HUMAN FACTORS ANALYSIS AND MONITORING TO ENHANCE HUMAN-ROBOT COLLABORATION

Dissertation Defense

by

AKILESH RAJAVENKATANARAYANAN

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at The University of Texas at Arlington

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2021

To Mom and Dad.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Fillia Makedon for giving me this opportunity, supporting and believing in me. Dr. Maria Kyrarini for her constant support and motivation. I would also like to thank my colleagues and labmates at the Heracleia Lab for making my time at UTA fun and productive. A special thanks to Dr. Maher Abujelala, Dr. Michalis Papakostas, and Dr. Konstantinos Tsiakas for their constant pep talks and reality checks. Moreover, I would like to thank my Ph.D. supervising committee members, Dr. Nicholas Gans, Dr. Vassilis Athitsos, and Dr. Farhad Kamangar, for their support and valuable insights during my Ph.D.

Last but not least, I would like to thank my parents Kala and Rajan, and my cousin Abhi for the continuous support and encouragement throughout my Ph.D. It would not have been possible without them.

05-April-2021

ABSTRACT

HUMAN FACTORS ANALYSIS AND MONITORING TO ENHANCE HUMAN-ROBOT COLLABORATION

AKILESH RAJAVENKATANARAYANAN, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Fillia Makedon

Human-Machine Interaction (HMI) can be defined as a way for us to communicate with machines through user interfaces. User interfaces have evolved from complicated punch cards and levers in the first analog computers to a more natural way of interaction using speech or gestures in today's digital assistants. Technological advancements in computing devices have paved the way for smart, powerful computers to be part of our everyday lives. There is also an increasing trend of using smart computing devices and robots in manufacturing lines, medical procedures, rehabilitation, and personal care.

The umbrella of HMI typically covers several areas like Human-Robot Interaction (HRI) and Human-Computer Interaction (HCI), but a new paradigm of Human-Robot Collaboration (HRC) is required to cover the growing research in collaborative robots. Collaborative robots or cobots are used where humans and robots work together as a team to achieve a common goal. Such a setup requires the robot system to understand several aspects of the human partner's behavior, including their physical and mental states, based on the area of application. Advancements in wearable sen-

sors, artificial intelligence, and robotics have made these collaborative systems smart, personalizable, and safe. Despite the abundance of research in this field, there is a lack of research to understand the different human factors, such as human behavior and cognition, to create better HRC systems.

The central focus of this work is to advance research in the field of human factors for HRC. It revolves in two axes that first explore the different cognitive and behavioral assessment systems and finally exploit the domain expertise gained to build a cognitive assessment system that simulates a real-world task. Different intelligent cognitive assessment systems are built that are capable of using physiological data to predict a specific cognitive ability effectively. Sensors such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electrodermal Activity (EDA), and RGB cameras have been used to assess the user's state. Subsequently, physiological sensors are used in an industrial collaborative assembly scenario to predict user performance and cognitive load to enhance HRC. A collaborative system is built using advanced HMI concepts to simulate a real-world scenario and collect data from human subjects. Several data, including system-specific performance metrics and multimodal sensor data, are collected to perform a data-driven evaluation of the developed HRC system for cognitive load prediction.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT v

LIST OF ILLUSTRATIONS x

LIST OF TABLES xiii

1. **Introduction** 1

    1.1 A Gentle Introduction to Human-Robot Collaboration 1

    1.2 Understanding Human Factors in Human-Robot Collaboration 2

        1.2.1 What is Cognition? 3

        1.2.2 Introduction to Cognitive Assessments 4

    1.3 Challenges and Motivation 5

    1.4 Dissertation Structure 6

2. **Cognitive Assessments and Monitoring Systems** 8

    2.1 Introduction 8

    2.2 Components of an Intelligent Assessment System 8

        2.2.1 Cognitive Assessments 9

        2.2.2 Sensors 13

        2.2.3 Intelligent Cognitive Assessment Systems 17

        2.2.4 Intelligent Interfaces 20

    2.3 Conclusion and Discussion 23

3. **Task-based Cognitive Assessment Framework using Physiological Sensors** 25

    3.1 Introduction 25

3.2 Cognitive Fatigue Prediction using EEG and Subjective User Feedback 26
3.2.1 Framework for Fatigue Analysis Using Serious Games 27
3.2.2 Experimental Setup 29
3.2.3 Machine Learning Analysis and Results 32
3.3 Conclusion and Discussion 36
4. Human Factors Analysis for Task Performance Prediction 38
4.1 Introduction 38
4.2 Predicting Task Performance using Implicit and Explicit Cues 39
4.2.1 The Sequence Learning Task 41
4.2.2 The Expert GUI System 41
4.2.3 User Performance and Engagement 42
4.3 Convolutional Neural Networks to Predict Task Performance 43
4.3.1 Task Outcome Prediction using EEG 43
4.3.2 Task Outcome Prediction using Facial Expression and Body Postures 45
4.3.3 Results 49
4.4 Everyday Activities that Affect Task Performance 51
4.4.1 Sleep and its Effect on Cognitive Performance 51
4.4.2 Experimental Study 52
4.4.3 Preliminary Results 52
4.5 Conclusion and Discussion 53
5. CogniSmart: An Intelligent Human-Factors Monitoring Framework to Enhance Human-Robot Collaboration 56
5.1 Introduction 56
5.2 Background: Human Factors Modeling for Safe Human-Robot Interaction 57

5.3 Proposed Framework 59
5.3.1 The Robot-Assisted Assembly Task 60
5.3.2 Sensors and Data Stored 61
5.4 User Study - Preliminary Analysis 69
5.4.1 Preliminary Results from Proposed Framework 71
5.4.2 Limitations and Research questions 75
5.5 Modeling Cognitive Load from Public Datasets 76
5.5.1 9PM Cognition Dataset 76
5.5.2 CLAS - Cognitive Load, Affect and Stress Detection Dataset . 78
5.5.3 Differences Between the CLAS and 9PM Datasets 79
5.5.4 Machine Learning Analysis to Predict Cognitive Load 80
5.5.5 Machine Learning Results 85
5.6 Conclusion and Discussion 92
6. Concluding Remarks and Future Directions 94
6.1 Concluding Remarks 94
6.2 Future Directions 96
6.3 Publicly Available Datasets and Implementations 97
REFERENCES 98
BIOGRAPHICAL STATEMENT 117

LIST OF ILLUSTRATIONS

Figure		Page
1.1	Research Overview	6
2.1	A sample Sequence Learning task setup. A computer plays audio of different sequences. The participant memorizes it and responds by pressing the labeled buttons in front of them.	10
2.2	The Wisconsin Card Sorting Task - Computerized version built by [1]. A stimulus card is shown on the bottom left, the user must now choose a matching card from the presented options on the top right based on color, number, or shape of the presented stimuli	11
2.3	The N-back Task. In this example, a 2-back task, sequences of shapes are used	12
2.4	MUSE EEG Headset [2]	14
2.5	ECG signal sample signal snapshot	16
2.6	EDA signal, Raw EDA and tonic component signal snapshot(top). Phasic component signal snapshot(bottom)	17
2.7	Cameras for behavioral data collection	18
2.8	Game-based intelligent cognitive training system. Experimental setup of the Towers of Hanoi Game	18
2.9	Different training methods. (a) - Computer-aided training. (b) - Human trainer. (c) - Game-based training.	19

2.10 The participant interacts with three different GUI during a robot-assisted assessment session. The *control-only* GUI (a) , does not have any monitoring features. The *history-based* GUI (b) provides a history of task performance over past rounds. The *model-based* GUI (c), provides a visualization of the performance model as a set of success probabilities at each level. 22

2.11 Participants' feedback on the enjoyability and the effectiveness of the three interfaces. 23

3.1 The WCST version implemented in [1]. Participants must play different versions **a**–**d**) of the game. In V1, the game starts with two options (**a**), and then gradually increases to 5 options (**d**), until the game is over. In V2, options a, b, c, and d change randomly after every four rounds according to the same decision rules. 27

3.2 The Data Collection Experimental Setup. 29

3.3 Analysis of the Self-reported CF and User Performance During game play [1]. (a) Average *P_Errors* during V1 and V2 WCST (b) Analysis of Self-Reported Cognitive Fatigue during V1 and V2 versions of WCST 33

4.1 Experimental Setup [3] 40

4.2 The Expert User-Interface [3]. The GUI displays performance metrics such as current sequence, user response, and engagement value. 42

4.3 Neural Network Architecture for (a) EEG Signal from Neural Network (ENN), (b) Emotions from Facial Expression (EFE), (c) Emotions from Body Pose (EBP) taken from Ramesh Babu, A. and Rajavenkatanarayanan A. et al. [4] 44

5.1 Overview of the proposed *CogniSmart* system architecture for HRC . . 59

5.2 Setup of the Proposed Collaborative Assembly Scenario, *RoboAssist*. Top-left image: Final Assembly Product - A Miniature Sanding Machine. [5] ©2020 IEEE. 60
5.3 Sensor Placement of ECG and EDA sensors. The EDA sensor is placed on the right shoulder. The ECG sensor is placed in a Lead II setup of the Einthoven's triangle. 62
5.4 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, T indicate the Q-wave, R-wave, S-wave, and T-wave component of the ECG signal. ©2020 IEEE. 64
5.5 Downsampled EDA signal acquired from the Biosignalsplux sensor. The top plot shows the preprocessed EDA signal in blue and the tonic component of the signal in orange. The bottom plot shows the respective phasic component in blue. 68
5.6 Subjective feedback from user feedback. (a) A comparison of user response across the three Baseline, post_task1, and post_task2 surveys. (b) A comparison of user response after post_task1, and post_task2 surveys. 73
5.7 Illustration of the distribution of stimulus across the different quadrants of the valance and arousal scale [6], ©2019 IEEE. 79
5.8 A Graph of Task Difficulty from User Survey Data [7]. 81

LIST OF TABLES

Table Page

3.1 Average classification results in all folds for different classifiers [1]. The column S indicates the EEG feature stream that achieved the best results after a comprehensive search in the grid. In the last row, the best results are obtained by combining the predictions of all trained models. Values in bold correspond to the methods that gave the best and most stable results. The abbreviations for Table 3.1 are as follows: Cl: Classifier, S: Signal, Pr: Precision, Rc: Recall and Ac: Accuracy. . . . 36

4.1 Task outcome prediction from EEG signal as presented in [4]. Abbreviations: SVM-Support Vector Machines, GB-Gradient Boosting, RF-Random Forests, ET-Extra Trees 50

4.2 Prediction from individual modalities and combined as presented in [4]. Abbreviations: EFE-Emotion from facial Expression module, EBP-Emotions from body postures and ENN-EEG signal with Neural Network 51

4.3 Summary of correlation analysis between stages of sleep and task performance. (τ/ρ) denotes the degree of correlation while P denotes if the correlation is significant. Initially presented by A. Rajavenkatanarayanan et al. in [8] 53

5.1 Time Domain Feature Extraction from ECG Data [5]. ©2020 IEEE. . 65

5.2 Frequency Domain Feature Extraction from ECG Data [5]. ©2020 IEEE. 66

5.3 Preliminary results using Support Vector Machines (SVM) to predict cognitive load using data collected from RoboAssist. Results are presented SVM with and without Principal Component Analysis (PCA) for each signal: ECG, EDA Phasic, and a combination of both signals. Abbreviation: F1 - F1 Score; Acc - Accuracy 74
5.4 Classification results of predicting cognitive load on the RoboAssist data. For each combination, PCA was performed and the least best C value is shown. 87
5.5 Classification results of predicting cognitive load on the RoboAssist data. For each combination, Univariate feature selection was performed and the least best K value is shown. **Best Features**: Mean, Min, Range, Energy, 1st Mean, 2nd Mean, 2nd Std, Spectral Centroid, Spectral Entropy 87
5.6 Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from the Math test, PCA was performed and the least best C value is shown. 88
5.7 Classification results of predicting cognitive load on the RoboAssist data. For each sensor data combination fom Math Test, Univariate feature selection was performed and the least best K value is shown. **Best Features:** Spectral Centroid, Spectral Rolloff, Spectral Entropy. 88
5.8 Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from Logic test, PCA was performed and the least best C value is shown. 89

5.9 Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from Logic test, Univariate feature selection was performed and the least best K value is shown. **ER Best Features:** Max, Range, Standard deviation, Variance, Zero Crossing Rate **E+ER Best Features:** BPM, IBI, medianNN, rangeRR, SDRR, SDSD, SDHR, CVRR, RMSSD, MAD_RR, Mean, Min, Spectral_Centroid 90

5.10 Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from the combined dataset, Univariate feature selection was performed and the least best K value is shown. **Best Features:** EDA Raw - RF - Mean, Min, Max, Energy, 2nd Std, Spectral_Centroid, Spectral_Rolloff, Spectral_Entropy, Entropy_of_Energy **Best Features:** EDA Raw - SVM - Mean, Min, Energy, Spectral_Centroid, Spectral_Rolloff, Spectral_Entropy, Entropy_of_Energy 91

CHAPTER 1

Introduction

1.1 A Gentle Introduction to Human-Robot Collaboration

Human-Robot Collaboration (HRC) is an interdisciplinary research area that includes robotics, data science, psychology, and cognitive sciences, to name a few. Technological advancements, consumer needs, and the need to meet personalized demands paved the way for a fourth industrial revolution. This revolution is commonly referred to as industry 4.0 [9]. Several companies have already employed robots in their manufacturing line for mundane, repetitive tasks. Complete automation has not yet been achieved in these industries because robots are not yet able to handle intricate parts for assembly or are too costly to deploy. There is also a necessity to ensure that the need to increase productivity does not remove human workers from the manufacturing industry [10]. To address these challenges, a new paradigm of collaborative robots or cobots is being used instead, where humans and robots work synchronously to achieve a common goal. Such a setup ensures the incorporation of the creativity and dexterity of the human worker and the intelligence and efficiency of the robot system. This setup enables the human worker to be more creative in the workflow rather than doing mundane tasks. This synergy between humans and autonomous machines is also called the 5th Industrial revolution, 'Industry 5.0' [10, 11].

In this dissertation, the main focus is on the human factors that affect the human worker's performance while using cobots for manufacturing industries. This setup requires the robot system to understand the human-partner-specific requirements to provide personalized adaptation to the human partner, resulting in increased

productivity and reducing errors. To this end, different cognitive human factors that affect performance in an HRC setup are explored. In the following section (Section 1.2), we will go through the types of Human Factors that are important in an HRC setup.

1.2 Understanding Human Factors in Human-Robot Collaboration

'Human Factors' may refer to several different things. For instance, it may refer to 'the human factor,' which refers to the human as a factor in a system's performance. In this research, we refer to human factors as the different factors of a human that focus on the humans' abilities, limitations, and characteristics. According to a blog post by Steven Shorrock [12], several types of factors exist that includes:

- Cognitive abilities such as attention, working memory, and reasoning that are used to perform a task,
- Cognitive systems such as the dual-process theory, which implies that cognition may involve the coordinated activity of two independent, but connected systems. For instance, this may aid in a person's problem-solving ability.
- Types of performance such as knowledge-based, skill-based, and rule-based performance,
- Error types such as a lapse in judgment, reason's slips, and mistakes,
- Physical functions and qualities such as strength, speed, accuracy, balance, and reach that aids in task performance,
- Subjective behaviors and non-technical skills such as situation awareness, decision making, and teamwork that help enhance performance in HRC,
- Physical, cognitive, and emotional states such as stress, emotion, and fatigue that may affect performance.

Such factors must be understood at a basic level to design and implement an effective collaboration system. Human factors engineering is a branch of psychology that deals with the application of psychological principles in the design and development of products or systems that involve a person and their working environment [13]. Research in this domain aims to minimize human error, enhance safety, increase comfort, reduce workload, and at the same time, increase productivity [13, 14].

