

Towards a Real-Time Cognitive Load Assessment System for Industrial Human-Robot Cooperation

Akilesh Rajavenkatanarayanan^{1§}, Harish Ram Nambiappan^{1§}, Maria Kyrarini^{1§}, and Fillia Makedon¹

Abstract—Robots are increasingly present in environments shared with humans. Robots can cooperate with their human teammates to achieve common goals and complete tasks. This paper focuses on developing a real-time framework that assesses the cognitive load of a human while cooperating with a robot to complete a collaborative assembly task. The framework uses multi-modal sensory data from Electrocardiography (ECG) and Electrodermal Activity (EDA) sensors, extracts novel features from the data, and utilizes machine learning methodologies to detect high or low cognitive load. The developed framework was evaluated on a collaborative assembly scenario with a user study. The results show that the framework is able to reliably recognize high cognitive load and it is a first step in enabling robots to understand better about their human teammates.

I. INTRODUCTION

Robots have become part of our everyday lives and have several roles such as helping workers in their work duties [1], [2], [3], assisting people with impairment in activities of daily living [4], [5], entertaining and keeping company to children and the elderly [6], [7], and assisting rehabilitation procedures [8], [9]. In industrial environments, such as assembly lines, a strong level of interaction and cooperation is reached where humans and robots are required to work synergistically on a specific task and have different roles and complementary abilities. Researchers focus on achieving safe Human-Robot Cooperation (HRC), which will not threaten or harm the physical health of the human teammate [10], [11]. However, there is limited research on understanding the psychological influence on the human who cooperates with a robot on a daily basis. To ensure the efficiency and productivity of the overall HRC, the human teammate needs to feel comfortable while working with a robot. Therefore, it is essential for robots to understand how their human teammates feel and adapt accordingly. The difficulty of the task at hand, time restrictions, and arousal are shown to be some of the factors that affect cognitive load [12]. Cognitive load is the amount of energy a person has to spend to achieve a specific goal. This type of cognitive load is called an extraneous cognitive load. Prior research [13] indicate that different types of biosignals from physiological sensors can

help monitor the human partner while working, for changes in performance due to stress, lack of sleep, or increase in cognitive load.

In our previous work [14], we proposed a cognitive performance assessment framework for HRC based on sleep data collected from a commercially available wrist band. Inspired by our previous work, the main focus of this paper is on the development of a real-time HRC framework, called *RoboAssist*, that enables robots to monitor the cognitive load of their human teammates using multimodal sensor data. The *RoboAssist* system consists of commonly used minimal-invasive sensors, such as Electrocardiography (ECG) and Electrodermal Activity (EDA) [13], [14]. The main contributions of the presented work are:

- the research and development of a machine learning based framework that assesses cognitive load during HRC;
- a comprehensive list of features that are extracted from physiological sensors, such as ECG and EDA.

To demonstrate the feasibility of *RoboAssist*, a study was conducted with 25 participants. The study was performed in a collaborative assembly scenario using the collaborative robot Sawyer developed by Rethink Robotics [15] (Fig. 1).

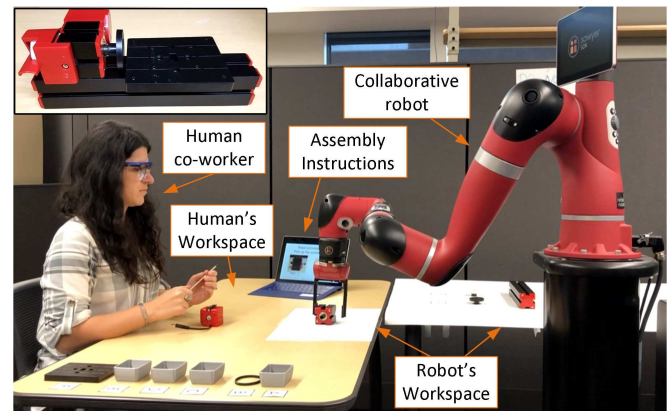


Fig. 1: Setup of the Presented Collaborative Assembly Scenario. Top-left image: Final Assembly Product - A Miniature Sanding Machine

The paper is organized as follows. In section II, we present related work, section III describes *RoboAssist* system for HRC, and section IV presents the experimental results of the study and evaluation. In the final section, we discuss our future research direction and conclude.

This work was supported in part by the National Science Foundation under award number 1719031.

¹Akilesh Rajavenkatanarayanan, Harish Ram Nambiappan, Maria Kyrarini, and Fillia Makedon are with the Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington 76019, Texas. akilesh.rajavenkatanarayanan@mavs.uta.edu, harishram.nambiappan@mavs.uta.edu, maria.kyrarini@uta.edu, makedon@uta.edu

[§]Akilesh Rajavenkatanarayanan, Harish Ram Nambiappan, and Maria Kyrarini equally contributed to this work.

