2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS) | 978-1-6654-5238-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICHMS56717.2022.9980676

Cognitive Workload Classification of Upper-limb Prosthetic Devices

Junho Park
Wm Michael Barnes '64
Department of Industrial &
Systems Engineering
Texas A&M University
College Station, TX, USA
junho.park@tamu.edu

Joseph Berman
Joint department of Biomedical
Engineering
North Carolina State University
and University of North Carolina
at Chapel Hill
Raleigh, NC, USA
jmberman@ncsu.edu

Albert Dodson
Joint department of Biomedical
Engineering
North Carolina State University
and University of North Carolina
at Chapel Hill
Raleigh, NC, USA
aedodson@ncsu.edu

Yunmei Liu
Department of Industrial and
Systems Engineering
University of Florida
Gainesville, FL, USA
liu.yunmei@ufl.edu

Armstrong Matthew
Intercollegiate School of
Engineering Medicine
Texas A&M University
Houston, TX, USA.
matt_armstrong12@tamu.edu

He Huang
Joint department of Biomedical
Engineering
North Carolina State University
and University of North Carolina
at Chapel Hill
Raleigh, NC, USA
hhuang11@ncsu.edu

David Kaber
Department of Industrial and
Systems Engineering
University of Florida
Gainesville, FL, USA
dkaber@ise.ufl.edu

Jaime Ruiz
Department of Computer &
Information Science &
Engineering
University of Florida
Gainesville, FL, USA
jaime.ruiz@ufl.edu

Maryam Zahabi
Wm Michael Barnes '64
Department of Industrial &
Systems Engineering
Texas A&M University
College Station, TX, USA
mzahabi@tamu.edu

***Abstract*— Limb amputation can cause severe functional disability in performing activities of daily living (ADLs). Using prosthetic devices as aids for such activities requires substantial cognitive resources. Machine Learning (ML) algorithms can be used to predict cognitive workload (CW) of prosthetic device prototypes early in the design process and serve as a tool for improving device usability. The objective of this study was to explore subsets of input features that can be easily captured during early stages of the design cycle to classify CW of electromyography (EMG)-based upper-limb prostheses. An experiment was conducted with 30 participants to collect task performance and pupillometry data, and to provide a basis for generating cognitive performance model (CPM) outcomes. Three ML algorithms, including the random forest (RF), support vector machine (SVM), and naïve Bayesian (NB) classifier were developed. The most important subset of features was selected based on classification accuracy and computational and experimental cost. Findings revealed that the CPM outcomes and prosthetic device configuration were the most important features for reasonably classifying CW responses under low cost. Also, the SVM classifier can be used for near-real time classification of CW. Future studies should include additional data and improve hyperparameter tuning parameters, as well as advanced CPM techniques to improve the performance of algorithms.**

Keywords—prosthesis, cognitive workload, machine learning, cognitive performance modeling, classification

I. Introduction

Approximately 2.1 million Americans live with amputations and about 185,000 amputation surgeries performed each year [1]. Limb amputation can cause severe functional disability for the performance of activities of daily living (ADLs). Amputees use prosthetic devices on a regular basis to perform ADLs. Prosthetic devices require substantial amount of cognitive resources [2], which can lead to device rejection. Prosthetic devices with high cognitive workload (CW) can have a negative impact on task performance, which can reduce user satisfaction and device usability [3]. Thus, assessing CW of prostheses are critical to ensure device usability [4].

To classify CW, machine learning (ML) algorithms can be used with several advantages compared to inferential statistics such as ANOVA. First, ML algorithms can be used to find relationships among features in high dimensional spaces and deal with non-linear factors and uncertainty without strict assumptions in inferential statistics [5]. Second, the method allows for classification of CW in near real-time [6]. With these advantages, several ML algorithms have been used to classify CW of operators in various domains such as construction or aviation. The most frequently used methods were support vector machine (SVM) [7], random forest (RF) [8], and Naïve Bayes (NB) algorithms [9]. A majority of

This project was funded by the National Science Foundation (No. IIS-1856676/1856441/1900044).

