, and 268/299 (89.63%) stair descent (SD) steps. Overall, across all modes in the DEP case, XGBoost correctly classified 5024/5223 steps (96.19%). In the IND case, XGBoost correctly classified 3098/3227 (96.00%) level walking (LW) steps, 516/633 (81.52%) ramp ascent (RA) steps, 457/643 (71.07%) ramp descent (RD) steps, 391/421 (92.87%) stair ascent (SA) steps, and 233/299 (77.93%) stair descent (SD) steps. Overall, across all modes in the IND case, XGBoost correctly classified 4695/5223 steps (89.89%).

IV. DISCUSSION

Our study explored two key features in enhancing the locomotion mode classification performance by 1) comparing different model complexities of current state-of-the-art models to XGBoost and evaluating the performance of these algorithms across users and different stair heights and inclination angles, and 2) understanding whether the user-independent classifiers with the addition of extra sensors could achieve similar performance to the dependent classifiers.

As we explored the effect of model complexity across different case studies, XGBoost outperformed both LDA and NN in the steady state error across independent and dependent models ($p<0.05$), while only outperforming NN in the transitional case ($p<0.05$). Therefore, our first hypothesis

was partially accepted in that the most complex algorithm (XGBoost) performed the best, but simply adding complexity did not yield benefits as LDA still outperformed more complex NN in certain situations, and is rejected that more complex algorithms are better for improving performance. Thus, an optimal level in complexity can reduce classification error. Across all case studies performed, XGBoost showed best performance which is one step closer to creating algorithms that can generalize across multiple grades of terrain. Note that RM-1-cases were generally more difficult for a machine learning algorithm to predict modes compared to a DEP or IND setup. This is because the classifier must generalize to a ramp incline or stair height that does not exist in the training data. Our goal was to understand if these algorithms could learn on limited data and generalize to unknown environmental conditions.

Although sensor (Fig. 5), channel (not shown) and feature type (not shown) selection were analyzed, no sensors/channels or feature types all were useful for reducing user-independent classification errors. Our second hypothesis on the addition of sensor information was accepted. It was shown that the inclusion of sensors continually improved model performance across both DEP and IND cases. Results indicated that the 6-DOF loadcell was the most essential sensor. Additional analysis could be performed to determine the influence of specific feature components from each sensor to minimize the amount of information that is extracted from the sensors while maintaining model performance.

We believe that direct comparisons cannot be made to prior literature due to our dataset's unique inclusion of multiple stair heights and inclination angles which do not exist in previous studies. Hence, we found other literature methods that created intent recognition systems; LDA and NN was commonly used as a baseline and we applied that same method to make an equal comparison to our dataset. The results indicate that XGBoost shows potential in generalizing across subjects when employing a user independent intent recognition system. We observed that XGBoost had some difficulty in differentiating between LW and RA (Fig. 6). Future work should look into combining these modes as one label as seen in prior work to see if there is an improvement in model performance [21], [24]. Similar results were seen in Young et al., where steady state and transitional errors of 8% and 13% were achieved but required more complicated dynamic bayesian networks (DBN - useful methods that incorporate time history information using current observations and prior probabilities) and mode specific architectures which are much more challenging to implement than the methods presented here [16], [24].

One limitation of our study is the small number of subjects (N=8) especially when trying to create a user independent system. The purpose of each ambulation circuit was to capture the behavior of traversing different terrain types from level walking to allow for inclusion of both steady state and mode transitional steps. However, the amount of training data is relatively small. For every ambulation circuit, there are only 4 transitional steps compared to 12-16 steady state steps. Future work is still required to address the issue of achieving smaller transitional errors with a small training dataset. Lastly, our study was limited in that it was just an offline analysis. Previous studies have indicated that implementing these models in real-time must be done properly in order to avoid the dangerous outcomes of misclassification errors [10], [15], [22]. To make these algorithms prevalent in prosthetic controllers for clinical applications, real-time validation of these models is imperative.

V. CONCLUSIONS

Our study investigated machine learning models for improving user independent and dependent locomotion mode classification. We found that XGBoost had the lowest errors compared to our other models (DEP SS Error: 2.93% $\pm$ 0.49%, DEP TS Error: 7.03% $\pm$ 0.74%, IND SS Error: 10.12% $\pm$ 3.16%, and IND TS Error: 15.78% $\pm$ 2.39%). Our approach showed that finding an optimal model complexity could improve model performance and generalize across different stair heights and inclination angles. Future work will continue to explore enhancing these classification models for them to be used in real-time applications and progress the long-term goal of promoting this technology into clinical settings.

ACKNOWLEDGMENT

We would like to thank our undergraduate team who helped with data collection and analysis which included Alice Bok and Daniel de Matheu. We would also like to thank all of our participants who took part in this study.