II. RELATED WORK

Nelles et al. [16] summarized the available research work on evaluation metrics for human well-being in Human-Robot Interaction (HRI) and highlights that the experimental design, the questionnaires, and measures used are very heterogeneous. Well-being is the state of feeling comfortable, healthy, happy, and it is connected with the ability of an individual to manage stress. Most of the research presented by Nelles et al. focuses on self-reporting surveys and questionnaires to provide feedback for the design of HRI regarding trust [17], [18], [19], usability [20], [21], feeling of safety [22], cognitive and physical workload [23], and well-being [24]. However, self-reporting is a subjective measure of cognitive workload and it is difficult to quantify for real-time applications.

Advances in sensor technology in recent years have the potential to help with the assessment of cognitive load. Weistroffer et al. [25] used pre-post questionnaires and pre-post measures of heart rate from a Photoplethysmogram sensor and skin conductivity from an EDA sensor during HRC. Their goal was to compare the acceptability of collaborative assembly between actual situations and their virtual reality counterparts combined through questionnaires. However, the authors were not able to record the physiological measurements during the actual HRC due to the positioning of the sensors. Novak et al. [26] identified physical and cognitive load during haptic interaction with a robot by using four measurements; heart rate measured by ECG, skin conductivity by EDA sensor, respiratory rate by the thermistor flow sensor, and peripheral skin temperature by a digital temperature sensor. Results from a user study of 30 participants show the combination of respiration and skin temperature appears to estimate cognitive workload in physically demanding interaction with haptic robots. However, the level of discomfort of wearing a thermistor flow sensor attached to the user's nose was not considered.

A framework to recognize user's cognitive load during HRI was presented by Villani et al. [27]. The framework analyzes the variability of the user's heart rate, which is measured by a smartwatch. The selected scenario was the interaction with a wheeled robot, and the user was providing commands to the robot via hand gestures. The authors used the analysis of 1000 Monte Carlo trials on segments of the heart rate signal to detect the rest and stress conditions. Experimental results showed that the algorithm could detect changes in mental workload. Moreover, The framework was tested by Landi et al. [28] in a teleoperation robotic task, where virtual fixtures are used as an assistive technology. However, in both works, there is no evaluation of the accuracy of the cognitive load recognition. Moreover, other characteristics of heart rate could provide valuable insights into the person's mental state.

The assessment of cognitive load during HRC is beneficial for the improvement of working conditions and human well-being. HRC requires the sharing of objects and/or environments between the human and the robot that may induce

stress and anxiety, and thereby increasing the cognitive load. Therefore, our primary focus is on developing a framework for collaborative robots, which are increasingly used in industry [29]. This work presents a multi-sensory minimal-invasive HRC system that analyzes the physiological state of the user and recognizes changes in mental state, more specifically, cognitive load. The framework is based on machine learning techniques, and differs from the previous presented work. Moreover, for the evaluation of the proposed framework, a collaborative assembly scenario was proposed, which is similar to real-world scenarios in the industry.

III. COGNITIVE LOAD ASSESSMENT FOR HRC

An overview of the proposed *RoboAssist* system is shown in Fig 2. The system assesses human cognitive load during HRC. In this section, the selected collaborative assembly task, the placement of the multi-modal sensory system, feature extraction, and machine learning approaches are explained in detail.

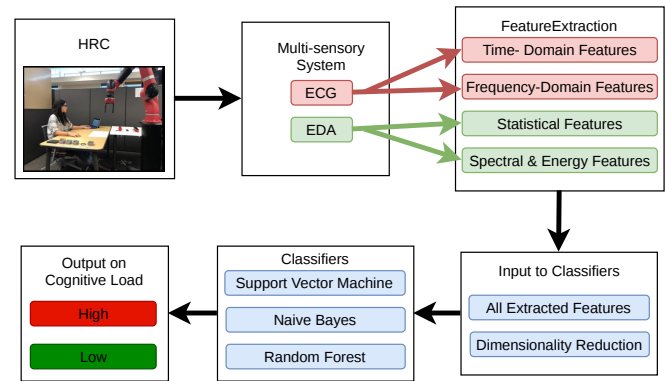


Fig. 2: Overview of the proposed *RoboAssist* system for HRC

A. Collaborative Assembly Task

For this assessment, we focused on the task of assembling a small sanding machine, shown in Fig 1, in an industrial HRC scenario. We divided the parts required to make the sanding machine into two sets; a set of components that a robot can handle and a set of pieces that only a human can handle.

As shown in Fig 1, the parts to the robot's right side are the parts that are handled by the robot. The robot picks it up and brings it towards the user while the user assembles the machine. Small pieces, which can only be handled by the user, such as nuts and bolts, were placed in separate small boxes called "bins" to the right side of the user along with a screwdriver. This setup is shown in Fig 1. A user interface was created for the assembly task, which showed the step-by-step instructions required to assemble the sanding machine. Each step provided the user with detailed instructions of the parts s/he needed to handle and information about the parts the robot would handle. The user interface is integrated with the robot using the Robot Operating System (ROS) [30] so that the robot provides the required components automatically to the user. A video

of the assembly task can be found in the following link: https://youtu.be/m_dkLHf1CUo.