978-1-6654-5238-0/22/$31.00 ©2022 IEEE

studies used physiological measurements (e.g., heart rate) as input features to classify CW [10, 11] and some used task performance outcomes (e.g., task completion time) [12, 13]. However, prior studies had several limitations. First, there has not been any investigation on classification of CW for prosthetic devices, although high CW is one of the major challenges with existing prosthetic devices. Second, although several measures such as physiological responses, task performance, and subjective responses have been used as input features in CW classification algorithms, no study used cognitive performance model (CPM) generated outcomes as input features to classify CW. CPM models and their outcomes can be generated by observation of different tasks and using knowledge elicitation approaches with small sample size and do not require extensive human subject experiments, and therefore can be used in early stages of the design cycle [14]. Third, there were limited number studies exploring the effect of a subset of features on ML outcomes. Some studies tested subsets of features, however, they are limited to only physiological [12] or task performance data [6]. Therefore, this research aimed to investigate a subset of multimodal input features to classify CW in using electromyography (EMG)-based prosthetic devices with acceptable accuracy and low computational and experimental cost.

II. Method

A. Human-subject experiment

Thirty able-bodied participants (18 males and 12 females) were recruited for this study (Age: *M*=22.9 yrs.; *SD*=2.8 yrs.). All participants had 20/20 vision without prior experience of participating in studies with prostheses or myoelectric exoskeleton for upper-limbs. The experiment protocol was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill. A commercial 2-DoF (Degree of Freedom) prosthetic device (Motion Control ETD, Filauer) in hand open/close and wrist pronation/supination was used with three control modes including: direct control (DC), pattern recognition (PR), and continuous control (CC). A custom prosthetic hand adapter was designed and fabricated as a bypass device, as shown in Figure 1 (Left).

For the DC mode, EMG signals were collected from two channels (hand close/wrist pronation for the flexor carpi radialis; hand open/wrist supination for the extensor carpi radialis longus) based on the mean absolute value (MAV) of each channel [15, 16]. Participants could only control the hook with one DoF (i.e., either rotation or open/close) by wrist flexion and extension. The experimenter manually adjusted thresholds and proportional control gains for each channel based on feedback from the participant. Participants were trained with five hand gestures (hand close/open, wrist pronation/supination, inactive) in the PR mode. EMG data were collected and labeled simultaneously with a certain movement class. Four commonly used time domain features (MAV, number of zero crossings, waveform length, and number of slope sign changes) were extracted from EMG signals following the methods used in our prior studies [15, 16]. A Linear discriminant analysis (LDA)-based classifier was trained based on the features and labels to predict one of the five-movement classes. The speed was set proportional to the sum of the magnitudes of all EMG signals. During the calibration, the experimenter manually adjusted control gains based on the classification performance. In the CC mode, EMG data were recorded simultaneously with kinematic data from a Leap Motion Controller (Leap Motion, Inc., USA). A camera was used to accurately estimate the positions of segments in the hand and forearm [17, 18]. Estimated positions of the phalangeal, palm, and forearm segments were recorded at 120 Hz and used to calculate wrist pronation/supination and metacarpophalangeal (MCP) flexion/extension joint angles. Muscle activations were estimated from the recorded EMG signals. Training data was collected from participants with three gestures: MCP flexion/extension only, wrist pronation/supination only, and simultaneous wrist and MCP. An artificial neural network was created for each participant using the Deep Learning Toolbox in MATLAB 2018b (Mathworks Inc., USA). Pupil dilation and blink rate were measured using the Pupil-Core eye tracking system (Pupil Labs, Germany).

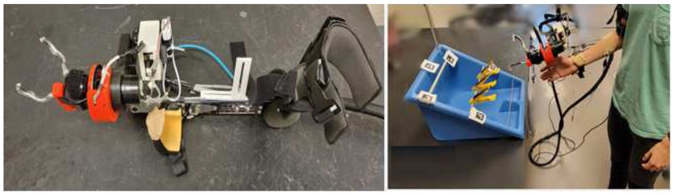


Figure 1. Prosthetic device (Motion Control ETB, Filauer, 4.54lb) (Left), A participant trying to pick up a pin (Right)

The experiment followed a between-subject design in which each participant was randomly assigned to one of the three prosthetic configurations (i.e., DC, PR, or CC) to avoid potential fatigue or learning effect from one trial to the next. Clothespin Relocation Test (CRT) was used as an ADL in this experiment as it is a widely used ADL for assessing usability of upper limb prostheses [19]. The CRT required participants to move as many pins as possible from the horizontal bar to the vertical bar and vice versa in 2 minutes with various hand gestures in each control mode [20] (Figure 1 – Right).