The focus of this dissertation is predominantly on analyzing cognitive factors and how to assess them during interaction for adaptation and personalization. There has recently been a growing interest in the monitoring and analysis of different signals generated by a person for safe interaction with robots and computers [15]. This research focuses on using physiological and behavioral sensors for monitoring a person's cognitive and behavioral cues. These cues help predict the implicit and explicit feedback from a person during interaction to assess physical, cognitive, and emotional states. In order to understand these different states, the first step is to understand the different cognitive abilities and how to monitor and assess them using physiological sensors like Electroencephalogram (EEG), Electrocardiogram (ECG), Electrodermal Activity (EDA), and behavioral sensors such as the RGB camera for the physical state.

The reminder of this section discuss what cognition is and briefly introduce why cognitive assessments are important.

1.2.1 What is Cognition?

According to Bence Ölveczky, at the Department of Organismic and Evolutionary Biology and the Center for Brain Science at Harvard University, cognition can also be defined as the ability of a person to be able to understand their surroundings using their cognitive abilities to perform a specific task [16, 17]. For instance,

to accomplish a simple task like crossing a road, a person may use several cognitive abilities like perception (e.g: recognize and interpret a crossing junction), attention (e.g: concentrate on the pedestrian signal to look out for change in the signal), motor skills (e.g: mobilize the muscles to walk), visual and spatial processing (e.g: ability to process the stop and walk signal appropriately), and executive functions (e.g: decide to walk fast or slow).

Human activity involves using several cognitive abilities such as working memory, attention, problem-solving, or set-shifting, and continuous usage of cognition leads to a cognitive load. The overuse of these abilities might lead to cognitive fatigue that causes many issues for a worker in a manufacturing line or a nurse in emergency care [18]. Thus, it is essential to monitor such loads in an automated closed-loop system that can provide the necessary assistance to reduce errors or mistakes due to cognitive fatigue or cognitive load.

1.2.2 Introduction to Cognitive Assessments

A cognitive assessment is commonly used to assess cognitive capabilities and to determine cognitive impairment levels. While cognitive assessments like the Montreal Cognitive Assessment [19] or the Mini-Mental State Exam [20] cannot give the exact cause of the impairment, it helps a doctor or caregiver to get an idea of the level of impairment so that more tests, if needed, can be performed to address the problem. In rehabilitation and healthcare, specifically for patients suffering from neurological conditions such as Traumatic Brain Injury or Multiple Sclerosis, cognitive assessments are regularly performed to track and discover mental impairments and learning difficulties. In this scenario, cognitive assessments are used to assess the mental capacity of a person to think, solve problems, memorize, execute complex behaviors, and carry out day-to-day activities. Several tests have been proposed to assess different types of

abilities in patients. The National Institute of Health Toolbox (NIH Toolbox) is one of the most popular initiatives that provide a standardized and comprehensive set of neuro-behavioral measures to assess cognitive, emotional, sensory, and motor functions [21, 22]. The NIH Toolbox provides several easy-to-use cognitive assessments that can be administered from a portable tablet. This research uses variations of different tests proposed in the NIH toolbox and others to build intelligent cognitive assessment systems to understand human cognition. The studies' results are then applied in a real-world setup.

1.3 Challenges and Motivation

Traditional Human Robot Interaction (HRI) design covers collision detection and avoidance issues and ensures that the robot should not cause immediate injury or harm the human user [23]. Recent research has shown how incorporating human factors in a manufacturing process can improve the HRC outcome [24, 25]. There is also some focus on Ergonomic HRI in the industry by developing frameworks to improve human posture and minimize the risk of developing work-related musculoskeletal diseases, and disorders [26]. However, it is not clear about the types of cognitive or behavioral factors that impact human performance in these systems.

To this end, a two-pronged research is adapted. Figure 1.1 shows a high-level outline of this research. In the first step, intelligent assessment systems using a closed-loop automated assessment system are built. These assessment systems incorporate physiological sensors to monitor the user's cognition using implicit or explicit feedback. Additionally, different intelligent user interfaces that help the caregiver or test administrator to make informed decisions during assessments are developed. Subsequently, other indirect factors from a human's day-to-day lives that may affect

cognitive performance are explored. In the second step, these results are studied and incorporated in an industrial assembly setup that resembles a real-world setup.

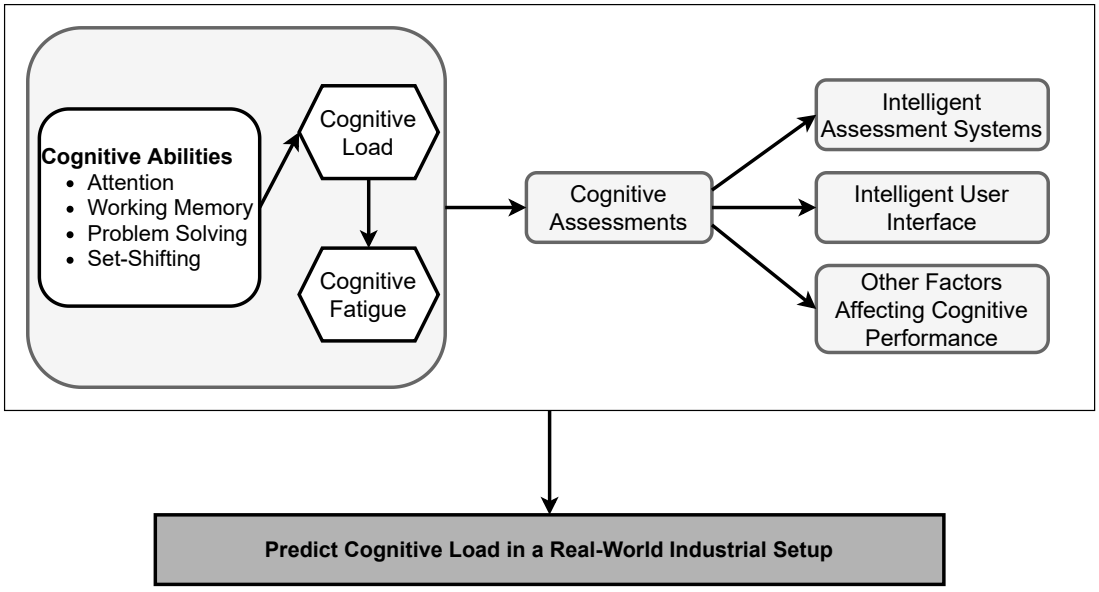


Figure 1.1: Research Overview

1.4 Dissertation Structure

The rest of this dissertation is outlined as follows. Chapter 2 makes an in-depth discussion of the different cognitive assessment systems and the components of an intelligent assessment system. It also presents how such systems are beneficial in studying the human cognitive model. Chapter 3 discusses and evaluates a task-based cognitive assessment system that tries to capture the user model using physiological sensors. Chapter 4 presents different studies that examine different human factors that affect task performance. Chapter 5 discusses the industrial assembly system, which simulates the real-world scenario. Challenges in achieving this system are also presented, along with the steps taken to address them. Finally, Chapter 6 concludes

this dissertation with future research directions and highlights the takeaways of this research.

CHAPTER 2

Cognitive Assessments and Monitoring Systems

2.1 Introduction

While cognitive processes have been a topic of research in philosophy for several centuries, recent advances in brain imaging and other technologies have shown that human cognition works using neuronal activity in the brain. People suffering from neurological diseases that affect the brain's neuronal activities like Multiple Sclerosis [27] and Parkinson's disease [28] have shown a decline in cognitive abilities that may lead to unemployment and disrupt the social functioning of patients. Although these conditions can not be cured, proper rehabilitation and treatment may help them manage several symptoms. Several cognitive assessments have been developed over the years to assess the level of cognitive functioning to provide the patients with a proper rehabilitation and treatment plan. This chapter introduces the different components of a cognitive assessment and monitoring system and briefly explains why there is a need for intelligent systems that can increase user engagement and automatically monitor user performance.

2.2 Components of an Intelligent Assessment System

An intelligent cognitive assessment system consists of multiple components such as cognitive assessments, an intelligent interface for the test administrator to easily interpret the results, and sensors to monitor the user's cognitive state. All these components are essential for an effective assessment system. This section will introduce

the cognitive assessments, sensors used, and our findings about intelligent cognitive tests and interfaces.

2.2.1 Cognitive Assessments

As more industries adopt robots to increase productivity, there is an increased need for effective HRI, especially when heavy and high precision robots are used. In an HRI-based manufacturing industry, the cost of human errors due to high cognitive workload, human decision making, situational awareness, and other cognitive processes are very high [24]. Mistakes in industrial scenarios involving heavy machinery and robots can turn out to be fatal. Cognitive ailments like cognitive fatigue, high cognitive load, lack of concentration, or engagement also possess considerable workplace safety risks [29]. Such an unsafe working condition may also induce health issues in workers that will be harmful to their everyday lives. Hence, in an HRC scenario, it is required to focus on human cognitive assessments that will help build a synergistic collaborative framework. To address these requirements, the first component of an intelligent assessment system is the cognitive assessment. In this research, we have focused on understanding these conditions through several cognitive tests that require the users to be vigilant and imposes a high cognitive load. The following sections briefly discuss the different cognitive assessments used in this dissertation.

2.2.1.1 Sequence Learning

Sequence learning (SL) tasks [30, 3] evaluate a person's ability to arrange thoughts and information in a meaningful order. This ability has been recognized as an essential ability for vocational assessment, especially in HRC, where the user has to exercise attention, good working memory, and decision-making in order to interact with robots safely and efficiently [31]. Several cognitive science research has

shown that sequence learning can be used to assess human behavior towards learning ability, working memory, and attention [32, 33]. Different types of SL tasks are used to predict and assess different abilities, such as sequence prediction, sequence generation, sequence learning, and sequence recognition. This work focuses on studying the users' sequence learning and working memory, which is essential in industries involving assembly line work. The SL task involves listening or seeing a set of character sequences and repeating them correctly in a certain amount of time. The sequences are delivered via audio (speech) or image on a computer screen. In this research, we used three alphabets such as a, b, and c to provide the participant sequences of lengths varying from three to nine. The participant must remember the sequence and then responds by pressing the buttons labeled A, B, C in front of them. Figure 2.1 shows the sample setup of the SL task. The performance outcomes from the SL task can help therapists and other experts determine what particular treatment or rehabilitation an individual might need to enhance his/her performance in a given domain or application.

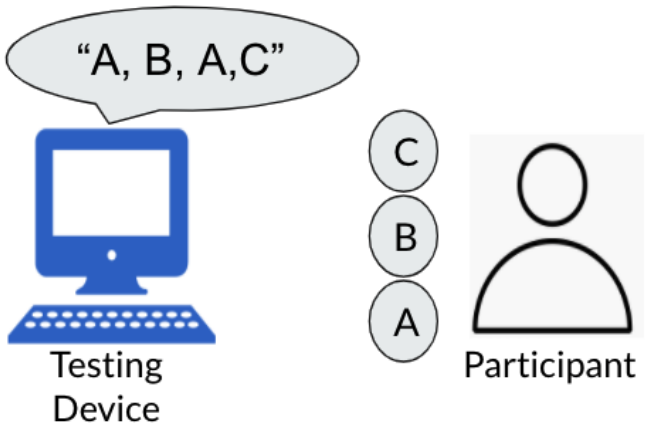


Figure 2.1: A sample Sequence Learning task setup. A computer plays audio of different sequences. The participant memorizes it and responds by pressing the labeled buttons in front of them.

2.2.1.2 Wisconsin Card Sorting Task

The Wisconsin Card Sorting Test (WCST) is a cognitive test used to test set-shifting. Set-shifting is the ability of a person to display flexibility in the face of changing schedules of reinforcement [34]. The WCST is a cognitive task that incorporates cognitive challenges such as short-term memory, adaptive decision-making, and problem-solving that can be found in a great variety of daily living activities. In this task, the user is shown several stimulus cards in random order. The stimulus card can contain different shapes that vary in number and color. The user is told to match the cards but not how to match them. The only feedback given to the user is whether the match is right or wrong while they can choose to use several rules based on color, shape, or number of the symbols. At each turn, only one rule applies,

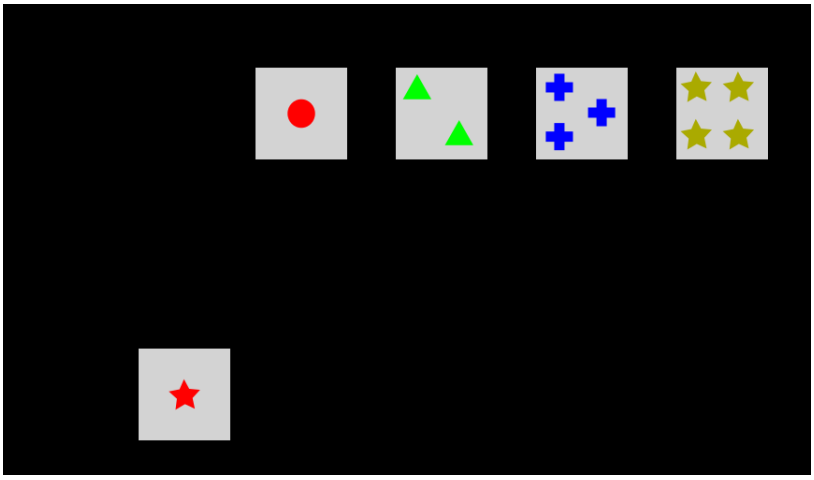


Figure 2.2: The Wisconsin Card Sorting Task - Computerized version built by [1]. A stimulus card is shown on the bottom left, the user must now choose a matching card from the presented options on the top right based on color, number, or shape of the presented stimuli

and the user's goal is to derive the rule based on the feedback provided. The rule keeps changing periodically after a few rounds, and the user must guess the new rule.

Figure 2.2 shows a sample screenshot of the WCST, which shows the computerized version of the task based on the test offered by PsyToolkit [35].

2.2.1.3 N-back Task

The N-back task is commonly used in cognitive neuroscience to measure working memory and attention, two essential skills in an assembly task [36]. It has also been used in the past to induce cognitive load, and cognitive fatigue [37, 8]. It is a sequential cognitive task where the participant is presented with stimuli sequentially one-by-one. For each stimulus, the participant needs to decide if it is the same as the one presented N stimuli back.

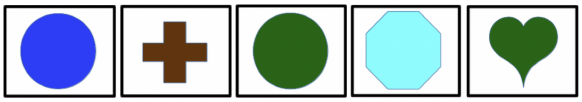


Figure 2.3: The N-back Task. In this example, a 2-back task, sequences of shapes are used

Several versions of the N-back task exist, such as the N-back task with alphabets for sequences and the visual N-back task where stimuli appear on different screen positions. In our proposed version in [8], we present the user with a sequence of shapes one at a time, as shown in Figure 2.3, from a pool of eight shapes with eight different colors. Shapes were chosen to induce additional cognitive processing, using different shaped targets of different colors. Figure 2.3 depicts a 2-back task based only on the shape. The user responds by pressing the space bar on the keyboard when they see the second circle, which repeats after exactly two stimuli. The participant was presented with a total of 64 stimuli, of which 12 were targets. Each stimulus lasted for 2500 ms. Several levels of the task are possible starting from 0-back, where

a participant is pre-informed with the target stimuli to respond to and can go up to how much ever is required.

2.2.2 Sensors

This dissertation follows a data-driven approach involving several user studies that comprise data collection from human subjects, data analysis, and user modeling. Data collection in our user studies involves physiological and behavioral data. Physiological data is measured using sensors capable of measuring the autonomic nervous system's involuntary response to stimuli [38]. Thus, the second component of an intelligent assessment system that monitors users' physiology is the sensors. This research uses data from EEG, ECG, and EDA sensors capable of recording user data non-invasively. Physical reactions to stimuli such as body postures and facial expressions are classified as behavioral data. RGB webcam and RGB-D RealSense sensors are used to capture behavioral data.