B. Multi-Sensor System and Sensor Placement

The *RoboAssist* system consists of an ECG and an EDA sensor developed by BioSignalsplux [31]. The data are transmitted via Bluetooth at 1000 Hz. ECG data were collected from a standard 3-point bipolar standard limb leads configuration of the Einthoven's triangle [32]. We used a Lead II setup in this configuration, where a positive electrode is on the left leg, a negative electrode on the right arm, and a reference electrode on the right leg for recording purposes. The electrodes were placed on the participant's right shoulder and lower torso for this study to make wearing the sensor and recording easier. The EDA sensor measures the electrical potential produced on the skin surface due to the activity of the sweat glands [33]. The best locations to acquire such signals are spots where sweat glands are most active, like the palms and the soles [34]. Since our application involves participants using their arms for the collaborative assembly task, EDA was collected from the shoulder, which was proven to be one of the best alternative locations for skin conductance measurement [35]. The selected placements for ECG and EDA enable the sensors to be integrated into an easily wearable smart T-shirt.

C. Feature Extraction

In this section, the data processing and feature extraction for the ECG and EDA sensor data will be discussed.

1) *ECG Signal*: Because of the position of the ECG electrodes on the shoulder and lower torso, the ECG signal is inverted when compared to the standard Lead II setup. Hence, the signal is inverted to represent the original waveform and then is pre-processed for feature extraction. For this study, we extracted time and frequency domain features from the QRS complex as shown in Fig 3 and RR interval (the time elapsed between two successive R-waves) of the ECG signal, which are commonly used features in heart rate variability analysis for mental stress detection [36]. The QRS complex forms the major component of the ECG signal and it represents the electrical activation in the sensor due to the ventricles contracting in the heart. It is the main spike seen on the ECG waveform and is used for computing the heart rate and several heart-disease states [37], [38]. We used the peak detection algorithm developed by Van Gent et al. [39]. To improve the peak detection, a notch filter is applied before the peak detection algorithm with a cut-off frequency of a threshold value, empirically selected as 0.05 Hz to minimize the T-wave and other unwanted low-frequency noise. Subsequently, the time and frequency-domain features are extracted. The time domain features that are proposed by Boonnithi et al. in [36] are as follows; the mean RR interval or mean Inter-beat Interval (mRR), the mean heart rate (mHR), the standard deviation of RR interval (SDRR), the standard deviation of heart rate (SDHR), the coefficient of variance of RR intervals (CVRR), the root mean square successive difference (RMSSD), the proportion of successive

differences above 20 ms in percentage (pRR20), and the proportion of successive differences above 50 ms in percentage (pRR50). Moreover, additional time domain features were extracted as follows; the median RR Interval ($\overline{RR}$), the range of the RR Interval (rRR), and median absolute deviation of RR intervals (MAD). Table IV summarizes the formulas of all the time domain extracted features.

The frequency domain features extracted from ECG are the low and high frequency, the Sympathetic modulation index, the Vagal modulation index and the Sympathovagal balance index and Table I summarizes their formulas.

TABLE I: Frequency Domain Feature Extraction from ECG Data

Features	Computation
Low Frequency (LF)	LF = Power spectrum from 0.04 to 0.15 Hz
High Frequency (HF)	HF = Power spectrum from 0.15 to 0.5 Hz
Sympathetic modulation index (SMI)	$SMI = LF / (LF + HF)$
Vagal modulation index (VMI)	$VMI = HF / (LF + HF)$
Sympathovagal balance index (SVI)	$SVI = LF / HF$

2) *EDA Signal*: The collected EDA data are downsampled to 200Hz to reduce computation and then filtered using a Butterworth filter to remove high-frequency noise using methodology proposed by Bizzego et al. [40]. Research indicates that EDA signals comprise of two different superimposed components; the phasic or the skin conductance response (SCR) and the tonic or the skin conductance levels (SCL) [34], [41]. The phasic component varies based on the provided stimulus, where changes in the signal imply activation of the sudomotor nerve due to activity in the sweat glands. Whereas the tonic component is the baseline level of skin conductance, which varies from person to person [41]. In this paper, the phasic component is used to extract features as we were interested in modeling the user's cognitive load to the presented task.

The shape of the EDA signal is important in signifying a change in nervous response. Statistical features related to the amplitude, the first derivative and the second derivative of the signal were extracted. Additional spectral and energy features that are commonly used to describe the characteristics of one-dimensional (1D) signals were extracted. The following features were extracted from the SCR signal; Mean Value, Standard Deviation, Maximum Value, Minimum Value, Range, Variance, first Derivative Mean, first Derivative Standard Deviation, second Derivative Mean, second Derivative Standard Deviation, Zero Crossing Rate (the rate at which the signal changes sign in a given window), Spectral Centroid, Spectral Rolloff, Spectral Entropy, Energy, and Entropy of Energy.