CW of participants was measured by primary task performance, pupillometry data, CPM outcomes, and perceived workload ratings. Task performance was captured by counting the numbers of pins moved and the shortest time to move one pin from one bar to another in each trial. The number of training trials to achieve mastery in the task was also included as another task performance measure. To develop CPMs, hierarchical task analysis (HTA) trees were initially created for each type of control mode. Based on the findings of the HTA, three Cognitive-Perceptual-Motor GOMS (CPM-GOMS) [21] models were developed in Cogulator for each control scheme [22]. The models generated outcomes including the task completion time (TCT) for one cycle, number of cognitive, perceptual, and motor operators, and the number of memory chunks required to perform the task. Each participant rated their CW using the NASA Task Load Index (NASA-TLX) scale after each trial and it was used as a ground truth or target variable in the ML algorithm, as this measure has been used extensively in prior studies using prosthetic devices [23, 24].

978-1-6654-5238-0/22/$31.00 ©2022 IEEE

Once participants arrived at the lab, they signed the informed consent form, and filled out the demographic questionnaire. Then, they completed the Edinburgh Handedness Test (EHT) [25] and the Purdue Pegboard Test (PPT) [26, 16] to ensure that they were eligible for the study with a certain level of hand dexterity. Once participants completed the EHT and PPT, they were equipped with the prosthesis and EMG electrodes were placed on their skin based on the assigned control mode. Participants were allowed to interact with the device until they reported comfort with the control mode and the classifier was sufficiently trained. Once the participants received training for their assigned control mode, they were trained on the CRT, which assessed the mastery of device handling and the respective control mode. If the average TCT to move three pins in three sequential trials was within 15–25s for the PR, 20–35s for the DC, and 16-23s for the CC mode, the participant was allowed to proceed to the experimental trials. These thresholds were defined based on our prior studies with a similar prosthetic device. Upon completion of the training trials, the eye-tracking system was calibrated for the participants, and they could begin the experimental trials after having 5 minutes of rest. In experimental trials, participants were instructed to move as many clothespins as possible between the two bars within 2 minutes. All participants completed three trials with a 5-min rest period after each trial. After each trial, participants filled out the NASA-TLX questionnaire.

B. *Cognitive Workload Classification*

Three algorithms (RF, SVM, and NB) were selected to classify CW as they exhibited high prediction accuracy (> 80%) with small datasets in previous studies [27]. The prediction outcome was the level of CW which was categorized in three classes of "high", "moderate", and "low" CW based on the findings of the clustering analysis on NASA-TLX scores collected from the experiment.

The dataset included 90 datapoints (10 participants per each control scheme ✕ 3 control schemes ✕ 3 trials). To avoid overfitting, we split the dataset into training (70% of the data) and testing (30% of the data) groups. To ensure generalizability of the predictive algorithms, 10-folds cross validation (CV) was employed to optimize the number of hyperparameters [28]. Across 10-folds CV, a hyperparameter grid search method was used employing the *sklearn* Python library [29] and a *pipeline* function to streamline testing across three classifiers. This is because scanning over every permutation of a wide range of values for each set of hyperparameters increases out-of-sample test performance [30].

Feature selection methods were used to eliminate less-contributory features from the dataset, to make modeling more efficient, and to improve classification accuracy. K-Best method was used as a representative method of the univariate filter class of selectors [31]. K-Best estimates and stratifies the contribution of each variable to the target class and chooses the *k* (number of features specified) best ranking features to model from. In addition, the recursive feature elimination (RFE) and sequential feature selection (SFS) methods were employed as members of the wrapper class of selectors [32, 33]. RFE, similar to backwards feature selection, considers multivariate feature contribution as a whole and iteratively removes the least contributory features until the desired count is obtained [34]. SFS performs the opposite, adding features by order of significance until the number of features is obtained. Both RFE and SFS have demonstrated a decent performance in improving model accuracy and efficiency in prior studies [32].

Classification accuracy on the test dataset was considered as the foremost important metric of algorithm performance [6, 12]. In addition, computational cost was defined as the time for grid search in each run. Lastly, experimental cost was selected as the third measure to evaluate the best subset of features, which included resources such as time, monetary cost, difficulty of modeling, and potential risks of having low quality data due to poorly designed experimental protocol (e.g., unexpected noise in eye-tracking data).

Four groups of input features were tested in this study including pupillometry data, device configuration, task performance, and CPM outcomes. Therefore, fourteen subsets of features were composed as shown in Table 1 (i.e., $\binom{4}{1} + \binom{4}{2} + \binom{4}{3}$). Since the purpose of this study was to explore the optimal subset of features, it excluded the set with all of the features.