The following subsections describe the sensors used in our studies classified based on the type of data.

2.2.2.1 MUSE EEG headset

Muse EEG headset is a non-invasive wearable device, widely used for Brain-Computer Interface systems [39]. Muse has four electrodes, two over the prefrontal lobe and two behind the ears. It allows us to record EEG activation at a sampling rate of 220 Hz. Using the digital signal processing unit embedded in the device, we can store other information and features extracted from the individual EEG frequency bands namely: gamma 32-100 Hz (g), beta 13-32 Hz (b), alpha 8-13 Hz (a), theta 4-8 Hz (t) and delta 0.5-4 Hz (d) at a sampling rate of 10 Hz. For each of the four sensors, we record different types of data streams like Raw EEG, Absolute Frequency

Bands (A), Relative Frequency Bands (R), Session Score for each Frequency Band (s), and Signal Quality Indicator (h).

Figure 2.4: MUSE EEG Headset [2]

- **Absolute Frequency Bands (A):** The absolute band power for a given frequency range is the logarithm of the sum of the Power Spectral Density of the EEG data over that frequency range.

$$xA = \log \sum_{i=f\_low}^{f\_high} |G(f_i)|^2 \tag{2.1}$$

  where f_low and f_high are the minimum and maximum frequencies of frequency band x and G is the fast fourier transform (FFT) of the EEG signal g.

- **Relative Frequency Bands (R):** The relative band powers are calculated by dividing the absolute linear-scale power in one band over the sum of the absolute linear-scale powers in all bands.

$$xR = \frac{10^{xA}}{10^{aA} + 10^{bA} + 10^{dA} + 10^{gA} + 10^{tA}} \tag{2.2}$$

  where x is one of the five frequency bands.

- **Session Score for each Frequency Band (s):** is a value computed by comparing the current value of a band power to its history in sampling frequency of 10 Hz. This value is mapped to a score between 0 and 1 using a linear function that returns 0 if the current value is equal to or below the 10th percentile of the distribution of band powers, and returns 1 if it's equal to or above the 90th percentile. Linear scoring between 0 and 1 is done for any value between these two percentiles.
- **Signal Quality Indicator (h):** is an integer value from 1 (optimal quality) to 3 (very bad quality).

2.2.2.2 Biosignalsplux Kit

Biosignalsplux [40] is a wearable body sensing kit that is capable of collecting data from different types of sensors. In our studies, we use the ECG and EDA sensors. The data are transmitted via Bluetooth at 1000 Hz. ECG data were collected from a standard 3-point bipolar limb leads configuration of the Einthoven's triangle [41]. A Lead II setup in this configuration is used, 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 EDA sensor measures the electrical potential produced on the skin surface due to the sweat glands' activity [42]. The best locations to acquire such signals are spots where sweat glands are most active, like the palms and the soles [43]. Raw data from the sensor is stored on a data server in a Comma-Separated Values (CSV) format. Unit conversion is required for the data transmitted by the sensor since the data is in digital units. For instance, ECG is measured in millivolts and hence a unit conversion is applied with the formula:

$$ECG(V) = \frac{(\frac{ADC}{2^n} - \frac{1}{2}) * VCC}{G_{ECG}} \tag{2.3}$$

$$ECG(mV) = ECG(V)/1000 \tag{2.4}$$

where $ECG(V)$ is the ECG value in volt (V), $ECG(mV)$ is the ECG value in milli volts, ADC is the value sampled from the channel, n is the number of bits per channel equal to 16, VCC is the operating voltage equal to 3 V, and G_{ECG} is the sensor gain equal to 1000. Figure 2.5 shows a sample snapshot of the ECG signal after pre-processing.

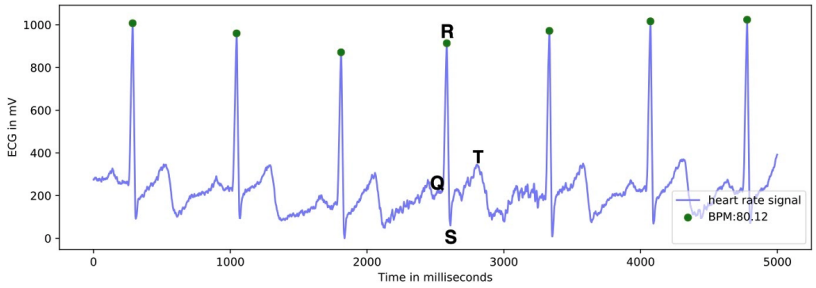


Figure 2.5: ECG signal sample signal snapshot

Similarly, EDA signals are measured in microsiemens (μS) and hence a unit conversion is required using the formula:

$$EDA(\mu S) = \frac{\frac{ADC}{2^n} \cdot VCC}{k} \tag{2.5}$$

where $EDA(\mu S)$ is the EDA value in microsiemens, ADC is the value sampled from the channel, n is the number of bits per channel equal to 16, VCC is the operating voltage equal to 3 V, and k is a constant value 0.12. Figure 2.6 shows a sample snapshot of the EDA signals after pre-processing the raw signals. The different components of the signals are be explained in Chapter 5.

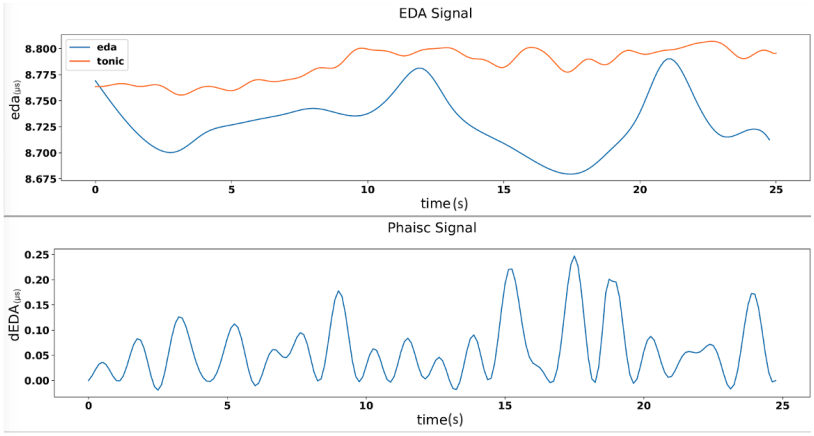


Figure 2.6: EDA signal, Raw EDA and tonic component signal snapshot(top). Phasic component signal snapshot(bottom)

2.2.2.3 Behavioral Data

User behavioral data such as body postures and facial expressions are captured using sensors such as an RGB web camera (Figure 2.7a) and an RGB-D RealSense D435i camera (Figure 2.7b). We use a Logitech HD Pro web camera that records RGB data at the rate of 30 fps with 1920 x 1080 pixels. The Intel RealSense D435i camera consists of depth and tracking technologies designed to give machines and devices depth perceptions capabilities. The D435i version, as shown in fig 2.7a, consists of a pair of depth sensors, an RGB camera, an inertial measurement unit, and infrared projectors. The camera uses the active IR stereo technology [44] for depth data and can capture data at the rate of up to 90 frames per second (fps) and 1280 x 720 pixels. The RGB camera can capture data at the rate of 30 fps with 1920 x 1080 pixels.

2.2.3 Intelligent Cognitive Assessment Systems

Popular libraries such as the NIH Toolbox [21] and the Psy-Toolkit library [35] have designed and released online tests for several cognitive abilities. These libraries have proven to be very useful in a clinical setup. One of the major drawbacks

(a) Intel RealSense D435i camera (b) Logitech HD Pro Webcam C920

Figure 2.7: Cameras for behavioral data collection

of such tests is the monotonous nature of these tests. The repetitive nature may lead to frustration and boredom, thereby leading to ineffective data due to a lack of engagement towards the task [45]. This drawback leads to the third component of an intelligent assessment system, cognitive assessments that can innovatively engage participants. Game-based training and assessment techniques have been developed to improve the engagement towards the task while at the same time improving the data quality. In our study [46], we developed an intelligent game-based cognitive

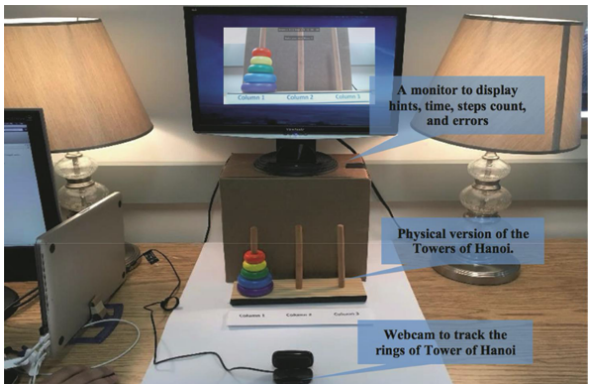


Figure 2.8: Game-based intelligent cognitive training system. Experimental setup of the Towers of Hanoi Game

training system, shown in Figure 2.8 using a popular game, the Towers of Hanoi

(TOH). The developed system provides personalized assistance to the trainee and improves cognitive abilities like planning and problem-solving skills.

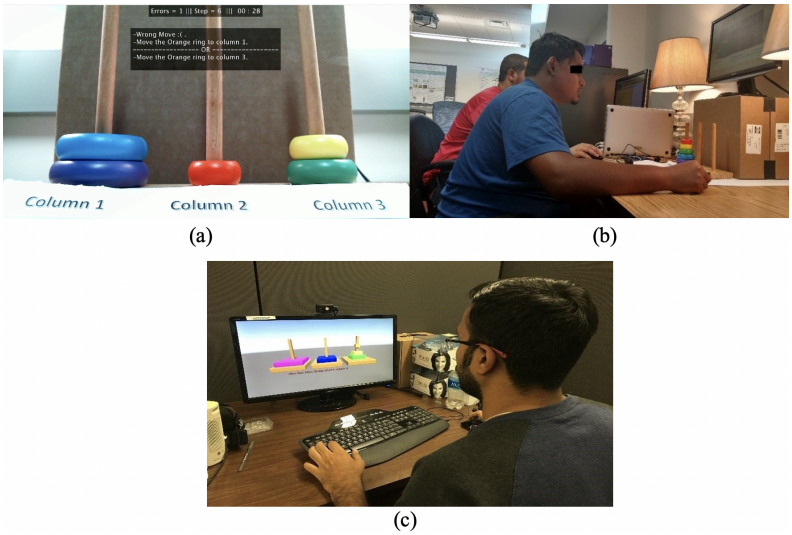


Figure 2.9: Different training methods. (a) - Computer-aided training. (b) - Human trainer. (c) - Game-based training.

In this setup, the participants were trained to solve the TOH game using three different training methodologies, as shown in Figure 2.9, and then tested to evaluate each training method's effectiveness. The different training methods were (a) traditional (with a human trainer), (b) gamification (game-based training simulation), and (c) computer-aided training. In the training sessions, the participants were given step by step instructions to solve the task using the optimal number of steps (31 steps). However, in the testing session, there were no restrictions on the number of steps. Participants had to solve the game from memory.

In the *traditional method*, the participants were trained with a personal human trainer as shown in Figure 2.9b. The trainer went through the steps verbally to solve the TOH with the participants. This training was timed and the number of errors while solving was recorded manually. The *game-based training* (Figure 2.9c), used a

TOH game was designed using Unity game engine. The game provided instructions using audio and the score and number of steps were displayed on screen. In *computer-aided training* method (Figure 2.9a), the participants were trained with instructions that appeared on the screen. That is, instead of an individual trainer, participants were asked to solve TOH with instructions flashing on the monitor in front of them. Every step performed was captured through a webcam to evaluate the accuracy of the steps from the instructions that appeared.

Results from this study indicated that the users not only liked the game-based and computer-aided training better than the traditional human-based training, but they also felt that the game-based training allowed the participants to think on their own rather than someone often giving pointers at each mistake. Cognitive training and abilities development systems need to allow participants to learn to solve the problem independently and develop their knowledge. This also leads to an improvement in their cognitive abilities.

2.2.4 Intelligent Interfaces

Traditionally, the focus of an intelligent assessment system is the end-user (patients or trainees). Few researchers have proposed cognitive assessment systems that incorporate a secondary user (e.g., supervisor, teacher, therapist), who can monitor and support the interaction between the intelligent system and the primary user through Graphical User Interfaces (GUI) [47], the fourth component of an intelligent assessment system. The secondary users conduct these assessments in most cases, and it is important to focus some attention towards the development of intelligent GUI that can be exploited to enhance secondary users' decision-making.

In our study [48], we investigate which GUI feature (e.g., visualization, transparency, etc.) enhances human decision making and efficiency in a user skill assess-

ment task. To this end, we used an adaptive robot-assisted cognitive assessment and training framework proposed by Tsiakas *et al.* [49]. Robot-Assisted Training (RAT) is an HRI research area that studies how to use robots to assist users during a training task [50]. Data collected using this framework was used to create a user simulation model instead of a real human player (primary users).

In the user study, 30 participants were recruited to monitor and guide the robot-assisted assessment session using the NAO, a socially-assistive robot [1]. We used the sequence learning task described in Section 2.2.1.1 for assessment. The participant pool consisted of 24 male and six female participants, most of whom did not have prior experience interacting with a socially-assistive robot. The study administrators acted as "fake" primary users to provide a realistic environment for the participants. Participants thought that the research administrators performed the task while, in reality, they interacted with the user simulation model.

The participants run the assessment session by selecting different difficulty levels for the assessment, using the GUI. Three different types of GUI were developed, as shown in Figure 2.10. The *control-only* GUI includes only the buttons that control the next difficulty level and a label that displays the outcome at the end of each round. The *history-based monitoring* GUI displayed the score obtained at each round in a graph apart from the label for outcome and buttons to control difficulty. The graph displayed score as a function of outcome and difficulty level computed using the formula: $s = outcome * level$, where $outcome = [-1, 1]$ and $level = [1, 2, 3, 4]$. The *model-based monitoring* GUI visualizes an estimation of the performance model $P(success|level)$, where the x-axis represents the difficulty level and the y-axis the probability (frequency) of success at each level, updated after each turn. The goal of the secondary user in this scenario was to be a trainer or a therapist who would want

[1] `https://www.ald.softbankrobotics.com/en/cool-robots/nao`

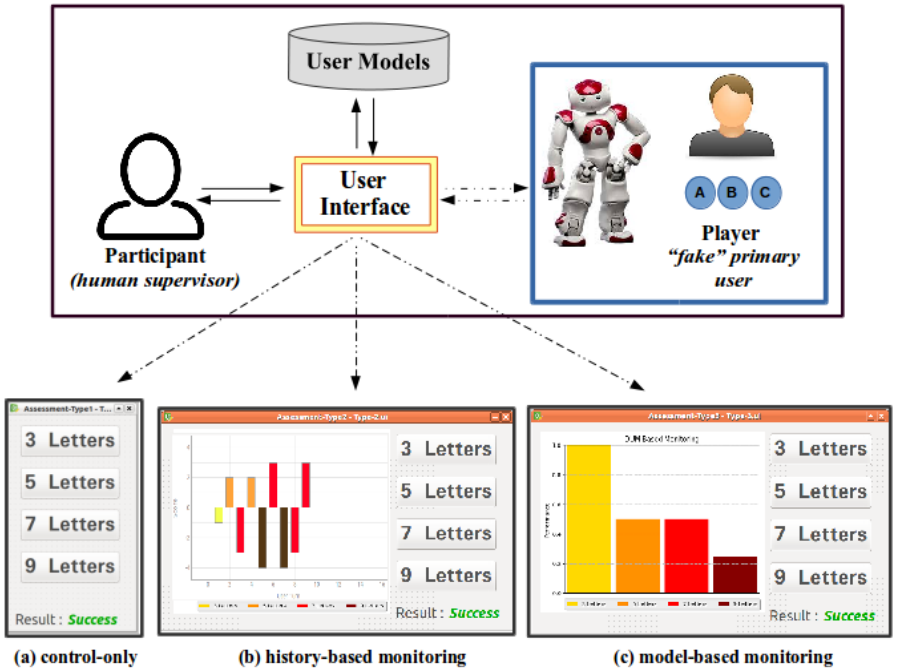


Figure 2.10: The participant interacts with three different GUI during a robot-assisted assessment session. The *control-only* GUI (a) , does not have any monitoring features. The *history-based* GUI (b) provides a history of task performance over past rounds. The *model-based* GUI (c), provides a visualization of the performance model as a set of success probabilities at each level.

to increase the difficulty of the task gradually, but at the same time, they should take in to account the errors and the user's performance to decrease the difficulty level if needed. Our goal was to evaluate what level of information helps the user better understand the primary user's performance.