The spectral centroid of the given frame of the spectrum is computed by the following equation:

$$C = \sum_{i=0}^{N-1} X_i p(X_i),$$

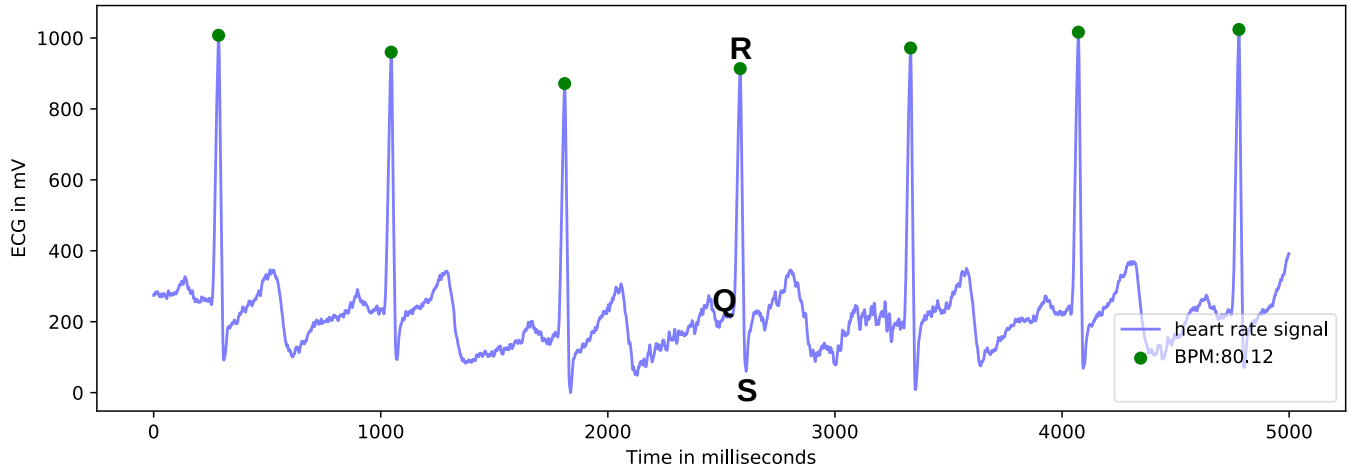


Fig. 3: Sample ECG signal acquired from the Biosignalsplux sensor. The green dots indicate the peak detected using which the heart rate of the signal was estimated. Q, R, S indicate the Q-wave, R-wave and S-wave component of the ECG signal.

where N is the size of the spectrum, X is the observed frequencies and $p(X)$ is the probability to observe a specific value in X .

Spectral Rolloff corresponds to the frequency below which 90% of the magnitude distribution of the spectrum is concentrated. It is given by the equation:

$$R = 0.9 \sum_{i=0}^{N-1} |X_i|,$$

where X is the spectrum of the signal and N is the size of the positive spectrum.

Spectral Entropy is the entropy of the normalized spectral energy of the given signal and is computed by the formula:

$$S_E = - \sum_{f=0}^{f_s/2} P(f) \log_2 P(f),$$

where f_s is the sampling frequency and P is the normalized power spectral density.

Energy is the sum of squares of the signal divided by the length of the frame and it is calculated by the formula:

$$E = \frac{1}{N} \sum_{l=0}^{N-1} |X_l|^2,$$

where N is the length of the signal window and X is the observed frequencies.

The entropy of energy of the given signal is given by the formula

$$E_E = - \sum E \log_2 E,$$

where E is the energy of the signal given a window. These spectral and energy features have also been for other signals like EEG [42], and speech [43]. The total extracted features from ECG and EDA data are 17 and 16, respectively.

D. Machine Learning Approach

Machine Learning analysis was performed using the features extracted from each ECG and EDA data and a combination of these data. A total of 3 combinations of modalities were tested for classification performance. Our main goal is to prove that the data collected from the proposed *RoboAssist* framework is enough to detect the cognitive load of the participant given the HRC task. Three widely used classifiers were selected; Support Vector Machine (SVM) and Naive Bayes (NB), the two commonly used algorithms in Human-Computer Interaction [44]. We also chose the Random Forest (RF) algorithm to compare the performance using an ensemble algorithm. These algorithms are then evaluated using accuracy and F1-scores. Accuracy is defined as a measure of the total number of correctly identified cases and F1-score is a measure of the harmonic mean of the precision and recall [45]. F1-score gives a better understanding of the misclassified cases as it is critical in the design of the framework. In the next section, the evaluation results of the classification models are presented and discussed.