III. Results

A. *Classification Performance*

The algorithm accuracy differences were statistically evaluated using Cochran's Q test ($\alpha = 0.05$) which is an extension of McNemar's test [35]. The Cochran's Q test was only performed under the K-Best feature selector because it was the fastest selector. There was no significant difference in accuracy between the RF and NB algorithms and between the RF vs. SVM algorithms ($p > .05$). However, SVM significantly outperformed NB ($p < .05$). In addition, RF and SVM outperformed random guessing ($p < .05$), while NB was not significantly different from random guessing ($p > .05$). It was also revealed that the accuracy of RF algorithm heavily relied on the primary task performance feature as subsets 1, 2, 4, 5, 7, 9, and 12 exhibited higher accuracy (>0.80) than the other subsets. Good models were defined as models that exhibited at least 70% accuracy [36].

B. *Computational Cost*

As shown in Table 1, the SVM and NB algorithms were faster than the RF in terms of training the model. Most cases under SVM and NB algorithms could be trained in near-real time (less than 1 sec.). Meanwhile, RF required several minutes to train, except for some cases with only one subset of features (Subset ID 11-14).

C. *Experimental Cost*

If a feature subset included pupillometry data, it was labeled as "high" in terms of experimental cost, as it requires conducting human subject experiments equipped with devices to collect bio signals which can be time consuming and costly [37]. If the feature subset only included CPM outcomes or device configuration, it was assumed to have "low"