Participant's survey results shown in Figure 2.11 indicate that they preferred the history-based monitoring GUI compared to the others. They reported that the history-based GUI was more enjoyable and effective in judging a user's performance. It can be observed that it is essential to provide the users with as much information as possible and, at the same time, present it in a transparent and simple manner. The performance graphs and other details for the secondary users like supervisors,

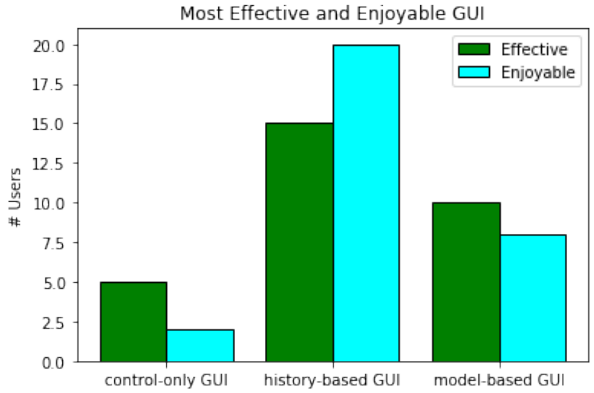


Figure 2.11: Participants' feedback on the enjoyability and the effectiveness of the three interfaces.

teachers, and therapists must be easy to understand, facilitating fast and reliable assessment.

2.3 Conclusion and Discussion

In this chapter, we discussed the different components of an intelligent assessment system, such as cognitive assessments, sensors, intelligent cognitive assessment systems, and intelligent interfaces. We saw why these components are essential in an intelligent system and the results of user studies that indicated the importance and user-enjoyability of these components. Specifically, the user study discussed in Section 2.2.3 indicated that the users felt an intelligent computerized system was more helpful than a traditional system. These results indicate that our intelligent system can provide the same functionality as the traditional system and provide more long-term benefits to the users. Similarly, Section 2.2.4 discussed a user study that explores a novel component for a cognitive assessment system, the expert user interface. Such interfaces can provide all the test details in real-time during the assessment transparently and straightforwardly.

Despite these results, the assessment systems discussed so far lack a major component, human-sensing. The assessment system discussed in Section 2.2.3 considers just the user responses for assessment. While this is similar to traditional cognitive assessments, we are left to assume how the user felt or get a subjective idea of how a user might feel after/during the assessment via surveys. While surveys are currently a gold standard to collect subjective user feedback in clinical studies, it suffers from extreme bias from the user who takes the assessment and administers it. The subjective nature of these surveys makes it complicated for an accurate assessment and sometimes tends to delay the disease or assessment prognosis. In the following chapters, we will discuss how this research address this problem.

CHAPTER 3

Task-based Cognitive Assessment Framework using Physiological Sensors

3.1 Introduction

Fatigue is a common psychophysiological condition that is prevalent in people caused due to physical or mental exertion. In the manufacturing industry, working in shifts is common to increase production to meet global demands. This practice tends to affect the sleep cycle and rest time of employees and may, in turn, result in fatigue, which is a very unsafe workplace condition [51]. It is also a common disabling symptom among several medical conditions like Multiple Sclerosis (MS) [52], Traumatic Brain Injury [53] and Parkinson Disease [54], among others. Despite its common occurrence in everyday life and different fields, there is no consensus on the definition of fatigue. Cognitive fatigue may manifest as a loss in cognitive control, high-level information processing, and attention [55]. In an HRC scenario, the human partner must be vigilant and cognitively alert, which is an important human factor to maintain the task's safety and productivity [24].

While there is no golden standard for measuring fatigue, it is commonly measured in rehabilitation as a subjective measure using surveys and clinical tests based on observation and patient self-report. These measurements are susceptible to errors due to bias that may lead to a wrong diagnosis and, consequently, delays progress [56]. Similarly, in an industrial environment, detecting that a person's performance is affected due to fatigue would be a great asset to an employer in avoiding workplace-related injuries or fatalities. This chapter presents an intelligent, objective cognitive assessment techniques that address these concerns. EEG signals are used to detect

cognitive fatigue using a serious-game based framework to assess cognitive fatigue's impact on user performance.

3.2 Cognitive Fatigue Prediction using EEG and Subjective User Feedback

Cognitive Fatigue (CF) is a very common symptom caused due to mental exertion in healthy adults. In medicine, CF and physical fatigue are one of the most disabling symptoms in patients suffering from diseases such as Multiple Sclerosis (MS), Lupus [57], Parkinson's disease [58], Chronic Insomnia or bad sleep quality [59], Traumatic Brain Injury (TBI) [60], and others.

In behavior modeling, detecting and predicting CF is not a new problem. In the past, various studies have tried to solve this problem by adopting different methods under different experimental assumptions. However, due to its high degree of ambiguity, this is still an unresolved problem. Although it is vital for many applications, there are very few (if any) datasets that can be used to solve this problem. In 2004, Hursh et al. [61] was one of the first research groups that tried to use computer modeling techniques to predict CF. Specifically they proposed FAST, a fatigue prediction tool designed to help operators in the transportation industry. FAST used the SAFTE model, a computational architecture used to model fatigue based on signal analysis related to the operator's sleep activity and task performance. In 2007, Donovan et al. [62] further emphasized the potential of cognitive models to predict fatigue in a user study of 256 women receiving early breast cancer treatment. A few years later, Gonzalez et al. [63] used the ACT-R [64] cognitive architecture to predict the fatigue level of users when performing data entry task. Their method uses the principles described in the ACT-R architecture and uses a rule-based decision-making method to assess the impact of fatigue on certain performance parameters such as task accuracy and response time. ACT-R has also inspired other advanced methods

related to monitoring fatigue and performance in intelligent driving and occupational safety applications. In 2018, Golan et al. [65] emphasized the importance of subjective reporting of CF and its impact on the cognitive function of MS patients. In this work we take advantage of a cognitive assessment task to induce CF. Data from this setup is then used to build a machine learning-based analysis of EEG signals towards identifying CF.

3.2.1 Framework for Fatigue Analysis Using Serious Games

Serious games usually refer to virtual games used for training, simulation, or education and can engage users in cognitive and physical tasks. For this study [1], we used the Wisconsin Card Sorting Task (WCST) described in Section 2.2.1.2.

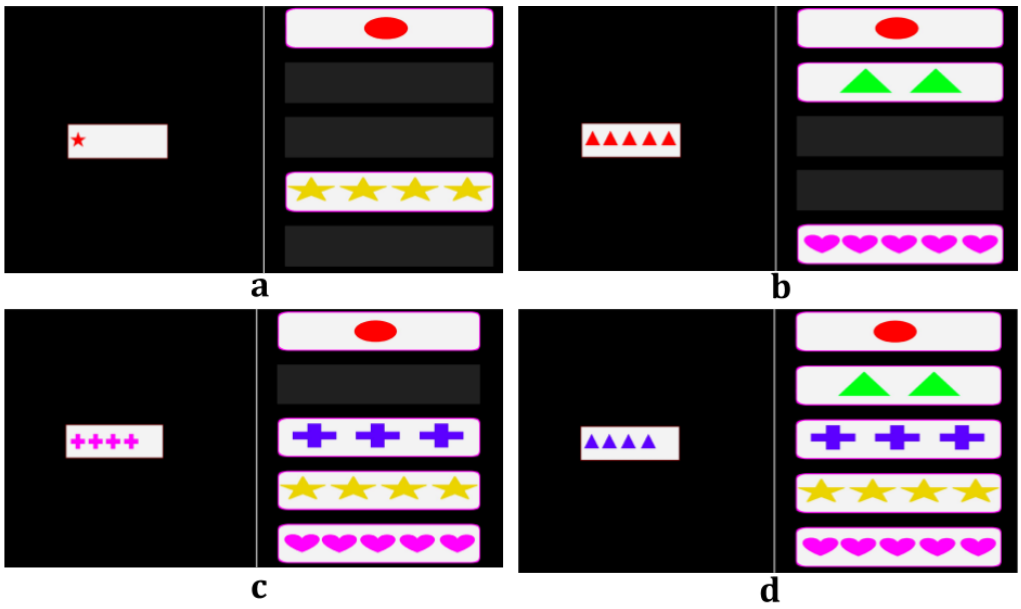


Figure 3.1: The WCST version implemented in [1]. Participants must play different versions **a**–**d**) of the game. In V1, the game starts with two options (**a**), and then gradually increases to 5 options (**d**), until the game is over. In V2, options a, b, c, and d change randomly after every four rounds according to the same decision rules.

To induce CF, two different versions of the WCST task were developed, as shown in Figure 3.1, which aims to increase the complexity and user demands. In version one (V1), the user is given two options to choose from at the start of the game, in contrast to the original task's standard four choices. The number of possible choices is increased gradually up to five possible choices as the game progresses. In version two (V2), the total number of choices displayed is changed randomly as the decision rule changes. In both versions, the number of possible options ranged from two to five. Compared to the original, the number of rounds in the game increased from 60 to 128; the decision rule changes every four rounds instead of six rounds in the original, and the maximum response time is four seconds while it is 6 seconds in the original.

3.2.1.1 Real-Time User Reports on Cognitive Fatigue

During each session, participants were asked to report if they had difficulty keeping up with the task by pressing a button in front of them (see Figure 3.2). The button can be used at any time during a session as often as participants deem appropriate. Therefore, the button push event would serve as an indicator that the user is feeling overwhelmed with the game and could result from a person's inability to pay attention, boredom, difficulty remembering decision rule, or any other reason/condition that could potentially affect the performance of the task according to the participant's subjective opinion. For this study, all conditions mentioned above were considered indicators of cognitive fatigue and were thus used to label the dataset for cognitive fatigue.

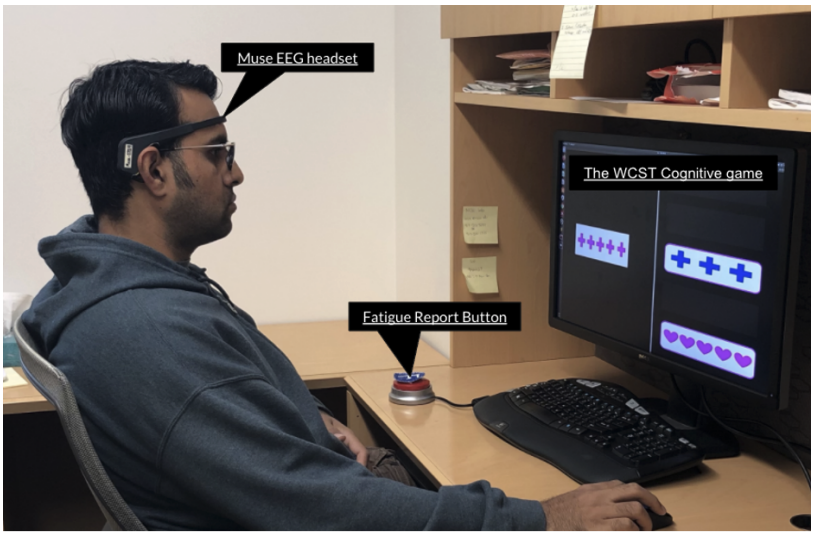


Figure 3.2: The Data Collection Experimental Setup.

3.2.2 Experimental Setup

For this study, data was collected from 19 participants that include 13 male and six female participants. The participants had no prior experience with the task, and it took approximately up to 30 minutes to complete the game. The game was designed to be demanding in terms of user engagement and attention. Hence, the participants had to pay close attention to complete the game. Data were collected over two sessions in order to include data from the two different versions of the task. During each session, the participants played the original WCST designed by our team, which follows the same rules and guidelines described by the original WCST [35]. Then, the participants had to play either the V1 or the V2 of the modified WCST. The main difference between the two sessions was in the second part of the task. During the second part of the first session, participants were asked to play the V1 version of the WCST, while in the second session, they had to play V2. The order of the tasks presented to the users was randomized in order to avoid the order effect

in the dataset.

During each collection, the following steps were followed:

1. Understand the instructions of the original WCST task as described by the researcher
2. Play the original WCST task
3. Complete a post-completion questionnaire to report subjective fatigue
4. Understand the instructions of the modified WCST (V1 or V2) task.
5. Play the modified version of WCST
6. Complete a post-completion questionnaire to report subjective fatigue

No resting time was incorporated between the two tasks to incorporate the participants' cumulative cognitive load to ensure the onset of cognitive fatigue.

3.2.2.1 Physiological and Task-based Data

Data was collected from the MUSE EEG headset (Section 2.2.2) to capture the participant's mental state during the interaction. The data collection resulted in 76 sessions that were later used to build a machine-learning model to predict CF. Moreover, task-based performance metrics were also stored for preliminary analysis.

During each of the different tasks, the designed game stores different metrics such as:

- *Outcome*: A binary flag to indicate if the participant's response in a round was correct.
- *P_Errors*: The cumulative number of perseverative errors until the current round. Perseverative errors are errors committed by the participant using the wrong rule despite several negative feedback.
- *NP_Errors*: The cumulative number of non-perseverative errors until the current round. Non-perseverative errors are errors committed by the participant

when the game changes rules. Since there are three possible decisions in a round (based on color or shape, or number), the participant must realize the correct rule before the third round. All errors before the third round are non-perservative errors, while after the third round are perservative errors.

- *C_Answers*: The total number of correct answers.
- *Response_Time*: Time to respond to stimuli at each round.
- *Score*: An indicative round-based score computed as:

$$score = \frac{\#available\_choices}{response\_time \times \#round\_under\_same\_rule} \tag{3.1}$$

The score is only computed if the participant's answer was correct; otherwise, the score is 0.

In addition, for every round the system logs the following game parameters:

- The number of possible choices offered by the system: 2, 3, 4, or 5.
- The type of the correct stimulus: color, shape, or number.
- The value of the correct stimulus:
  - If color: green, yellow, blue, red, or magenta
  - If shape: triangle, star, cross, circle, or heart
  - if number: one, two, three, four, or five

3.2.2.2 Preliminary Analysis

A preliminary analysis based on the survey results found that of the 38 data collection sessions (19 participants $\times$ two sessions) in 28 of them ($\sim$ 74% of the time), respondents reported being more tired by the end of the process versus what they felt just before the experiment started. In addition, most participants suggested that they had to work harder to adapt to the variety of options that the modified versions

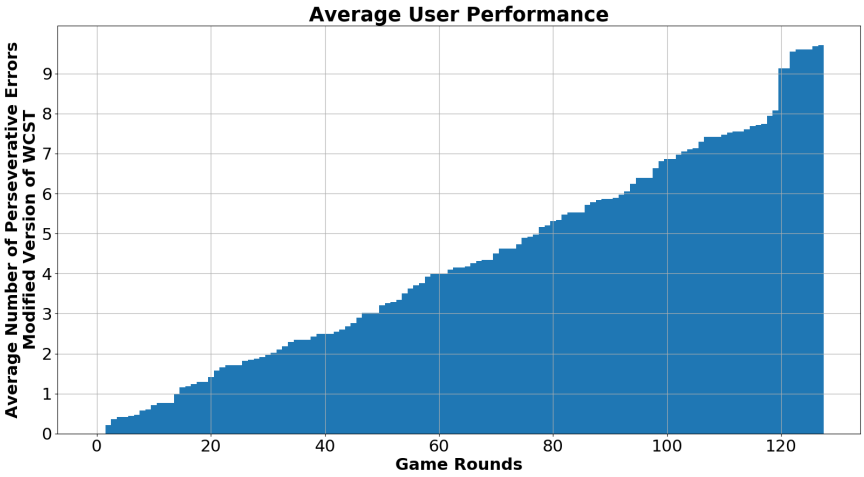
of the games offered. Survey results also showed that on a scale of 1 (no fatigue) to 5 (very tired), there was an average increase in fatigue of 1.05 points with a standard deviation of 3.54 over the 38 data collection sessions. This analysis shows that the data acquisition framework could induce fatigue.