IV. EXPERIMENTAL RESULTS

A. User study

Twenty five participants from The University of Texas at Arlington (UTA), participated in the user study where fifteen were male and ten were female participants. Among these participants, 23 were right-handed participants and two were left-handed participants. Six of the participants had prior experience in HRC. Except for one participant, who was in the 31 to 40 age range, all other participants were in the 19 to 30 age range. Each user study lasted for about 30 to 40 minutes and it was approved by the Institutional Review Board at UTA.

The participants completed a baseline survey before starting the user study. Subsequently, they performed the first session, which is performing the assembly task. On completion, they were asked to fill the first post-task survey form. Next,

they performed the second session which is performing the same assembly task with time constraints to induce stress and high cognitive load. The participants were required to complete each step of the assembly task within 30 seconds. After the second session is completed the participants filled out the second post-task survey form. In each and every session, the time taken by the user in each and every step of the task was recorded along with the sensor data, which is used for further evaluation along with the survey responses. To avoid the cumulative effect of stress, there was a short 3 minute break between the two sessions for the participants to relax.

In real-world assembly and production lines, time is a very critical parameter as factories are required to increase productivity and to reduce production time [46], [47]. Traditionally, human co-workers receive training before they start working with a robot daily. Hence, practice effect is always present in the real world and our framework reflects this design. Since our goal is to simulate a real-world setup, the participants were required to undergo a training session (first session) before they were allowed to work on a timed session (second session).

User Feedback: Fig 4 summarizes the responses of the participants on some of the survey questions. The participants were asked to rate their level of sleepiness, stress, and attention from 0, meaning very low to 10, meaning very high. Fig 4 shows that the users felt slightly more stressed during the second session (with timer) in comparison to the first session (no timer) and the baseline. Similarly, the users felt less sleepy during the second session in comparison to the baseline and the first session indicating that the participants were more attentive as the stress levels increased. However, the users felt that their attention was almost similar for both sessions and higher than the baseline. The survey also shows that continuous work can lead to an increase of mental exertion, and this may lead to stress and increased cognitive load. Thus, it is important in an HRC scenario to be able to monitor the user's cognitive load and adapt accordingly, especially in an industry. It is also important to note that the values obtained from the user surveys were not significantly different between the two tasks and this can be attributed to the relatively small number of participants.

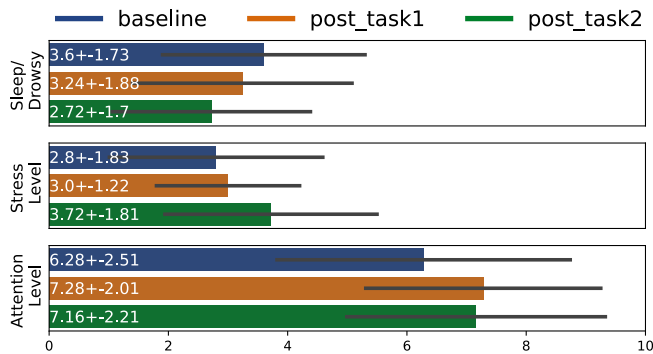


Fig. 4: Subjective feedback

B. Evaluation of the Proposed Framework

Data from three participants were rejected due to sensor malfunction. The data from the remaining 22 participants were organized into two classes; the first session was the class "low cognitive load" while the second session was the class "high cognitive load". Hence in total, we had 44 data-points with 22 in one class and 22 in the other. Sensors are inherently noisy. Hence, the collected sensor data were preprocessed to remove noise. Owing to the small size of the dataset, the research team manually verified the data to see if the preprocessing steps discussed in section III-C was able to reduce noise and did not disturb the important characteristics of the signal.

As discussed in section III-C, using the proposed *RoboAssist* framework, a total of 17 ECG and 16 EDA features were extracted from the data. Three machine learning algorithms were utilized to classify the participant's stress state using both a multimodal and a unimodal approach. The following data combinations were used for classification; only ECG, only EDA, and both ECG and EDA data. Since our main goal is to build a real-time cognitive load assessment system, it is important to have a lower computational load on the system. Hence, we first perform preliminary analysis on the entire dataset after which we performed PCA to reduce the dimensionality of the dataset to avoid overfitting and thereby reduce computation.