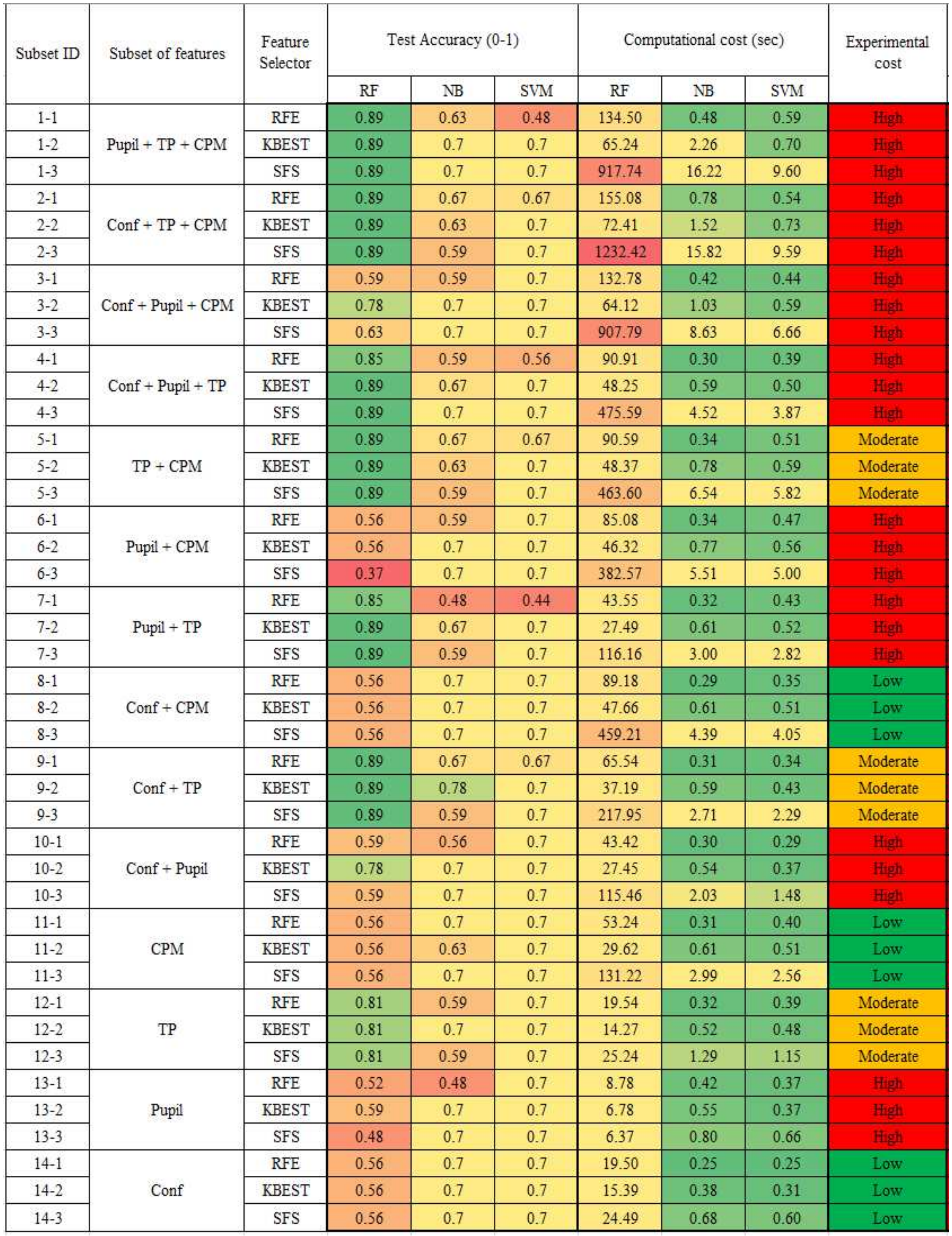
978-1-6654-5238-0/22/$31.00 ©2022 IEEE

experimental cost because the data can be generated from CPM without the need for conducting human subject experiments. If the feature included only task performance measures, the subset was labeled as having “moderate” experimental cost because it needs conducting a study with human participants but does not require physiological data collection.

IV. Discussion

This study developed ML algorithms and explored a subset of input features that can predict CW when using EMG-based prosthetic devices with acceptable accuracy and low computational and experimental cost. This was the first study that estimated CW of prosthetic devices with ML algorithms. There are several studies which measured CW of prosthetic

Table 1. Summary of machine learning results



Subset ID	Subset of features	Feature Selector	Test Accuracy (0-1)			Computational cost (sec)			Experimental cost
			RF	NB	SVM	RF	NB	SVM	
1-1	Pupil + TP + CPM	RFE	0.89	0.63	0.48	134.50	0.48	0.59	High
1-2		KBEST	0.89	0.7	0.7	65.24	2.26	0.70	High
1-3		SFS	0.89	0.7	0.7	917.74	16.22	9.60	High
2-1	Conf + TP + CPM	RFE	0.89	0.67	0.67	155.08	0.78	0.54	High
2-2		KBEST	0.89	0.63	0.7	72.41	1.52	0.73	High
2-3		SFS	0.89	0.59	0.7	1232.42	15.82	9.59	High
3-1	Conf + Pupil + CPM	RFE	0.59	0.59	0.7	132.78	0.42	0.44	High
3-2		KBEST	0.78	0.7	0.7	64.12	1.03	0.59	High
3-3		SFS	0.63	0.7	0.7	907.79	8.63	6.66	High
4-1	Conf + Pupil + TP	RFE	0.85	0.59	0.56	90.91	0.30	0.39	High
4-2		KBEST	0.89	0.67	0.7	48.25	0.59	0.50	High
4-3		SFS	0.89	0.7	0.7	475.59	4.52	3.87	High
5-1	TP + CPM	RFE	0.89	0.67	0.67	90.59	0.34	0.51	Moderate
5-2		KBEST	0.89	0.63	0.7	48.37	0.78	0.59	Moderate
5-3		SFS	0.89	0.59	0.7	463.60	6.54	5.82	Moderate
6-1	Pupil + CPM	RFE	0.56	0.59	0.7	85.08	0.34	0.47	High
6-2		KBEST	0.56	0.7	0.7	46.32	0.77	0.56	High
6-3		SFS	0.37	0.7	0.7	382.57	5.51	5.00	High
7-1	Pupil + TP	RFE	0.85	0.48	0.44	43.55	0.32	0.43	High
7-2		KBEST	0.89	0.67	0.7	27.49	0.61	0.52	High
7-3		SFS	0.89	0.59	0.7	116.16	3.00	2.82	High
8-1	Conf + CPM	RFE	0.56	0.7	0.7	89.18	0.29	0.35	Low
8-2		KBEST	0.56	0.7	0.7	47.66	0.61	0.51	Low
8-3		SFS	0.56	0.7	0.7	459.21	4.39	4.05	Low
9-1	Conf + TP	RFE	0.89	0.67	0.67	65.54	0.31	0.34	Moderate
9-2		KBEST	0.89	0.78	0.7	37.19	0.59	0.43	Moderate
9-3		SFS	0.89	0.59	0.7	217.95	2.71	2.29	Moderate
10-1	Conf + Pupil	RFE	0.59	0.56	0.7	43.42	0.30	0.29	High
10-2		KBEST	0.78	0.7	0.7	27.45	0.54	0.37	High
10-3		SFS	0.59	0.7	0.7	115.46	2.03	1.48	High
11-1	CPM	RFE	0.56	0.7	0.7	53.24	0.31	0.40	Low
11-2		KBEST	0.56	0.63	0.7	29.62	0.61	0.51	Low
11-3		SFS	0.56	0.7	0.7	131.22	2.99	2.56	Low
12-1	TP	RFE	0.81	0.59	0.7	19.54	0.32	0.39	Moderate
12-2		KBEST	0.81	0.7	0.7	14.27	0.52	0.48	Moderate
12-3		SFS	0.81	0.59	0.7	25.24	1.29	1.15	Moderate
13-1	Pupil	RFE	0.52	0.48	0.7	8.78	0.42	0.37	High
13-2		KBEST	0.59	0.7	0.7	6.78	0.55	0.37	High
13-3		SFS	0.48	0.7	0.7	6.37	0.80	0.66	High
14-1	Conf	RFE	0.56	0.7	0.7	19.50	0.25	0.25	Low
14-2		KBEST	0.56	0.7	0.7	15.39	0.38	0.31	Low
14-3		SFS	0.56	0.7	0.7	24.49	0.68	0.60	Low