Further analysis of the user fatigue self-report and task performance data are shown in the Figures 3.3. Figure 3.3a shows the average *P_Errors* during the game. On average, each participant made 9.3 *P_Errors* in each session and increased across all users. *P_Errors* are "unwanted" errors committed by the participants due to choosing a wrong decision rule for a round, despite the system's negative feedback. An increasing number of perseverative errors in a healthy individual can be considered as a clear sign of cognitive fatigue.

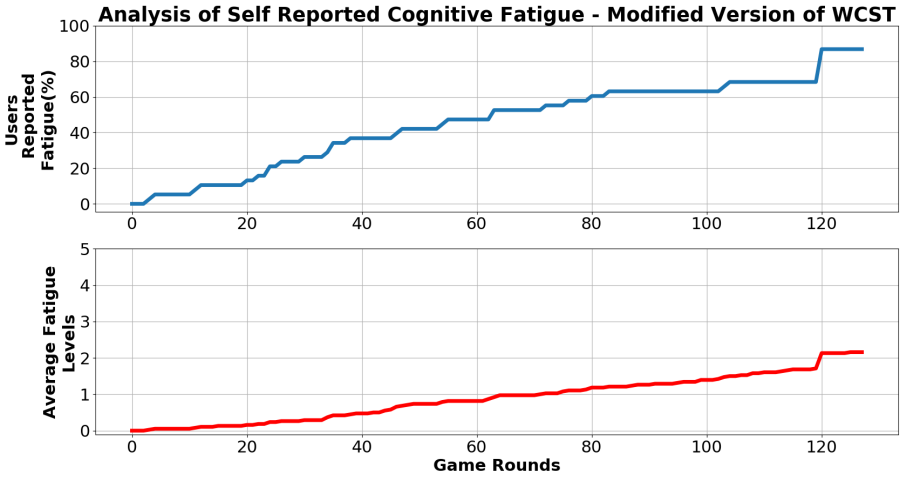
Figure 3.3b shows the analysis of user self report of fatigue during the game. The percentage of participants fatigued during the game is shown in the top plot. Out of the 38 sessions recorded, users reported experiencing at least some levels of CF by the end of the game in 35 sessions. The bottom plot in Figure 3.3b shows the average number of times the user reported fatigue which averaged to 2.2 times. As illustrated by these plots, it is clear that the users felt more fatigued toward the end of the game, reinforcing our hypothesis that the modified WCST does induce CF.

3.2.3 Machine Learning Analysis and Results

This study is exploratory, and there is still a long way to fully decipher and model the concept of CF and its impact on human performance. However, this work uses EEG data and subjective feedback from users to make cognitive fatigue predictions. More specifically, the focus was on identifying the presence of fatigue in a single round of WCST. For this analysis, all rounds of the three variations (original WCST, V1, and V2) were combined into a single data set, resulting in 76 sessions.



(a)



(b)

Figure 3.3: Analysis of the Self-reported CF and User Performance During game play [1]. (a) Average *P_Errors* during V1 and V2 WCST (b) Analysis of Self-Reported Cognitive Fatigue during V1 and V2 versions of WCST

Instances where the participant did not press the self-report button were considered "no fatigue," while the remainder were considered "fatigue" samples. The temporal relationship between successive rounds was not taken into account in the preliminary experiments.

3.2.3.1 EEG Feature Extraction

In order to represent EEG signals as a feature vector within a round, we extracted several spectral and time-domain features. These features are known for their ability to describe the core behaviors of 1D signals and are used extensively in other EEG classification tasks [66, 67]. In particular, the following six features were extracted for a given sequence of EEG measurements within a round of each cognitive game:

1. Mean Value
2. Standard Deviation
3. Maximum Value
4. Minimum Value
5. Spectral Centroid

$$C = \sum_{i=0}^{N-1} X_i p(X_i), \tag{3.2}$$

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

6. Spectral Rollof

$$R = 0.9 \sum_{i=0}^{N-1} |X_i|, \tag{3.3}$$

   where X is the spectrum of the signal and N is the size of the positive spectrum. Spectral Rollof corresponds to the frequency below which 90% of the magnitude distribution of the spectrum is concentrated.

Since the MUSE has a total of 4 electrodes, the final representation for each round was a vector of features of the size $4_{electrodes} \times 6 \frac{features}{electrode} = 24$ features.

While experimenting, other popular features such as zero-crossing rate, signal energy, spectral spread, the entropy of energy, and spectral entropy were used, but no significant improvements were seen in the classification results. In most cases, the rating performance fell by 5% to 8% of the average F1 score when additional features were added.

3.2.3.2 Classification Results

Different machine learning algorithms like Support Vector Machines (SVM), SVMs with an RBF kernel (SVMr), Random-Forests (RF), Extra-Trees (ET), and Gradient-Boosting (GB)that are commonly used in modeling similar problems were used [68]. A 10-fold cross-validation approach was used to evaluate the model using the dataset, split into a 20% testing set and 80% training set. In each fold, the distribution of the two classes (Fatigue and No-Fatigue) in the training and testing samples varied based on the total number of times users reported fatigue in the specific sessions. However, in all cases, the two classes were highly unbalanced towards the "No-Fatigue" class. To avoid overfitting and a bias toward one class in the models, the majority ("No-Fatigue") class was undersampled. The classifiers were then trained using a balanced dataset with the total number of samples for each class being equal to the available "Fatigue" samples in each fold. For testing, the original sample ratio was retained to have a realistic representation of the targeted problem.

Table 3.1 shows the results of the machine learning analysis. All the available feature streams (see Section 2.2.2) were used for an exhaustive grid search analysis to choose the best signal representation. The table presents only the signal which obtained the best results. The results indicate that the best results were using the feature streams related to the beta 13–32 Hz (b) and gamma 32–100 Hz (g) wavelengths and, in particular, their absolute (A) and relative (R) values. This result is

Table 3.1: Average classification results in all folds for different classifiers [1]. The column S indicates the EEG feature stream that achieved the best results after a comprehensive search in the grid. In the last row, the best results are obtained by combining the predictions of all trained models. Values in bold correspond to the methods that gave the best and most stable results. The abbreviations for Table 3.1 are as follows: Cl: Classifier, S: Signal, Pr: Precision, Rc: Recall and Ac: Accuracy.

Cl	**S**	**Rc**		**Pr**		**F1**		**Avg F1**	**Ac**
		NF	**F**	**NF**	**F**	**NF**	**F**		
SVM	gA	0.6	0.7	0.83	0.43	0.7	0.53	0.61	0.63
SVMr	gA	0.58	0.65	0.8	0.40	0.67	0.49	0.58	0.6
RF	bA	0.75	0.46	0.7	0.51	0.72	0.48	0.60	0.64
ET	dS	0.58	0.62	0.72	0.47	0.64	0.53	0.59	0.6
GB	bR	0.59	0.64	0.74	0.40	0.66	0.54	0.60	0.61
combined		0.72	0.56	0.79	0.46	0.75	0.51	**0.63**	**0.67**

in line with previous research that suggests that beta and gamma waves are highly related to mental states such as alert, normal alert consciousness, active thinking, and problem-solving [69].

3.3 Conclusion and Discussion

As confirmed by the study discussed in this chapter, physiological signals can model implicit user feedback while interacting with a system. This chapter proposed a novel system to build a statistical model for cognitive fatigue detection in healthy individuals. The system design is very unique compared to the state-of-the-art setups discussed in Section 3.2, which comprises physiological data acquisition for objective modeling of CF and collecting subjective user feedback of fatigue while playing the game. The subjective user feedback is also incorporated into the data set as labels for modeling (see Section 3.2.1.1). The results also provide confidence on off-the-shelf sensors such as the MUSE EEG headset to retrieve learnable patterns in data. This

is a beneficial finding because traditional EEG sensors can be cumbersome to set up and collect data. It is also not practical to use in a real-world setup.

Additionally, the data collection platform used in this study and the collected data are publicly available[1] to advance research in the domain and to tackle the problem of limited data availability for researchers. However, it should be noted that the total number of subjects offered by the dataset is relatively small to verify critical observations related to CF. One can also argue that the low accuracy and F1 values (Section 3.2.3.2) reported in the study could be because of the small dataset size. Nevertheless, this problem is prevalent in human-centric studies where getting approved participants for the study is difficult. Therefore, there is a need to explore other user data modeling techniques such as a multi-modal approach to see if we can incorporate various information of the same event to improve model performance and robustness. We must also look into how such user modeling techniques can be incorporated in an end-to-end system to assess user's performance. The following chapter addresses these issues by using studies that examine human factors like physical, behavior, task-specific engagement, and other factors that impact performance. A multi-modal approach to model user's task-specific attention and engagement to predict performance is also discussed.

[1] https://github.com/MikeMpapa/CogBeacon-MultiModal_Dataset_for_Cognitive_Fatigue

CHAPTER 4

Human Factors Analysis for Task Performance Prediction

4.1 Introduction

Affective robotics is a branch of robotics that falls under the realm of HRC that uses a human operator's psychophysiological measurements to improve human interaction by measuring his/her human factors [70]. Such robots consider the user's affective state, which can be measured physiologically or visually as a factor in the design process. In HRI, a robot often supports the performance of a human partner or provides feedback on the execution of a task. In such an interaction scenario, the robot system must perceive the cognitive state of the human teammate, which can influence the task performance. While human cognition is a critical human factor that affects several essential mental capabilities [24], it is important to note that the impairment in cognition can result from several reasons.

As discussed in Section 1.2.1, any task performed by humans uses several cognitive abilities to achieve a task's goal. Continuous usage of these abilities could lead to an increase in cognitive load. Overuse of these abilities may lead to cognitive fatigue, a debilitating symptom for people suffering from neurological disorders, a known risk factor in motor vehicle, and workplace accidents [71]. Cognitive fatigue is the primary reason most people suffering from neurological disorders such as Traumatic Brain Injury [53], and Multiple Sclerosis [52] often tend to avoid any social interactions and lose jobs or stay unemployed very long. A qualitative study by Aaronson et al. [72] describes fatigue in healthy individuals as a temporary state of exhaustion that manifests physically or emotionally and that it takes a toll on one's life roles,

which disrupts activity. Thus, it is essential to monitor such factors while working with an HRC set up and in day-to-day life when possible. This chapter discusses a multimodal solution to monitor cognitive human factors like task-based attention and engagement using physiological sensors and engagement through body postures using an RGB camera. We also present a quantitative study that shows other factors in our daily life like lack of sleep also influence cognitive task performance.

Assessing and monitoring cognitive abilities using sensors and intelligent games for cognitive training [46] and rehabilitation [73] have been addressed in recent research. While unimodal approaches to monitor the user's cognitive state during assessment have been proposed [74, 75], multimodal approaches show improved results assessing the user's mental state [3, 4]. To address the lack of attention and engagement of participants in monotonous and repetitive tasks, researchers have proposed robot-assisted therapy and training systems capable of increasing user engagement [76]. Similar research has used robot-assisted training systems to assess and adaptively train users in specific cognitive skills using task engagement as a vital factor [77]. Implicit and explicit cues from the users have been considered to estimate task engagement by utilizing several factors, including self-reports, facial expressions, and task behavior. This section discusses how this research utilizes such factors to develop a multimodal cognitive assessment framework for cognitive assessment and discusses how a multimodal cognitive assessment system provides better task performance prediction accuracy within the robot-assisted training framework.

4.2 Predicting Task Performance using Implicit and Explicit Cues

The components discussed in section 2.2.3 should be considered to develop an effective cognitive assessment system. In this study [3, 4], a Multimodal Robot-assisted Assessment System (MARS) is proposed based on the framework proposed

by Tsiakas et al. [49]. The cognitive assessment framework, shown in Figure 4.1, comprises a socially assistive robot, the NAO, conducting the assessment, an expert GUI for supervisors, and the sequence learning cognitive assessment test. Multimodal data from sensors such as the MUSE EEG sensor (see Section 2.2.2.1) and an RGB web camera (see Section 2.2.2.3) are used to monitor the user's physiological and behavioral data during the assessment.

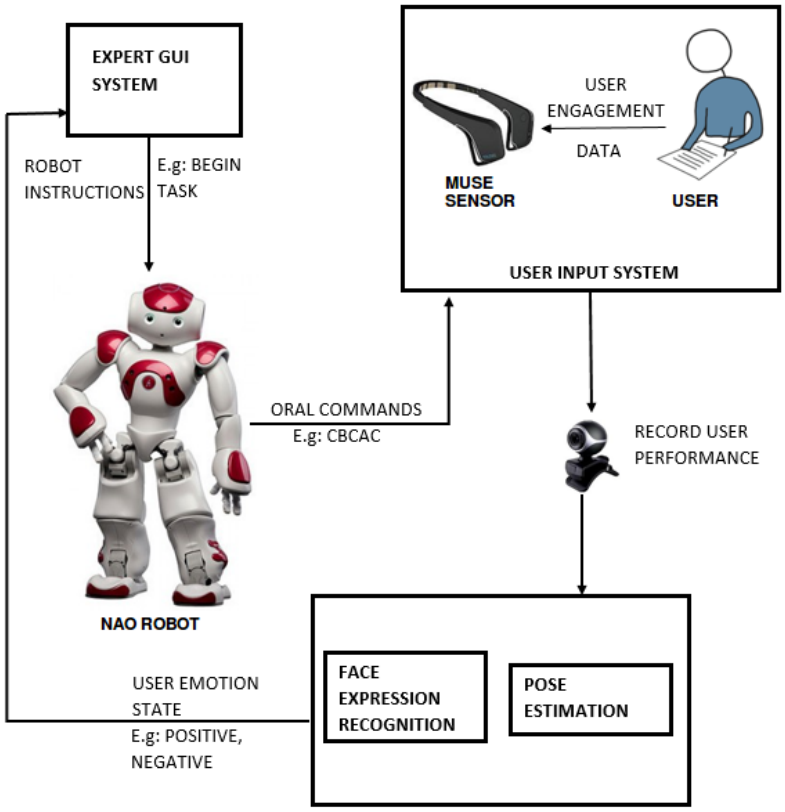


Figure 4.1: Experimental Setup [3]

4.2.1 The Sequence Learning Task

The proposed framework MARS uses the sequence learning task described in section 2.2.1.1. In this task, users were given sequences with three different levels of difficulty. Level 1 with sequence length 5, level 2 with sequence length 7, and level 3 with sequence length 9. For example, the robot generates the sequence $[c, a, c, c, b, c, a]$ at random as a sequence of level 2. These levels are chosen empirically to increase the user's cognitive load. Any lesser sequence length, the participants felt it was easy. Any longer sequence length, the participants felt it was tough and almost always failed.

The robot-assisted task was designed to provide audio feedback to the participant based on their responses. In order to capture the participant's natural facial expression and body pose changes to positive or negative feedback, two types of feedbacks were provided: positive and negative. The robot provides positive feedback like, "Great! Go ahead." or "Oh! you missed it, but go ahead," and negative feedback like "Maybe it was too easy" or "You do not seem to be paying attention." The robots' feedback helped create facial expressions and posture changes that we could capture with the RGB webcam.