The data from 16 participants were considered the training set and data from 6 participants were the testing set. For modeling the signals, we performed an exhaustive grid search in order to obtain the best machine learning model for evaluation. Tables II and III present the accuracy and F1-scores on the test set of the SVM (linear kernel), RF, and NB algorithms trained with all the features and with PCA, respectively. The number of PCA components, denoted by C, was selected empirically. The C value ranged between 3 and 15 and for each C value, the accuracy and F1-score were calculated for each ML algorithm. The minimum C value that provided the best accuracy and F1-score was selected and presented in Table III. The minimum number of required PCA components enables the system to work faster, which is crucial for real-time applications. Moreover, Chabathula et al. [48] have shown that the number of PCA components influences the accuracy of ML methods differently. Therefore, the number of PCA components that provide the best accuracy and F1-score is different for SVM, RF, and NB algorithms. The best result for each algorithm is highlighted in green. As shown in table II, SVM gives the best accuracy of 92.85 % and F1-Score of 0.941 when all the features of ECG and EDA are used. Moreover, the SVM gives similar results (accuracy of 92.85 % and F1-Score of 0.933) with the PCA applied to ECG and EDA features. The same accuracy and F1-score can also be achieved by using only the ECG features with PCA and RF. Hence, the use of only ECG could provide similar results with the combination of ECG and EDA. As for NB, the results in Table II show very low accuracy whereas it is slightly improved on application of

PCA as show in Table III. These results could be explained by the fact that our dataset consists only 32 data points in the training set but contains 33 features. This may have resulted in overfitting and there by these results. Table III also shows that ECG performs better when compared to EDA in NB. We think that this is because of the generic features extracted for EDA which may have been redundant and thus creating a highly correlated features. In the future, we plan a larger data collection, including industrial workers, that could help to improve the accuracy of the system.

TABLE II: Machine learning evaluation results for all extracted features.

	SVM		RF		NB	
	Acc	F1	Acc	F1	Acc	F1
ECG	42.85	0.333	71.42	0.714	57.14	0.667
EDA	71.42	0.714	78.57	0.769	57.14	0.400
ECG+EDA	92.85	0.941	78.57	0.800	57.14	0.400

TABLE III: Machine learning evaluation results with PCA. C is the number of PCA components.

	SVM			RF			NB		
	C	Acc	F1	C	Acc	F1	C	Acc	F1
ECG	4	57.14	0.667	10	92.85	0.933	10	78.57	0.800
EDA	10	78.57	0.833	4	64.28	0.667	5	57.14	0.500
ECG+EDA	15	92.85	0.933	5	64.28	0.706	15	78.57	0.823

V. CONCLUSION

In this paper, we presented *RoboAssist*, an HRC framework that would enable the robot to assess the user's cognitive load. We believe such a framework is vital to ensure a robotic system with a heart and soul. The classification results of the machine learning algorithms indicate that the data collected using this framework is valid, and the framework can be applied not only to industrial HRC scenarios but also to other HRC domains, such as assistive and service robotics, rehabilitation robotics, and others. The results of the user survey provided us with a piece of valuable information that the users felt an increase in cognitive load exertion, which is indicated by the high values of their response to the attention level question. The answer to the question of sleepiness also suggests that as stress increases, they felt less sleepy. It is important to note that we focused only on cognitive load detection in this paper. As indicated in our prior work [14], there are several factors that can affect the user's performance. Further research is necessary to expand this framework to include several other factors that may affect performance. One of the critical questions we are tackling right now is that "*Do the extracted features represent the data properly?*" Even though the features extracted for ECG are specific to ECG signals, the features extracted for the EDA signal are generic and are used in most cases of 1D signal, such as speech and EEG. Further research is needed to identify significant features or create an automated feature extraction step before classification. Some research in this domain is already underway but progress needs to be made

for a real-time system to ensure safe HRC and improving the well-being of the human partner.

ACKNOWLEDGMENT

The authors would like to thank all the participants of the study. This paper is based upon work supported by the National Science Foundation under Grant No 1719031. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] L. Rozo, S. Calinon, and D. G. Caldwell, "Learning force and position constraints in human-robot cooperative transportation," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 619–624.
- [2] M. Kyrarini, M. A. Haseeb, D. Ristić-Durrant, and A. Gräser, "Robot learning of industrial assembly task via human demonstrations," *Autonomous Robots*, vol. 43, no. 1, pp. 239–257, 2019.
- [3] L. Peternel, N. Tsagarakis, and A. Ajoudani, "A method for robot motor fatigue management in physical interaction and human-robot collaboration tasks," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 2850–2856.
- [4] F. F. Goldau, T. K. Shastha, M. Kyrarini, and A. Gräser, "Autonomous multi-sensory robotic assistant for a drinking task," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 210–216.
- [5] M. Kyrarini, Q. Zheng, M. A. Haseeb, and A. Gräser, "Robot learning of assistive manipulation tasks by demonstration via head gesture-based interface," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 1139–1146.
- [6] D. Silvera-Tawil and C. Roberts-Yates, "Socially-assistive robots to enhance learning for secondary students with intellectual disabilities and autism," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 838–843.
- [7] L. P. E. Toh, A. Causo, P.-W. Tzuo, I.-M. Chen, and S. H. Yeo, "A review on the use of robots in education and young children," *Journal of Educational Technology & Society*, vol. 19, no. 2, pp. 148–163, 2016.
- [8] A. Lioulemes, M. Theofanidis, and F. Makedon, "Quantitative analysis of the human upper-limb kinematic model for robot-based rehabilitation applications," in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2016, pp. 1061–1066.
- [9] W.-Y. G. Louie, S. Mohamed, and G. Nejat, "Human-robot interaction for rehabilitation robots," in *Robotic Assistive Technologies*. CRC Press, 2017, pp. 25–70.
- [10] P. A. Lasota, T. Fong, J. A. Shah, *et al.*, "A survey of methods for safe human-robot interaction," *Foundations and Trends® in Robotics*, vol. 5, no. 4, pp. 261–349, 2017.
- [11] S. Robla-Gómez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, "Working together: A review on safe human-robot collaboration in industrial environments," *IEEE Access*, vol. 5, pp. 26 754–26 773, 2017.
- [12] E. Galy, M. Cariou, and C. Mélan, "What is the relationship between mental workload factors and cognitive load types?" *International Journal of Psychophysiology*, vol. 83, no. 3, pp. 269–275, 2012.
- [13] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, 2019.
- [14] A. Rajavenkatanarayanan, V. Kanal, M. Kyrarini, and F. Makedon, "Cognitive performance assessment based on everyday activities for human-robot interaction," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 398–400.
- [15] Sawyer black edition. [Online]. Available: <https://www.rethinkrobotics.com/sawyer>