*Note: Pupil features include: pupil diameter and blink rate. Task performance (TP) features include: number of moved pins from the experiment, shortest time to move one pin, and number of training trials. Device configuration (Conf) features include: DC, PR, and CC. Cognitive performance model (CPM) features include: estimated time to move one pin, number of cognitive, perceptual, or motor operators, and number of memory chunks

978-1-6654-5238-0/22/$31.00 ©2022 IEEE

devices, however, they employed human-subject experiments to collect physiological data, task performance, or subjective assessment and relied on inferential statistics and not ML [14]. It was found that the SVM algorithm can be used to classify CW in near-real time with a reasonable classification accuracy and therefore, can be used for predicting the workload of using EMG-based prosthetic devices early in their design and development process possibly further alleviating amputee challenges in performing ADLs.

This study was the first to include CPM outcomes as input features in the ML algorithm. There are several advantages of using CPM over human subject experiments. The CPM does not require extensive human-subject experimentation. Furthermore, it can quantify and estimate human behavior in simple tasks with tools such as Cogulator [38] or CogTool [39] based on well-established theories (e.g., human information processing). Combining CPM outcomes with other features captured from human subject experiment can increase prediction accuracy. Furthermore, combination of CPM outcomes and device configurations was found as a good candidate for near-real time classification of CW for designing upper-limb prostheses, which can minimize the demand of having human-subject experiments and modeling efforts. Although subsets 11 and 14 in Table 1 have the same test accuracy as subset 8, we recommend subset 8 because this subset of features would have more stabilized classification results with larger datasets. That is, using only one type of feature, either device configuration or CPM outcomes can increase the risk of having low classification accuracy with larger datasets [12].

Another contribution of this research was inclusion of experimental cost to select the best subset of features, unlike other ML studies that only focused on computational cost. This is an important consideration as human subject studies are challenging due to the difficulties of collecting high quality data from participants, especially when physiological data are collected (due to the devices being intrusive) [40, 17].

Regarding ML algorithms, we found that while RF generated a large number of trees in its training process, the model performance heavily depended on task performance features as compared to other features. This was shown with high test accuracy of RF especially for subsets 1, 2, 4, 5, 7, 9, and 12. In RF, task performance was always selected and regarded as the most important feature in all subsets. This can be problematic as it can be an indicator of overfitting, although we tried to minimize the effect of small dataset with hyperparameter tuning and feature selection. Furthermore, RF required substantial computational cost to train the model as it builds many trees to combine the decision trees to determine a class. In addition, hyperparameter setting in RF was more complex than that of SVM and NB, which extended the model running time.

SVM algorithm exhibited more stabilized classification performance as compared to RF, even with the extremely small dataset in this study. Throughout all subsets of features (subset 1-14), SVM showed around 70% accuracy. It is also known that low-complexity models such as SVM generate the best results because it does not put much emphasis on a small outstanding characteristic of the dataset, which can exhibit low risk of overfitting [41]. In addition, a simple hyperparameter tuning enabled short grid search time. Regarding the feature selector, both RFE and K-Best worked well for SVM to train the model within one second. However, SFS demanded extensive computational time because it is a wrapper method which needs to add each feature to the model and train the classifier for each feature subset.