REFERENCES

[1] F. Sup, A. Bohara, and M. Goldfarb, "Design and control of a powered transfemoral prosthesis," *International Journal of Robotics Research*, vol. 27, no. 2, pp. 263–273, feb 2008. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364907084588

[2] B. E. Lawson, J. Mitchell, D. Truex, A. Shultz, E. Ledoux, and M. Goldfarb, "A robotic leg prosthesis: Design, control, and implementation," *IEEE Robotics and Automation Magazine*, vol. 21, no. 4, pp. 70–81, dec 2014.

[3] E. J. Rouse, L. M. Mooney, E. C. Martinez-villalpando, H. M. Herr, and M. Ieee, "Clutchable Series-Elastic Actuator : Design of a Robotic Knee Prosthesis for Minimum Energy Consumption," *2013 IEEE International Conference on Rehabilitation Robotics*, no. 1122374, 2013.

[4] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, "Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050," *Archives of Physical Medicine and Rehabilitation*, vol. 89, no. 3, pp. 422–429, mar 2008.

[5] L. Nolan and A. Lees, "The functional demands on the intact limb during walking for active trans-femoral and trans-tibial amputees," *Prosthetics and Orthotics International*, vol. 24, no. 2, pp. 117–125, aug 2000. [Online]. Available: http://journals.sagepub.com/doi/10.1080/03093640008726534

[6] R. Gailey, K. Allen, J. Castles, J. Kucharik, and M. Roeder, "Review of secondary physical conditions associated with lower-limb amputation and long-term prosthesis use," vol. 45, no. 1, 2008.

[7] K. A. Ingraham, N. P. Fey, A. M. Simon, and L. J. Hargrove, "Assessing the relative contributions of active ankle and knee assistance to the walking mechanics of transfemoral amputees using a powered prosthesis," *PLOS ONE*, vol. 11, no. 1, p. 0147661, jan 2016. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0147661

[8] D. C. Morgenroth, M. Roland, A. L. Pruziner, and J. M. Czerniecki, "Transfemoral amputee intact limb loading and compensatory gait mechanics during down slope ambulation and the effect of prosthetic knee mechanisms," *Clinical Biomechanics*, vol. 55, pp. 65–72, jun 2018.

[9] A. M. Simon, N. P. Fey, K. A. Ingraham, S. B. Finucane, E. G. Halsne, and L. J. Hargrove, "Improved Weight-Bearing Symmetry for Transfemoral Amputees during Standing Up and Sitting Down with a Powered Knee-Ankle Prosthesis," *Archives of Physical Medicine and Rehabilitation*, vol. 97, no. 7, pp. 1100–1106, jul 2016.

[10] L. J. Hargrove, A. J. Young, A. M. Simon, N. P. Fey, R. D. Lipschutz, S. B. Finucane, E. G. Halsne, K. A. Ingraham, and T. A. Kuiken, "Intuitive control of a powered prosthetic leg during ambulation: a randomized clinical trial," *JAMA*, vol. 313, no. 22, pp. 2244–2252, 2015.

[11] M. R. Tucker, J. Olivier, A. Pagel, H. Bleuler, M. Bouri, O. Lambercy, J. R. Del Millán, R. Riener, H. Vallery, and R. Gassert, "Control strategies for active lower extremity prosthetics and orthotics: A review," pp. 1–30, jan 2015.

[12] H. M. Herr and A. M. Grabowski, "Bionic ankle–foot prosthesis normalizes walking gait for persons with leg amputation," *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, no. 1728, pp. 457–464, feb 2012. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rspb.2011.1194

[13] E. Russell Esposito, J. M. Aldridge Whitehead, and J. M. Wilken, "Step-to-step transition work during level and inclined walking using passive and powered ankle–foot prostheses," *Prosthetics and Orthotics International*, vol. 40, no. 3, pp. 311–319, jun 2016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0309364614564021

[14] 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. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0099387

[15] H. A. Varol, F. Sup, and M. Goldfarb, "Multiclass real-time intent recognition of a powered lower limb prosthesis," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 542–551, mar 2010.

[16] A. J. Young, A. M. Simon, N. P. Fey, and L. J. Hargrove, "Intent recognition in apowered lower limb prosthesis using time history information," *Annals of Biomedical Engineering*, vol. 42, no. 3, pp. 631–641, mar 2013. [Online]. Available: http://link.springer.com/10.1007/s10439-013-0909-0

[17] A. J. Young, A. M. Simon, and L. J. Hargrove, "A training method for locomotion mode prediction using powered lower limb prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 671–677, 2014.

[18] 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 Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1032–1042, may 2019.

[19] R. B. Woodward, J. A. Spanias, and L. J. Hargrove, "User intent prediction with a scaled conjugate gradient trained artificial neural network for lower limb amputees using a powered prosthesis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2016-Octob. Institute of Electrical and Electronics Engineers Inc., oct 2016, pp. 6405–6408.

[20] 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 Transactions on Biomedical Engineering*, vol. 58, no. 10 PART 1, pp. 2867–2875, oct 2011.