4.2.2 The Expert GUI System

Based on the results of our previous study [48] presented in section 2.2.4, we developed an expert GUI, which is shown in Figure 4.2. With this GUI, a supervisor can manage tasks and view performance metrics such as user name, current sequence, user response, assessment history, and engagement score calculated using EEG data from the MUSE sensor (see Section 2.2.2.1).

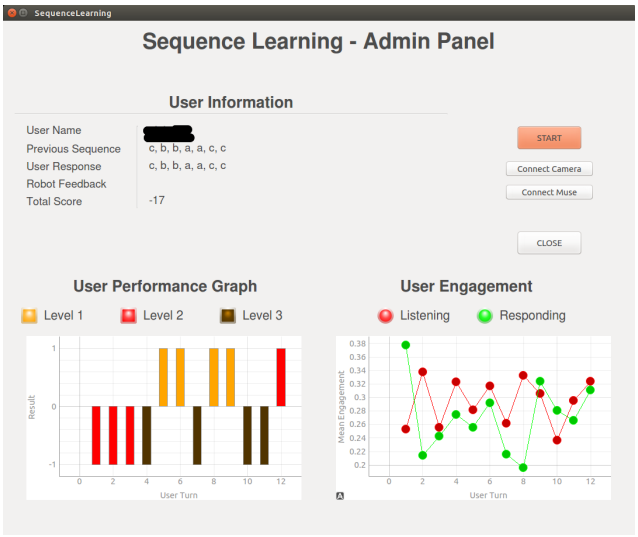


Figure 4.2: The Expert User-Interface [3]. The GUI displays performance metrics such as current sequence, user response, and engagement value.

4.2.3 User Performance and Engagement

Performance metrics, such as total score, current response, and user performance graph, are displayed on the GUI. The user performance graph displays the outcome (success = +1, failure = -1) of the round color-coded based on the difficulty level, as shown in Figure 4.2.

For the user performance graph, success is considered as +1, and failure is considered -1 and plotted for each turn. The total score is computed based on the score the participant achieved in each round. In level one, the participant gets +2 for a correct response or -2 for a wrong response. In level two, the participant gets +3 or -3 and +4 or -4 for level three.

For the user engagement plot, the engagement value is computed for each round using the alpha, beta, and gamma bands extracted using the MUSE headband at a sampling rate of 20 Hz. Data from the four EEG sensors on the headband as shown in Figure 2.4 (Section 2.2.2.1) are averaged to compute the mean engagement value

for each round using the formula $beta/alpha+theta$ [78]. During each round, the user had to listen to the stimulus first and then respond by pressing the buttons in front of them. The engagement value was computed during both these sessions separately, as shown in Figure 4.2.

4.3 Convolutional Neural Networks to Predict Task Performance

EEG sensor and RGB camera data were used to predict the task performance as success or failure. Convolutional Neural Networks (CNN) were used to build prediction models for individual modalities such as physiological data from EEG, facial expression, and body postures from the camera data. The output of these networks are then fused using a late-fusion technique to predict task performance results. The following sections will explain the individual networks and the data fusion technique briefly.

4.3.1 Task Outcome Prediction using EEG

The first modality is the EEG signals collected from the MUSE headband, which is used to predict task outcome. The MUSE headband collects five different bands of EEG signals (alpha, beta, gamma, delta, and theta) along with the raw EEG signal. Based on these signals, the features were extracted with a CNN. The NAO robot conducts the SL task (Section 4.2.1) and ground truth data for each sample is provided based on the success or failure in each round. Using this ground truth data and the extracted features, the network was trained to predict task performance. EEG signals are generally noisy. Hence, the noise was removed by applying an EWMA (Exponentially Weighted Moving Average) filter [79]. Figure 4.3a represents the architecture used to train and predict EEG signals.

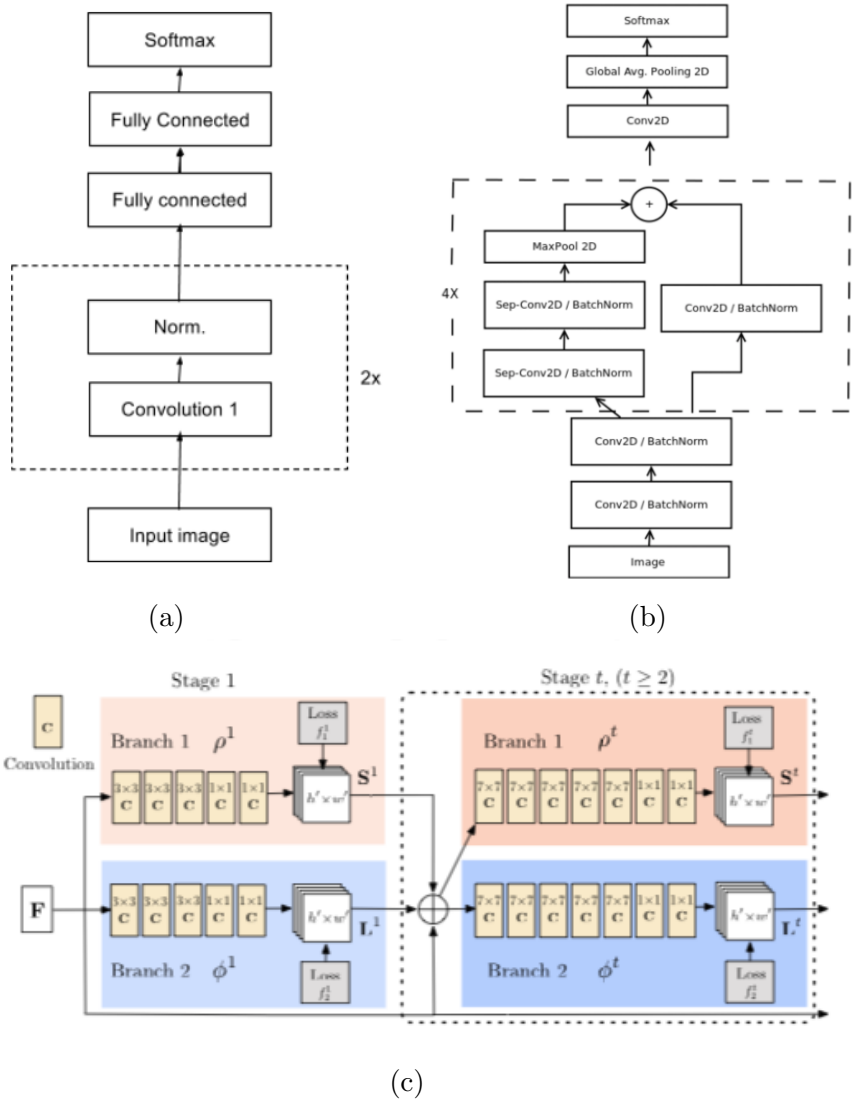


Figure 4.3: Neural Network Architecture for (a) EEG Signal from Neural Network (ENN), (b) Emotions from Facial Expression (EFE), (c) Emotions from Body Pose (EBP) taken from Ramesh Babu, A. and Rajavenkatanarayanan A. et al. [4]

The architecture, EEG signal from Neural Network (ENN), consists of two convolutional layers, each followed by a batchnorm operation with Rectified Linear Unit (ReLU) activation function, followed by two fully connected layers, and a softmax layer. The softmax layer predicts the probability of success or failure in a task, and the class that has the highest probability is considered the final prediction. The net-

work weights were initialized with Xavier initialization [80], and the Adam optimizer was used. The system was trained with 80% of the data and validated with 10% of the data. The remaining 10% was used to test the final multimodal network. A K-fold cross-validation process was performed ten times with different validation sets to check the consistency of the model in all samples. The proposed model produced 83% accuracy and 82% F1 score as explained in Section 4.3.3.

4.3.2 Task Outcome Prediction using Facial Expression and Body Postures

The second and the third modality used for task outcome prediction is facial expressions and body postures. To predict task performance as "success" or "failure", we follow a two-stage approach first to predict emotions from task performance first and then predict task performance. We define task performance as the outcome of a specific round of the task. The outcome, as mentioned earlier, can either be a success or failure. Two separate networks are created to train the data from facial expression and body postures. The following sections describe the individual networks first and then the data fusion algorithm used to predict task performance.

4.3.2.1 Emotion Prediction Using Facial Expressions

The first subsystem, Emotions from Facial Expressions (EFE), uses a CNN-based architecture proposed by Arriaga et al. [81] with minor modifications. The modified architecture is a CNN with four deeply separable residual folds; each convolution is followed by a batchnorm operation and a ReLU activation function. The last layer consists of a global average grouping and a softmax activation function, and the Adam optimizer was used for network optimization. This architecture has approximately 60,000 parameters representing the layer weights and the offset. The complete sequence consists of face recognition and an emotion classification module.

The emotion classification module extracts facial features such as eyes, eyebrows, mouth, and others to classify frames into three classes, positive, negative, or neutral. The network was trained and evaluated with the FER 2013 facial expression dataset [82]. For the entire input sequence, an array of emotion values are computed based on the above algorithm, which is then used for task outcome prediction. This subsystem produced an accuracy of 91% for the classification of the faces during testing using the FER 2013 dataset. Figure 4.3b shows the architecture for emotion recognition based on facial expression.

4.3.2.2 Emotion Prediction Using Body Postures

The second subsystem is the Emotions from Body Postures (EBP) module. Similar to EFE (Section 4.3.2.1), the pipeline for the emotion recognition from the body poses consists of two stages, body key-points detection, and emotion recognition. A deep-CNN was used to detect body key points [83, 84, 85]. The network's input is an RGB image of the format width x height x channels, and the output is a 2-dimensional (2D) location of the body keypoints. The feed-forward network predicts the 2D confidence of each keypoints and a set of 2D vector fields of part affinity fields (PAF) that holds the degree of association between the body parts. Using a greedy inference approach, the network predicts the final 2D keypoints using the confidence maps and affinity fields. In the network architecture shown in Figure 4.3c, which was initially proposed by Cao et al. [83], the confidence maps are predicted in the top branch, and the affinity fields are predicted in the bottom branch. The input image is first sent to a network (VGG-19 [86]) to generate a set of feature maps sent as input to the first phase of each branch. In the next phase, the predictions from the two branches are concatenated with the original features for the best results.

The second stage of the EBP module is the emotion recognition from the detected keypoints. The human body consists of several degrees of freedom and is capable of exhibiting emotions using different postures. Extensive research has been conducted in this field, and they indicate that humans exhibit different postures for different emotions such as anger, sadness, disgust, and others [87, 88, 89]. For example, Wallbott and Harald, in their paper [88], mention that humans tend to lean their bodies forward when angry and tend to bend their head down when sad. In our work, this critical information is utilized to predict emotions from the extracted body keypoints. Intermediate data such as the position of the hands, head tilt angle, body angles, and others were extracted from the body keypoints. The system then detects the postures and classifies them to one of the categories, i.e., positive, neutral, or negative, and produces an output of 1, 0, or -1, respectively. Similar to the EFE module, an array of output predictions of emotions are generated for each input sequence. Two hundred annotated images of different body postures were used to test the subsystem, producing an accuracy of 71%.

4.3.2.3 Data Fusion for Multimodal Task Outcome Prediction

Calculating task outcome prediction from EFE and EBP modules:

A simple algorithm to predict the task outcome from facial expressions and body postures was proposed for the second stage. The subsystems described in Sections 4.3.2.1 and 4.3.2.2 classifies the frames in the input sequence into one of the three classes, positive, negative, and neutral. Each network produces an array of emotion prediction for each input sequence which is then used as the input for the algorithm shown in Algorithm 1. In this algorithm, the total number of frames with positive, negative, and neutral emotions are calculated from the input sequence, and then the class and the confidence values are predicted. The subsystems EFE (Section 4.3.2.1)

Algorithm 1: Task Performance Prediction from Behavioral Data. Algorithm proposed in the work published by A.R. Babu, A. Rajavenkatanarayanan et al. [4]

Input: Frames of a sequence with emotions predicted, Threshold from training

Output: Individual modality prediction, classes(Success, Failure)

$neutral\_frames$ = total frames with neutral emotions in a sequence

$positive\_frames$ = total frames with positive emotions in a sequence

$negative\_frames$ = total frames with negative emotions in a sequence

$Total\_frames$ = total number of frames in a sequence

if *(number of predicted negative emotions* $\geq$ *Threshold)* **then**

$Prediction = Failure$;

$$Confidence = \frac{negative\_frames + neutral\_frames}{Total\_frames} \tag{4.1}$$

end

if *(number of predicted negative emotions* $<$ *Threshold)* **then**

$Prediction = Success$;

$$Confidence = \frac{positive\_frames + neutral\_frames}{Total\_frames} \tag{4.2}$$

end

and EBP (Section 4.3.2.2) were trained individually to find the optimal threshold cutoff to predict negative and positive emotions in order to maximize successful predictions effectively. The SL data was split into 80% training data and 10% testing data. The remaining 10% was used to test the final multimodal system. The system was tested ten times using a different split of the testing set to ensure the consistency of the model. The first subsystem EFE, produced a maximum accuracy of 75% while the EBP subsystem produced a maximum accuracy of 62.5% as discussed in Section 4.3.3.

Final Task Performance Prediction:

From the individual predictions of each of the three modalities, the EEG data, and image data from facial expressions and body postures, a combined decision is made

for the final task outcome prediction using the algorithm mentioned in Algorithm 2. Each of the individual modalities predicts the task performance outcome as success

Algorithm 2: Final Performance Prediction combined from three modalities. Algorithm proposed in the work published by A.R. Babu, A. Rajavenkatanarayanan et al. [4]

Input: Prediction output and confidence from 3 modalities, pred(EFE), pred(EBP), pred(ENN)

Output: Final Prediction, classes(Success, Failure)

if *(any of the predictions is Failure)* **then**

$$Confidence\_of\_that\_Modality \; * = -1; \tag{4.3}$$

end

$$Score = Confidence\_EFE + Confidence\_EBP + Confidence\_ENN \tag{4.4}$$

if *(Score is positive)* **then**
$FinalPrediction$ = Success;
else
$FinalPrediction$ = Failure
end

or failure. A confidence value is also computed, which is positive if the prediction is success and negative if the prediction is failure. With these predictions and the confidence values computed as part of each of the modalities, a score is computed in algorithm 2 as the sum of the confidence values as shown in Equation 4.4. If this score is still positive, the final prediction of the multimodal system will be "success". Else the prediction will be "failure."

4.3.3 Results

This chapter discussed how multimodal data could be used to predict task performance outcomes using emotions and engagement predicted from task-specific

interaction data. In Table 4.1, the results of the model developed for predicting task performance outcome using EEG, the ENN module, are shown. The results are compared with related research by Papakostas et al. [90] that uses similar EEG data from MUSE headband, collected from a Sequence Learning task to predict task performance outcome. In this work, the authors presented the results of task outcome prediction using several machine learning algorithms such as Support Vector Machines (SVM), Gradient Boosting (GB), Random Forests (RF), Extra Trees (ET). As shown in Table 4.1, our ENN model easily outperforms the best accuracy produced by the ET algorithm by 8% and the best F1 score produced by the GB algorithm by 13%. Table 4.2 shows the results of the individual modalities proposed in this chapter and

	SVM	**GB**	**RF**	**ET**	**ENN**
F1 Score	0.62	0.69	0.56	0.54	**0.82**
Accuracy	0.65	0.74	0.67	0.75	**0.83**

Table 4.1: Task outcome prediction from EEG signal as presented in [4]. Abbreviations: SVM-Support Vector Machines, GB-Gradient Boosting, RF-Random Forests, ET-Extra Trees

the combined multimodal approach. As mentioned earlier, the Emotions from Facial Expressions (EFE) module produces an accuracy of 75% and an F1 score of 73.8%. The Emotions from Body Pose (EBP) module produced an accuracy of 62.5% and an F1 score of 54%. The EEG from Neural Network module (ENN) produced an accuracy of 83% and an F1 score of 82%. These results show that a multimodal approach can outperform unimodal solutions. The multimodal approach also increases the robustness of the prediction, as shown in the higher F1 score compared to the unimodal solutions.