This study had some limitations. First, the experiment was conducted with able-bodied participants due to the limited number of amputee patients in the area [14]. The result can change if participants were experienced users of prosthetic devices. Future studies should validate the model outcomes with amputee patients. Future research should also include additional data to avoid the risk of overfitting. In addition, additional hyperparameter tuning method (e.g., random search) or parameters should be considered to improve the classification accuracy in SVM and NB, which currently could not go beyond 70%.

ACKNOWLEDGMENT

The funding for this study was provided by the National Science Foundation (No. IIS-1856676/1856441/1900044).

REFERENCES

[1] 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*, 2008. **89**(3): p. 422-429.

[2] M. M. Lusardi, M. Jorge, and C. C. Nielsen, Orthotics and Prosthetics in Rehabilitation-E-Book. 2013: Elsevier Health Sciences.

[3] A. Kannenberg and B. Zacharias. Difficulty of performing activities of daily living with the Michelangelo Multigrip and traditional myoelectric hands. in *American Academy of Orthotists & Prosthetists 40th Academy Annual Meeting &Scientific Symposium*, FPTH14. 2014.

[4] C. Gaskins, K. Kontson, E. P. Shaw, I. M. Shuggi, M. J. Ayoub, J. C. Rietschel, M. W. Miller, and R. Gentili, Mental Workload Assessment During Simulated Upper Extremity Prosthetic Performance. Archives of Physical Medicine and Rehabilitation, 2018. **99**(10): p. e33.

[5] K. Moustafa, S. Luz, and L. Longo. Assessment of mental workload: A comparison of machine learning methods and subjective assessment techniques. in 1st International Symposium on Human Mental Workload: Models and Applications, H-WORKLOAD 2017, June 28, 2017 - June 30, 2017. 2017. Dublin, Ireland: Springer Verlag.

[6] P. O. Braarud, T. Bodal, J. E. Hulsund, M. N. Louka, C. Nihlwing, E. Nystad, H. Svengren, and E. Wingstedt, An Investigation of Speech Features, Plant System Alarms, and Operator-System Interaction for the Classification of Operator Cognitive Workload During Dynamic Work. Human Factors, 2021. **63**(5): p. 736-756.

[7] D. Meyer, Support Vector Machines: The Interface to libsvm in package e1071, R package version 1.6-8. URL https://cran. r-project. org/web/packages/e1071/vignettes/svmdoc. pdf, 2017.

[8] A. Liaw and M. Wiener, Classification and regression by randomforest. R News 2 (3): 18–22. URL: http://CRAN. R-project. org/doc/Rnews, 2002.

[9] M. Majka, naivebayes: High performance implementation of the Naive Bayes algorithm. R package version 0.9. 2. 2018.

[10] Q. Meteier, M. Capallera, S. Ruffieux, L. Angelini, O. Abou Khaled, E. Mugellini, M. Widmer, and A. Sonderegger, Classification of Drivers' Workload Using Physiological Signals in Conditional Automation. Frontiers in Psychology, 2021. **12**: p. 18.

[11] R. Walambe, P. Nayak, A. Bhardwaj, and K. Kotecha, Employing Multimodal Machine Learning for Stress Detection. Journal of Healthcare Engineering, 2021. **2021**: p. 12.

[12] Y. Ding, Y. Q. Cao, V. G. Duffy, Y. Wang, and X. F. Zhang, Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. Ergonomics, 2020. **63**(7): p. 896-908.

[13] J. Li, H. Li, W. Umer, H. W. Wang, X. J. Xing, S. K. Zhao, and J. Hou, Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. Automation in Construction, 2020. **109**: p. 15.

[14] J. Park and M. Zahabi, Cognitive Workload Assessment of Prosthetic Devices: A Review of Literature and Meta-Analysis. IEEE Transactions on Human-Machine Systems, 2022.

[15] L. Resnik, H. H. Huang, A. Winslow, D. L. Crouch, F. Zhang, and N. Wolk, Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control. Journal of neuroengineering and rehabilitation, 2018. **15**(1): p. 23.

[16] M. M. White, W. Zhang, A. T. Winslow, M. Zahabi, F. Zhang, H. Huang, and D. B. Kaber, Usability comparison of conventional direct control versus pattern recognition control of transradial prostheses. IEEE Transactions on Human-Machine Systems, 2017. **47**(6): p. 1146-1157.