- [16] J. Nelles, S. T. Kwee-Meier, and A. Mertens, "Evaluation metrics regarding human well-being and system performance in human-robot interaction—a literature review," in *Congress of the International Ergonomics Association*. Springer, 2018, pp. 124–135.
- [17] R. E. Yagoda and D. J. Gillan, "You want me to trust a robot? the development of a human–robot interaction trust scale," *International Journal of Social Robotics*, vol. 4, no. 3, pp. 235–248, 2012.
- [18] B. Sadrfaridpour, H. Saeidi, and Y. Wang, "An integrated framework for human-robot collaborative assembly in hybrid manufacturing cells," in *2016 IEEE international conference on automation science and engineering (CASE)*. IEEE, 2016, pp. 462–467.
- [19] T. L. Sanders, T. Wixon, K. E. Schafer, J. Y. Chen, and P. Hancock, "The influence of modality and transparency on trust in human-robot interaction," in *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. IEEE, 2014, pp. 156–159.
- [20] S. Profanter, A. Perzylo, N. Soman, M. Rickert, and A. Knoll, "Analysis and semantic modeling of modality preferences in industrial human-robot interaction," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 1812–1818.
- [21] A. Weiss, R. Bernhaupt, M. Lankes, and M. Tscheligi, "The usus evaluation framework for human-robot interaction," in *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction*, vol. 4, no. 1, 2009, pp. 11–26.
- [22] D. Liu, J. Kinugawa, and K. Kosuge, "A projection-based making-human-feel-safe system for human-robot cooperation," in *2016 IEEE International Conference on Mechatronics and Automation*. IEEE, 2016, pp. 1101–1106.
- [23] A. Murata, "Ergonomics and cognitive engineering for robot-human cooperation," in *Proceedings 9th IEEE International Workshop on Robot and Human Interactive Communication. IEEE RO-MAN 2000 (Cat. No. 00TH8499)*. IEEE, 2000, pp. 206–211.
- [24] D. Bortot, M. Born, and K. Bengler, "Directly or on detours? how should industrial robots approximate humans?" in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 89–90.
- [25] V. Weistroffer, A. Paljic, P. Fuchs, O. Hugues, J.-P. Chodacki, P. Ligot, and A. Morais, "Assessing the acceptability of human-robot co-presence on assembly lines: A comparison between actual situations and their virtual reality counterparts," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 377–384.
- [26] D. Novak, M. Mihelj, and M. Munih, "Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction," *Robotica*, vol. 29, no. 3, pp. 367–374, 2011.
- [27] V. Villani, L. Sabatini, C. Secchi, and C. Fantuzzi, "A framework for affect-based natural human-robot interaction," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 1038–1044.
- [28] C. T. Landi, V. Villani, F. Ferraguti, L. Sabatini, C. Secchi, and C. Fantuzzi, "Relieving operators' workload: Towards affective robotics in industrial scenarios," *Mechatronics*, vol. 54, pp. 144–154, 2018.
- [29] C. Weckenborg, K. Kieckhäfer, C. Müller, M. Grunewald, and T. S. Spengler, "Balancing of assembly lines with collaborative robots," *Business Research*, pp. 1–40, 2019.
- [30] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.
- [31] biosignalsplux explorer. [Online]. Available: <https://plux.info/kits/215-biosignals-explorer-820201001.html>
- [32] R. E. Klabunde. Electrocardiogram standard limb leads (bipolar). [Online]. Available: <https://www.cvphysiology.com/Arrhythmias/A013a>
- [33] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 1934–1937.
- [34] R. Zangróniz, A. Martínez-Rodrigo, J. M. Pastor, M. T. López, and A. Fernández-Caballero, "Electrodermal activity sensor for classification of calm/distress condition," *Sensors*, vol. 17, no. 10, p. 2324, 2017.
- [35] M. van Dooren, J. H. Janssen, *et al.*, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations," *Physiology & behavior*, vol. 106, no. 2, pp. 298–304, 2012.
- [36] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *2011 Computing in Cardiology*. IEEE, 2011, pp. 85–88.
- [37] P. Brugada, J. Brugada, L. Mont, J. Smeets, and E. W. Andries, "A new approach to the differential diagnosis of a regular tachycardia with a wide qrs complex," *Circulation*, vol. 83, no. 5, pp. 1649–1659, 1991.
- [38] F. Ruschitzka, W. T. Abraham, J. P. Singh, J. J. Bax, J. S. Borer, J. Brugada, K. Dickstein, I. Ford, J. Gorcsan III, D. Gras, *et al.*, "Cardiac-resynchronization therapy in heart failure with a narrow qrs complex," *New England Journal of Medicine*, vol. 369, no. 15, pp. 1395–1405, 2013.
- [39] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Analysing noisy driver physiology real-time using off-the-shelf sensors: heart rate analysis software from the taking the fast lane project," *Journal of Open Research Software*, vol. 7, no. 1, 2019.
- [40] A. Bizzego, A. Battisti, G. Gabrieli, G. Esposito, and C. Furlanello, "pyphysio: A physiological signal processing library for data science approaches in physiology," *SoftwareX*, vol. 10, p. 100287, 2019.
- [41] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 237–249, 2011.
- [42] M. Papakostas, A. Rajavenkatanarayanan, and F. Makedon, "Cogbeacon: A multi-modal dataset and data-collection platform for modeling cognitive fatigue," *Technologies*, vol. 7, no. 2, p. 46, 2019.
- [43] T. Giannakopoulos, "pyaudioanalysis: An open-source python library for audio signal analysis," *PloS one*, vol. 10, no. 12, 2015.
- [44] N. Nourbakhsh, F. Chen, Y. Wang, and R. A. Calvo, "Detecting users' cognitive load by galvanic skin response with affective interference," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 3, pp. 1–20, 2017.
- [45] Y. Sasaki, "The truth oh the f-measure," *Manchester: School of Computer Science, University of Manchester*, 2007.
- [46] O. Salunkhe, O. Stensöta, M. Åkerman, Å. F. Berglund, and P.-A. Alveflo, "Assembly 4.0: Wheel hub nut assembly using a cobot," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 1632–1637, 2019.
- [47] M. Kyrarini, A. Leu, D. Ristić-Durrant, A. Gräser, A. Jackowski, M. Gebhard, J. Nelles, C. Bröhl, C. Brandl, A. Mertens, *et al.*, "Human-robot synergy for cooperative robots," *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 15, no. 3, pp. 187–204, 2016.
- [48] K. J. Chabathula, C. Jaidhar, and M. A. Kumara, "Comparative study of principal component analysis based intrusion detection approach using machine learning algorithms," in *2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)*. IEEE, 2015, pp. 1–6.