	EFE	**EBP**	**ENN**	**EFE+EBP+ENN**
F1 Score	0.738	0.540	0.820	**0.870**
Accuracy	0.75	0.625	0.83	**0.875**

Table 4.2: Prediction from individual modalities and combined as presented in [4]. Abbreviations: EFE-Emotion from facial Expression module, EBP-Emotions from body postures and ENN-EEG signal with Neural Network

4.4 Everyday Activities that Affect Task Performance

So far, this chapter discussed how to monitor implicit and explicit cues from a participant to predict task performance outcomes. The central underlying assumption is that the user's engagement and emotions directly affect performance. Despite the high accuracy and F1 score, as discussed in Section 1.2, there are several types of human factors, both physical and cognitive, that might affect performance in a system in the real-world. It is thus essential to monitor and assess these different factors as well. This section will take a brief look into a study by Rajavenkatanarayanan et al. [8] that considers one of the important activities of daily life and how it affects cognitive performance.

4.4.1 Sleep and its Effect on Cognitive Performance

It is a well known fact that a lack of good sleep or insomnia can cause an impact in the circadian rhythm and causes deterioration of cognitive performance [91]. However, it has been quite difficult to incorporate such metrics in to an interaction system such as the one discussed in Section 4.2.1. Advances in wearable technologies and smartwatches have paved the way for development of small sensors that can measure sleep quality. In this study, we take a look in to the correlation between the percentages of sleep in light, heavy, and Rapid Eye Movements (REM) cycle collected from the Fitbit smart watch [92] and performance in a cognitive task.

4.4.2 Experimental Study

This user study was conducted over five days with 30 participants, which consisted of 23 male and seven female participants. Over these five days, the participants were provided with a Fitbit smartwatch that recorded the participant's sleep patterns. The participants' cognitive performance was tested on two of these five days using the N-back task described in Section 2.2.1.3. This task was chosen because it assessed working memory and attention, which are essential cognitive abilities in everyday life. This study used two repetitions of the 0-back task and two repetitions of the 2-back task. In the 0-back task, the participant was presented with the target stimuli beforehand and pressed a button when they recognized it. On the other hand, in a 2-back task, the participant is shown a sequence of stimuli as shown in Figure 2.3. In this task, the user responds by pressing a button when they see the same shape, irrespective of color, precisely after two stimuli. The two repetetions within each type of task were played back to back. However, the order of each type of task was counterbalanced to avoid order effect among participants. The study lasted 45-minutes for each participant. During the task, data was recorded from the MUSE EEG headband (see Section 2.2.2.1), and task performance metrics such as errors, score, response time, and reaction time were recorded.

4.4.3 Preliminary Results

The objective of this study is to show a correlation between sleep pattern and task performance. Sleep quality data like total duration of sleep, percent of deep, light, and REM sleep, and awake time were extracted from the Fitbit data recording. Specifically, sleep data of the night prior to cognitive assessment was extracted for analysis. Before computing the correlation between sleep quality and task perfor-

Table 4.3: Summary of correlation analysis between stages of sleep and task performance. (τ/ρ) denotes the degree of correlation while P denotes if the correlation is significant. Initially presented by A. Rajavenkatanarayanan et al. in [8]

	Kendall		**Spearman**	
Variable	τ	**P**	ρ	**P**
total	-0.0483	0.760	-0.0261	0.901
%deep	0.0933	0.541	0.1169	0.578
%light	**-0.3519**	**0.018**	**-0.4593**	**0.021**
%rem	**0.3409**	**0.021**	**0.4689**	**0.018**
%awake	0.0104	0.962	0.0033	0.988

mance, to determine the normality of the data collected, D'Agostino's K^2 test [93] was performed, which suggested that the task performance metric did not follow a normal distribution. Following this result, Kendall and Spearman correlation analysis was performed between the three sleep quality features and performance metric, the average score in each assessment. Preliminary results, as shown in Table 4.3 indicated that the light sleep cycle had a moderate negative correlation with task performance, while the REM sleep cycle has a moderate positive correlation with high confidence. According to a blog post by Fitbit[94], on the types of the sleep cycle, research indicates that the light sleep cycle is responsible for processing memory, emotions, and metabolism regulation. On the other hand, the REM sleep cycle is responsible for emotion regulation, memory, and protein synthesis. It is thus clear that sleep affects memory, and the results presented indicate that the quality of sleep significantly impacts task performance.

4.5 Conclusion and Discussion

In summary, this chapter discussed the MARS - Multimodal Robot-assisted Assessment System framework to predict task performance outcome using multimodal data recorded during interaction in a cognitive assessment task. The Sequence Learn-

ing task is utilized that assesses the learning ability, working memory, and attention. Data recorded includes EEG from the Muse headband, facial expressions, and body pose from an RGB camera. For each of these data modalities, a separate network was trained. The EEG from Neural Network (ENN) module predicted task performance outcomes using the EEG data. The Emotions from Facial Expressions (EFE) and the Emotions from Body Posture (EBP) module classified emotions into the three classes positive, negative, and neutral. The classified emotions from these two modules were then used to compute a final task outcome as success or failure along with a confidence score. This confidence score and the prediction from all three modalities are used to make a final prediction using a late fusion approach. The results presented in Section 4.3.3 indicate that a multimodal approach outperforms the unimodal and traditional machine learning approaches and is robust. While such intelligent systems provide state-of-the-art performances, it is also important to consider other indirect factors that may affect task performance. Section 4.4 discuses one such factor, sleep. This longitudinal study over five days shows a correlation between sleep and cognitive task performance using the N-back task, where a relationship between poor sleep patterns and deteriorating cognitive performance is shown. Sleep is a well know human factor that affects our ability to function at our full capability. Further research is required in this field to incorporate improper sleep patterns and other similar factors into the adaptive system for a better understanding of human performance.

The dissertation so far discussed various frameworks developed in this research to monitor and predict different human factors during interaction in a specific cognitive assessment system. While these assessment systems take into account several cognitive factors that are relevant in an industrial scenario, one of the major points missing is how these results can be applied to the real world. Do the same assumptions hold? Can we still use the same sensors and data collection procedure to build an

effective HRC system? To address these questions and explore several practical limitations in implementing a real-world setup, the following chapter in this dissertation introduces an HRC framework that simulates a real-world task.

CHAPTER 5

CogniSmart: An Intelligent Human-Factors Monitoring Framework to Enhance Human-Robot Collaboration

5.1 Introduction

The role of robots in our lives has increased from just industrial robots to personal assistants. Several collaborative robots are now prevalent that help industrial workers [95, 96], people with disabilities in their daily activities [97], and children and elderly people with entertainment and providing company [98, 99]. There are also some collaborative robots that assist in rehabilitation [100]. With the advent of Industry 5.0 [10], several industrial applications are now introducing smaller collaborative robots rather than their large and costly counterparts. These emerging types of robotic devices enable companies to enrich their workflows with a robot's precision and a human worker's creativity to increase productivity. Research in this domain is now at its peak, where people are focusing on achieving a safe HRC setup. Such researches enable the adaptation of workflows that are ergonomic and do not threaten or harm the physical health of the human in the loop [101, 102].

Despite the abundance of research in this domain, there is still limited understanding of the psychological impact on the human partner collaborating with robots daily. It is essential to monitor cognitive human factors in an HRC setup so that the human partner feels safe and comfortable while working and ensures maximum efficiency and productivity. Cognitive Load (CL) is one of the critical human factors that affect an operator's performance in an assembly line [103]. Arousal, time limitations, and task difficulty are some of the factors that affect CL [104]. It is

thus important to monitor CL which affects performance times and thereby affects productivity. To this end, we propose CogniSmart, a framework to monitor CL and enhance HRC. In this research [5], the participant's extraneous CL, which can be defined as the amount of energy expended to achieve a specific goal, is predicted using the CogniSmart framework. Different types of physiological sensors are used in a novel robot-assisted assembly task to capture the user's cognitive state during interaction. Wearable sensors such as ECG and EDA are used because of their less obtrusive nature, and success in modeling user state [105, 8]. Some of the practical limitations and questions about implementing an effective HRC system and how this research addresses it are discussed in this chapter.

5.2 Background: Human Factors Modeling for Safe Human-Robot Interaction

Modeling human factors is not a new research area in behavioral modeling. Several research works have addressed this problem by adopting various methodologies under various assumptions. Despite its prevalence in research, very few datasets are available for modeling CL and the available datasets are very small in sample size. In 2018, Nelles et al. [106] published a review of the evaluation metrics used to assess human wellbeing in an HRI setup. The review highlights the heterogenic nature of the experimental design, qualitative surveys, and other measures. This heterogeneity could also be one reason for the limited availability of data for building a usable real-time system. The research works reviewed by Nelles et al. use surveys from users to evaluate the HRI design based on trust [107, 108, 109], usability [110, 111], safety [112], cognitive and physical workload [113], and well-being [114]. One of the major drawbacks of self-reported surveys is their subjective nature. This drawback makes it very difficult to design an objective assessment system that is usable in a real-time application.

Another important component of the proposed framework is the physiological sensors that help monitor CL during HRC. In a 2010 study by Novak et al. [115], different sensors such as ECG for heart rate measurement, EDA for skin conductivity, thermistor flow sensor for respiratory rate, and a digital temperature sensor for skin temperature were used to quantitatively evaluate the usability of the sensors to estimate cognitive workload in haptic interaction. The study showed that the mean and variability of the respiratory rate and skin temperature showed a significant difference between the difficulty levels, despite varying physical load levels. The study highlighted that physiological sensors are able to measure CL in an interaction task. However, the sensors used in the study such as the thermistor flow and skin temperature sensors can not be used in an actual industry setting because of the coplicated setup. Weistroffer et al. [116] used sensors such as Photoplethysmogram (PPG) and EDA sensors to measure heart rate and skin conductivity during HRC. Despite not being able to record data during interaction, the authors were able to compare the usability of a virtual and physical system using pre/post measurements. Recently, Villani et al. [117] proposed a CL assessment framework for HRI, which was used to control a mobile robot. In this framework, a smartwatch was used to extract the user's heart rate variability for detecting the rest and stress conditions. In combination with the accelerometer, gyroscope, and magnetometer on the smartwatch, an app was designed to provide commands to control the mobile robot. This framework was later evaluated in a study by Landi et al. [118] in a teleoperation task where virtual fixtures are used to operate a remote robot. Despite the usage of an unobtrusive sensor such as the smartwatch, these studies [117, 118] do not discuss the accuracy of the model that predicts CL.

5.3 Proposed Framework

Taking all the results and limitations of these studies into account, we propose CogniSmart, an intelligent framework (Figure 5.1) that aims to provide a reliable human factors monitoring system using as few unobtrusive sensors as possible to increase convenience. In contrast to the related works presented in Section 5.2, a data-driven approach is used to design machine-learning-based models of physiological sensors for identifying the cognitive human factor, Cognitive Load. A collaborative assembly task is also proposed, which simulates a real-world setup in the industry. In this framework, a user performs the collaborative robot-assisted assembly task

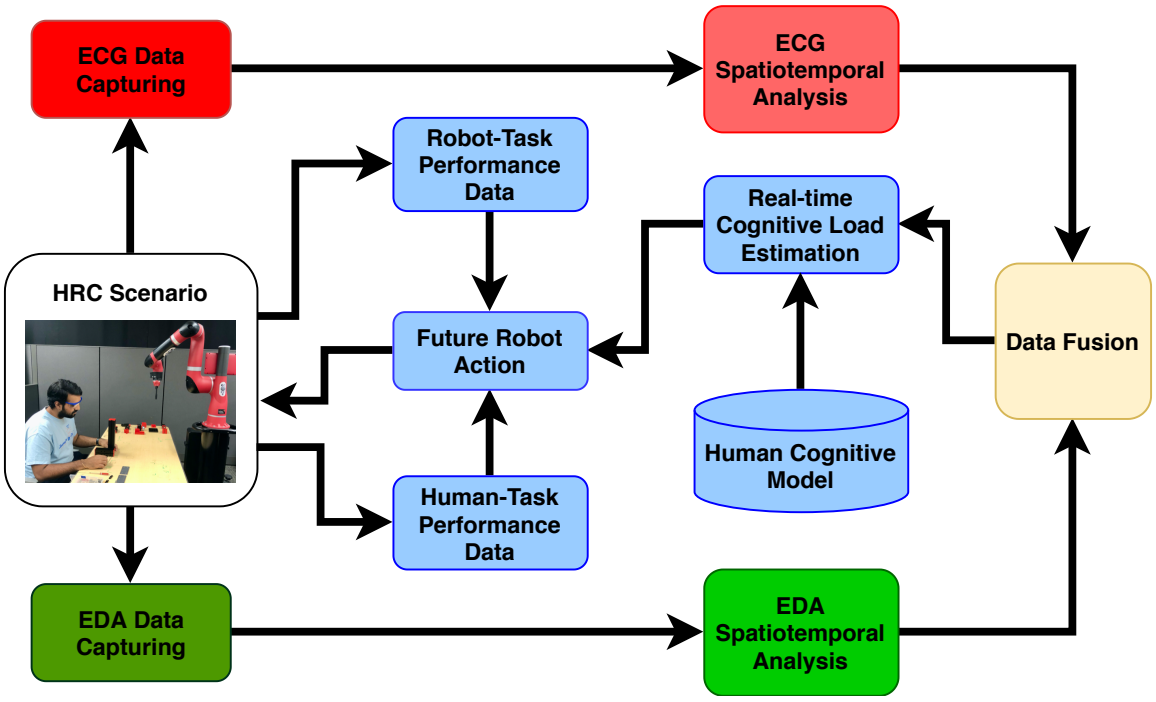


Figure 5.1: Overview of the proposed *CogniSmart* system architecture for HRC

during which different physiological sensors are used to record the user's affective state. This data is then used to predict CL, which helps decide the robot's future

actions like operation speed, the time interval between each assembly, and many others. The main focus of this chapter will be on signal processing and user modeling for CL prediction.

5.3.1 The Robot-Assisted Assembly Task

As shown in Figure 5.2, a robot-assisted assembly task was designed to assemble a small sanding machine. In this task, a collaborative robotic arm, Sawyer, developed by Rethink Robotics [119] was used to assist the participant in assembly. The motivation behind the task's design was to simulate a real-world assembly setup, where a robot and a human partner collaborate synchronously to achieve a common goal. In this task, referred to as *RoboAssist* from now on, the parts required for assembly of the sanding machine are split into two sets; a set large enough for the robot to handle (set 1), and a set very small that the robot is not able to handle accurately (set 2). For

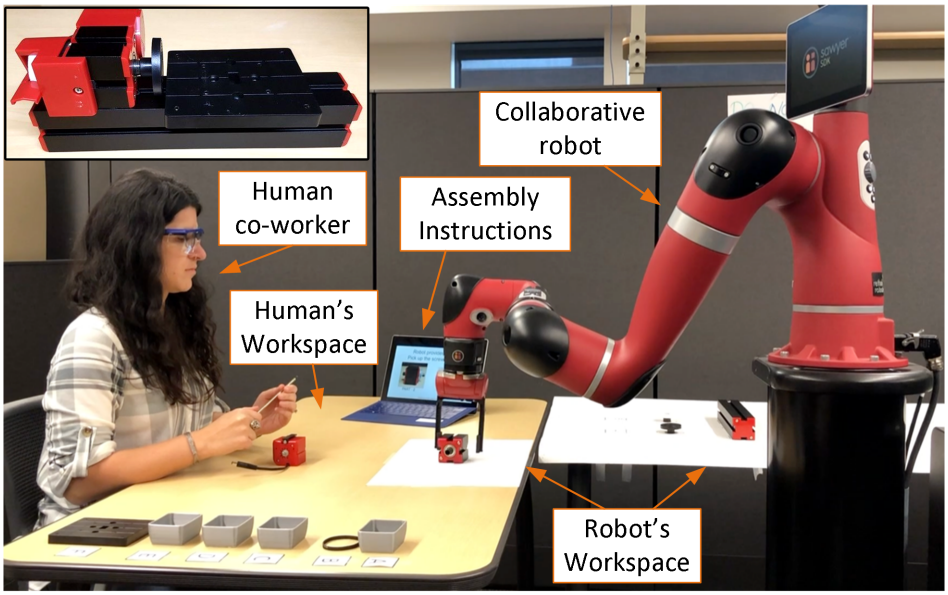


Figure 5.2: Setup of the Proposed Collaborative Assembly Scenario, *RoboAssist*. Top-left image: Final Assembly Product - A Miniature Sanding Machine. [5] ©2020 IEEE.

instance, as shown in Figure 5.2, the set 1 parts of the sanding machine were placed to the right side of the robot, while set 2 parts like a screwdriver, nuts, and bolts were placed in a separate bin on the right side of the user. The robot picks parts from set 1 and provides them to the user as the user uses the parts from set 2 for assembly. In addition to this, a simple user interface was designed to provide detailed step-by-step instructions to the user to assemble the sanding machine. Each instruction provides detailed information regarding the parts the robot handles and the parts the user needs to handle to complete an assembly step. Once each step was completed, the user was required to press a button to proceed to the next step. The user interface was integrated to the robot using the Robot Operating System (ROS) [120] in order to recognize a step completion and move on to the next step. A video of the assembly task can be found in the following link: `https://youtu.be/m_dkLHflCUo`.