[21] A. M. Simon, K. A. Ingraham, J. A. Spanias, A. J. Young, S. B. Finucane, E. G. Halsne, and L. J. Hargrove, "Delaying Ambulation Mode Transition Decisions Improves Accuracy of a Flexible Control System for Powered Knee-Ankle Prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1164–1171, 2017.

[22] J. A. Spanias, A. M. Simon, S. B. Finucane, E. J. Perreault, and L. J. Hargrove, "Online adaptive neural control of a robotic lower limb prosthesis," *Journal of Neural Engineering*, vol. 15, no. 1, p. 016015, feb 2018. [Online]. Available: http://iopscience.iop.org/article/10.1088/1741-2552/aa92a8

[23] L. J. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 847–853, may 2007.

[24] A. J. Young and L. J. Hargrove, "A Classification Method for User-Independent Intent Recognition for Transfemoral Amputees Using Powered Lower Limb Prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 2, pp. 217–225, feb 2016.

[25] D. Nielsen, "Tree Boosting With XGBoost Why Does XGBoost Win "Every" Machine Learning Competition?" Tech. Rep., 2016.

[26] B. Semiz, S. Hersek, D. C. Whittingslow, L. A. Ponder, S. Prahalad, and O. T. Inan, "Using Knee Acoustical Emissions for Sensing Joint Health in Patients with Juvenile Idiopathic Arthritis: A Pilot Study," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9128–9136, 2018.

[27] H. Lu, M. Pinaroc, M. Lv, S. Sun, H. Han, and R. C. Shah, "Locomotion recognition using XGboost and neural network ensemble," in *UbiComp/ISWC 2019- - Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. Association for Computing Machinery, Inc, sep 2019, pp. 757–760.

[28] Y. Ye, C. Liu, N. Zemiti, and C. Yang, "Optimal Feature Selection for EMG-Based Finger Force Estimation Using LightGBM Model," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2019*. Institute of Electrical and Electronics Engineers Inc., oct 2019.

[29] A. Kadrolkar and F. C. Sup, "Intent recognition of torso motion using wavelet transform feature extraction and linear discriminant analysis ensemble classification," *Biomedical Signal Processing and Control*, vol. 38, pp. 250–264, sep 2017.

[30] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-Augu. Association for Computing Machinery, aug 2016, pp. 785–794.

[31] S. Ben Taieb and R. J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," *International Journal of Forecasting*, vol. 30, no. 2, pp. 382–394, apr 2014.

[32] R. Mitchell and E. Frank, "Accelerating the XGBoost algorithm using GPU computing," *PeerJ Computer Science*, vol. 2017, no. 7, p. e127, jul 2017.

[33] N. Shawen, L. Lonini, C. K. Mummidisetty, I. Shparii, M. V. Albert, K. Kording, and A. Jayaraman, "Fall Detection in Individuals With Lower Limb Amputations Using Mobile Phones: Machine Learning Enhances Robustness for Real-World Applications," *JMIR mHealth and uHealth*, vol. 5, no. 10, p. e151, oct 2017.

[34] C. Wang, X. Wu, Y. Ma, G. Wu, and Y. Luo, "A Flexible Lower Extremity Exoskeleton Robot with Deep Locomotion Mode Identification," *Hindawi*, p. 9, 2018. [Online]. Available: https://doi.org/10.1155/2018/5712108

[35] F. Peng, W. Peng, and C. Zhang, "Evaluation of sEMG-Based Feature Extraction and Effective Classification Method for Gait Phase Detection," in *Communications in Computer and Information Science*, vol. 1006. Springer Verlag, nov 2019, pp. 138–149.

[36] S. Nakagome, T. P. Luu, Y. He, A. S. Ravindran, and J. L. Contreras-Vidal, "An empirical comparison of neural networks and machine learning algorithms for EEG gait decoding," *Scientific Reports*, vol. 10, no. 1, pp. 1–17, dec 2020.

[37] H. Zhao, E. Ambrose, and A. D. Ames, "Preliminary results on energy efficient 3D prosthetic walking with a powered compliant transfemoral prosthesis," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1140–1147.

[38] K. Bhakta, J. Camargo, and A. J. Young, "Control and experimental validation of a powered knee and ankle prosthetic device," in *ASME 2018 Dynamic Systems and Control Conference, DSCC 2018*, vol. 1. American Society of Mechanical Engineers (ASME), nov 2018.

[39] K. Bhakta, J. Camargo, P. Kunapuli, L. Childers, and A. Young, "Impedance Control Strategies for Enhancing Sloped and Level Walking Capabilities for Individuals with Transfemoral Amputation Using a Powered Multi-Joint Prosthesis," *Military Medicine*, vol. 185, no. Supplement_1, pp. 490–499, jan 2020.

[40] K. Englehart and B. Hudgins, "A Robust, Real-Time Control Scheme for Multifunction Myoelectric Control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848–854, 2003.