[17] Butt, A. H., Rovini, E., Dolciotti, C., De Petris, G., Bongioanni, P., Carboncini, M., & Cavallo, F. (2018). Objective and automatic classification of Parkinson disease with Leap Motion controller. BioMedical Engineering Online, 17(1), 1-21.

[18] Dyshel, M., Arkadir, D., Bergman, H., & Weinshall, D. (2015). Quantifying levodopa-induced dyskinesia using depth camera. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision Workshops.

[19] M. Zahabi, M. M. White, W. Zhang, A. T. Winslow, F. Zhang, H. Huang, and D. B. Kaber, Application of Cognitive Task Performance Modeling for Assessing Usability of Transradial Prostheses. IEEE Transactions on Human-Machine Systems, 2019. **49**(4): p. 381-387.

[20] J. Park, M. Zahabi, D. Kaber, J. Ruiz, and H. Huang. Evaluation of Activities of Daily Living Tesbeds for Assessing Prosthetic Device Usability. in 2020 IEEE International Conference on Human-Machine Systems (ICHMS). 2020. IEEE.

[21] B. E. John. Extensions of GOMS Analyses to Expert Performance Requiring Perception of Dynamic Visual and Auditory. in Empowering People: CHI'90 Conference Proceedings [on] Human Factors in Computing Systems: Seatle, Washington, April 1-5, 1990. 1990. Addison-Wesley.

[22] S. Estes, Cogulator. The MITRE Corporation, 2017.

[23] S. Deeny, C. Chicoine, L. Hargrove, T. Parrish, and A. Jayaraman, A Simple ERP Method for Quantitative Analysis of Cognitive Workload in Myoelectric Prosthesis Control and Human-Machine Interaction. Plos One, 2014. **9**(11).

[24] M. Markovic, M. A. Schweisfurth, L. F. Engels, T. Bentz, D. Wüstefeld, D. Farina, and S. Dosen, The clinical relevance of advanced artificial feedback in the control of a multi-functional myoelectric prosthesis. Journal of neuroengineering and rehabilitation, 2018. **15**(1): p. 28.

[25] R. C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 1971. **9**(1): p. 97-113.

[26] J. Tiffin and E. J. Asher, The Purdue Pegboard: norms and studies of reliability and validity. Journal of applied psychology, 1948. **32**(3): p. 234.

[27] M. Kaczorowska, M. Plechawska-Wojcik, and M. Tokovarov, Interpretable Machine Learning Models for Three-Way Classification of Cognitive Workload Levels for Eye-Tracking Features. Brain Sciences, 2021. **11**(2): p. 22.

[28] T. Götze, M. Gürtler, and E. Witowski, Improving CAT bond pricing models via machine learning. Journal of Asset Management, 2020. **21**(5): p. 428-446.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 2011. **12**: p. 2825-2830.

[30] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction. Vol. 2. 2009: Springer.

[31] C. C. Aggarwal, Machine learning for text. Vol. 848. 2018: Springer.

[32] A. J. Ferreira and M. A. Figueiredo, Efficient feature selection filters for high-dimensional data. Pattern recognition letters, 2012. **33**(13): p. 1794-1804.

[33] M. Raihan-Al-Masud and M. R. H. Mondal, Data-driven diagnosis of spinal abnormalities using feature selection and machine learning algorithms. Plos one, 2020. **15**(2): p. e0228422.

[34] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines. Machine learning, 2002. **46**(1): p. 389-422.

[35] Mangiafico, S. S. (2016). Summary and analysis of extension program evaluation in R, version 1.18. 1. New Brunswick: Rutgers Cooperative Extension.

[36] K. Barkved, How To Know if Your Machine Learning Model Has Good Performance. 2022.

[37] Lee, K., Caverlee, J., & Webb, S. (2010). Uncovering social spammers: social honeypots+ machine learning. Paper presented at the Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.

[38] Estes, S. (2017). Cogulator. The MITRE Corporation.

[39] John, B. E., & Suzuki, S. (2009). Toward cognitive modeling for predicting usability. Paper presented at the 13th International Conference on Human-Computer Interaction, HCI International 2009, July 19, 2009 - July 24, 2009, San Diego, CA, United states.

[40] J. Park and M. Zahabi. Comparison of Cognitive Workload Assessment Techniques in EMG-based Prosthetic Device Studies. in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2020. IEEE.

[41] A. Shenoy, Text Classification with Extremely Small Datasets. 2019.

978-1-6654-5238-0/22/$31.00 ©2022 IEEE