TABLE IV: Time Domain Feature Extraction from ECG Data

Features	Computation
mean Heart Rate	$mHR = \frac{\sum_{i=1}^N 60000/RR_i}{N}$ where N:number of RR interval terms
mean Inter-beat Interval	$mIBI = \frac{\sum_{i=1}^N RR_i}{N}$
Median RR Interval	$\widetilde{RR} = median(RR)$
Range RR Interval	$rangeRR = \max(RR) - \min(RR)$
Standard Deviation of RR intervals	$SDRR = \sqrt{\frac{\sum_{i=1}^N (RR_i - mIBI)^2}{N - 1}}$
Standard deviation of successive differences	$SDSD = \sqrt{\frac{\sum_{i=1}^N (RR_{i+1} - RR_i)^2}{N - 1}}$
Standard deviation of heart rate	$SDHR = \sqrt{\frac{\sum_{i=1}^N ((60000/RR_i) - mHR)^2}{N - 1}}$
Coefficient of variance of RR intervals	$CVRR = \frac{SDRR \times 100}{mIBI}$
Root mean square of successive difference	$RMSSD = \sqrt{\frac{(RR_{i+1} - RR_i)^2}{N}}$
Proportion of successive differences above 20 ms in percentage	$pRR20 = \frac{Count(RR_{i+1} - RR_i)_{>20ms} \times 100}{N - 1}$
Proportion of successive differences above 50 ms in percentage	$pRR50 = \frac{Count(RR_{i+1} - RR_i)_{>50ms} \times 100}{N - 1}$
Median absolute deviation of RR intervals	$MAD = median(RR_i - \widetilde{RR})$