5.3.2 Sensors and Data Stored

One of the main goals of this research is to explore sensors that are easy to use and unobtrusive. In the previous chapters, EEG sensor was used predominantly to monitor cognitive human factors for assessment. While the Muse EEG sensor provided promising results in cognitive fatigue prediction and engagement computation for task performance prediction, such a sensor may not be helpful in an industry setting. In an industrial setup, there are safety requirements expected of the employer. According to the Occupational Safety and Health Act of 1970 [121], employers are expected to provide protective equipments like goggles and helmets, making it difficult to use an EEG sensor. Moreover, the operator may not prefer additional sensors that create more distractions to the working environment. It is thus essential to use unobtrusive sensors that are easy and convenient to use. Previous research [105] has shown the usability of biosignals from ECG and EDA sensors in different stress de-

tection setups. In this research, ECG and EDA sensors are explored for cognitive modeling as they are easy to use and are usually not in the operator's way. Recent advancements in sensor technologies have also incorporated these sensors in easily wearable hand-worn devices like smartwatches or wearable smart T-shirts. The following sections will discuss in detail the different positions of the sensors used for data collection and features extracted for training a machine learning model.

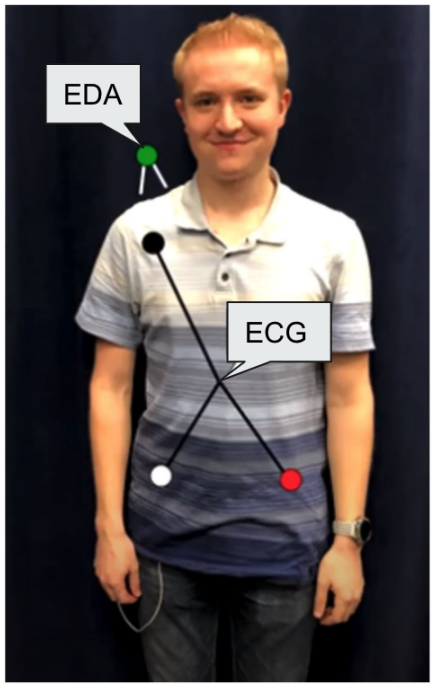


Figure 5.3: Sensor Placement of ECG and EDA sensors. The EDA sensor is placed on the right shoulder. The ECG sensor is placed in a Lead II setup of the Einthoven's triangle.

5.3.2.1 ECG Sensor - Placement and Feature Extraction

An Electrocardiogram (ECG) is an easy and painless way of measuring a person's heart activity by recording the heart's electrical signals. In medicine, it is commonly used to monitor the heart's function and detect any problem with the

heart. Advances in sensor technology have enabled the easy availability and accessibility of ECG sensors for researchers and the general public. In recent times, ECG sensors are also incorporated into commercially available smartwatches like the Apple Watch [122]. In this study, the participant's heart activity during interaction is recorded using a single-lead ECG sensor developed by BioSignalsplux (Section 2.2.2) at a sampling rate of 1000 Hz.

Sensor Placement: For this study, a lead II configuration of the standard 3-point bipolar limb leads configuration of the Einthoven's triangle was adapted [41]. Typically, in this configuration, 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. However, the RoboAssist setup requires the participant to sit and work on an assembly task. To facilitate this requirement and ensure ease of use by participants, the electrodes were placed on the right shoulder and the lower torso as shown in Figure 5.3.

Feature Extration: Due to the position of the ECG electrodes on the right shoulder and lower torso, the ECG signal is inverted compared to the actual data captured using the standard lead II configuration. Therefore, the signal is first inverted to represent the original waveform before preprocessing. Time and frequency domain features from the ECG signal are extracted from the QRS complex (Figure 5.4) and the RR interval (the time elapsed between two consecutive R waves). These features are commonly used in Heart Rate Variability (HRV) analysis in mental stress assessment studies [123]. The QRS complex forms the main component of the ECG signal that represents the electrical activation in the sensor due to the contraction of ventricles in the heart. It is the prominent peak in the ECG waveform and helps calculate heart rate and various heart-disease states [124, 125]. The peak detection algorithm developed by Van Gent et al. [126] is used to identify the R peaks. A notch

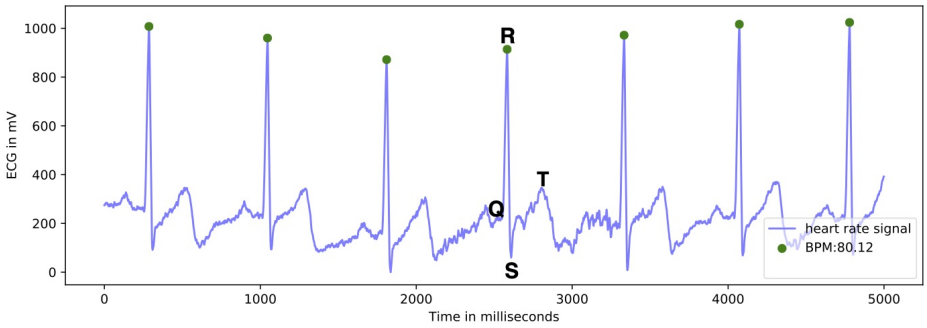


Figure 5.4: Sample ECG signal acquired from the Biosignalsplux sensor. The green dots indicate the peak detected using which the heart rate of the signal was estimated. Q, R, S, T indicate the Q-wave, R-wave, S-wave, and T-wave component of the ECG signal. ©2020 IEEE.

filter with a cut-off frequency threshold value of 0.05 Hz was empirically selected to be applied before the peak detection algorithm. This step helps in improving the peak detection accuracy due to additional noise by minimizing the T-wave and other unwanted low-frequency noise. After this step, time-domain and frequency-domain features are extracted from the preprocessed signal. Boonnithi et al. [123] proposed the use of some time domain features for HRV analysis such as, 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 ($\widetilde{RR}$), the range of the RR Interval (rRR), and median absolute deviation of RR intervals (MAD). A list of all extracted time domain features are listed in Table 5.1 with their formulas.

Table 5.1: Time Domain Feature Extraction from ECG Data [5]. ©2020 IEEE.

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	$pRR20 = \frac{Count(\lvert RR_{i+1} - RR_i \rvert)_{>20ms} \times 100}{N-1}$
Proportion of successive differences above 50 ms in percentage	$pRR50 = \frac{Count(\lvert RR_{i+1} - RR_i \rvert)_{>50ms} \times 100}{N-1}$
Median absolute deviation of RR intervals	$MAD = median\left(RR_i - \widetilde{RR}\right)$

Additionally, the frequency-domain features extracted from the ECG signal were Power Spectrum of Very Low Freq (LF) amd Power Spectrum of Very High Freq (HF). These powers have been largely used in the literature by associating them to autonomic nervous system activities (i.e., LF is associated with sympathetic activity and HF to parasympathetic). Other features extracted from these basic measures

are the Sympathetic modulation index, the Vagal modulation index, and the Symphatovagal balance index. A list of all the extracted frequency domain features is listed in Table 5.2 with their formulas. In total, for ECG signal, 17 features were extracted for machine learning analysis.

Table 5.2: Frequency Domain Feature Extraction from ECG Data [5]. ©2020 IEEE.

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
Symphathetic modulation index (SMI)	*SMI = LF / (LF+HF)*
Vagal modulation index (VMI)	*VMI = HF / (LF+HF)*
Symphatovagal balance index (SVI)	*SVI = LF / HF*

5.3.2.2 EDA Sensor - Placement and Feature Extraction

Electrodermal activity (EDA) is a measure of the change in electrical potential between different parts of the skin [127]. These changes are caused by alterations in sweat secretion and sweat gland activity due to changing sympathetic nervous system activity. Several studies [128, 42] have used EDA sensors to detect CL, stress, and other human factors. Recent research [129] has developed unobtrusive wearable solutions for using EDA sensors for long-term use. In this dissertation, the EDA sensors developed by BioSignalsplux (Section 2.2.2) is used for monitoring cognitive human factors, which is capable of accurately measuring the electrical properties of the skin. *Sensor Placement:* In a 2017 study by Zangróniz et al. [43], locations like the palm and the soles are suggested to be the best spots for EDA data recording because sweat glands are most active in these locations. However, due to RoboAssist's design, participants could not use their palms or soles for sensor placement. Hence, in

this study, data from the EDA sensor was collected from the shoulder, which was proven to be one of the best alternate locations for measuring skin conductance [130]. More specifically, the sensor was placed on the right shoulder of all the participants as shown in Figure 5.3.

Feature Extraction: Data collected from the EDA sensor is downsampled to 200 Hz to reduce computation. The downsampled data is then filtered using a Butterworth filter to remove high-frequency noise using methodology proposed by Bizzego et al. [131]. Research indicates that EDA signals comprise two different superimposed components; the phasic or the skin conductance response (SCR) and the tonic or the skin conductance levels (SCL) [43, 132]. 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. On the other hand, the tonic component is the baseline level of skin conductance, which varies from person to person [132]. Figure 5.5 shows a sample snapshot of the EDA signal with its phasic and tonic components.

The shape of the EDA signal is vital 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.

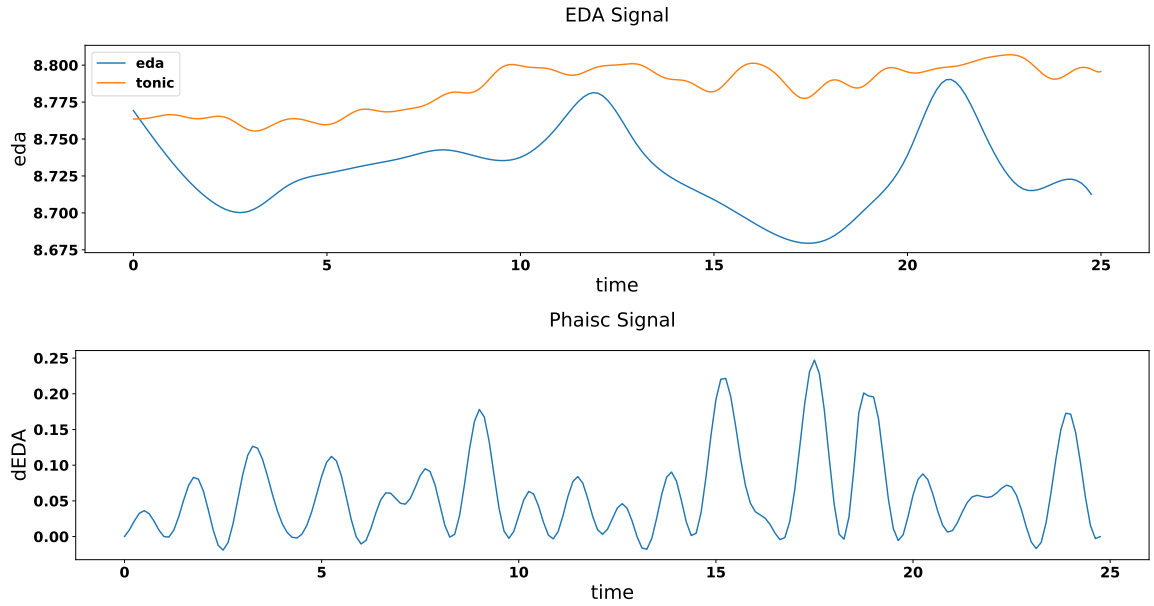


Figure 5.5: Downsampled EDA signal acquired from the Biosignalsplux sensor. The top plot shows the preprocessed EDA signal in blue and the tonic component of the signal in orange. The bottom plot shows the respective phasic component in blue.

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

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

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:

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

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:

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

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

The entrophy 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 1D signals like EEG [1], and speech [133]. In total, for EDA signal, 16 features were extracted for machine learning analysis.

5.4 User Study - Preliminary Analysis

A user study, approved by the Institutional Review Board at The University of Texas at Arlington, was conducted to validate the framework and demonstrate the feasibility of the RoboAssist task. Twenty-five participants enrolled for the study, of which 15 were male, and 10 were female participants. The participant's age range was 19 to 30, except for one who was in the 31 to 40 age range. The study lasted

for 40 minutes for each participant. The data collection for the study took place at the "Heracleia Human-Centered Computing Lab" over two sessions, one following the other with a short 3-minute break. In the first session (*S1*), the participants worked with RoboAssist, where the system provided detailed step-by-step instructions for each step. This session functioned as a practice round for the participants to get familiarized with the assembly setup. For the second session (*S2*), the participants were expected to remember the assembly steps from *S1*. Moreover, in *S2*, the participants were provided only 30 seconds per assembly step. This restriction was incorporated to induce stress and high CL.

During the study, once the experimental procedure is explained and the participants signed the consent form, a short baseline survey to assess the participant's baseline cognitive and physical state was conducted. Following the baseline survey, session *S1* was conducted, which enabled the participants to get familiarized with the setup and memorize the assembly instructions. A post-task survey to assess the user's subjective experience was conducted after *S1*. Next, session *S2* was conducted where the participant's stress to complete in time and CL was increased due to the timer for each step. After this session, a final post-task survey was conducted to assess the user's cognitive and physical state after the second session. During each session, data were recorded from ECG and EDA sensors. Other task-specific metrics like completion time and errors were also recorded. A 3-minute break was included between the sessions to avoid the cumulative effect of stress.

Time is a critical factor in an HRC assembly and production line where the robot and the human operator's moves must be in sync to ensure a successful assembly workflow. To get the human operator to work synchronously, companies often train their workers before working as a robot operator. This means that the practice effect [134] is always present in a real-world operator. It is thus important to incorporate this

factor into the modeling process. In order to simulate an actual assembly operator's work, in this study, all participants first worked on a training session (*S1*) before working on a timed session (*S2*) that simulates a real assembly setup.

5.4.1 Preliminary Results from Proposed Framework

5.4.1.1 Survey Results

Participants responded to three surveys in total. The first survey was "Baseline," which was collected before the start of the session. The second survey was "post_task1," which was conducted right after *S1*. The third survey was "post_task2," which was collected after *S2*. During these surveys, different questions were asked to assess the participant's cognitive and physical state, including some task-specific questions. For each question, the participants were asked to rate the response to each question from 0, meaning very low, to 10, meaning very high. The questions asked in the baseline survey focused on assessing the participant's physical and cognitive state before the start of the task. The questions were:

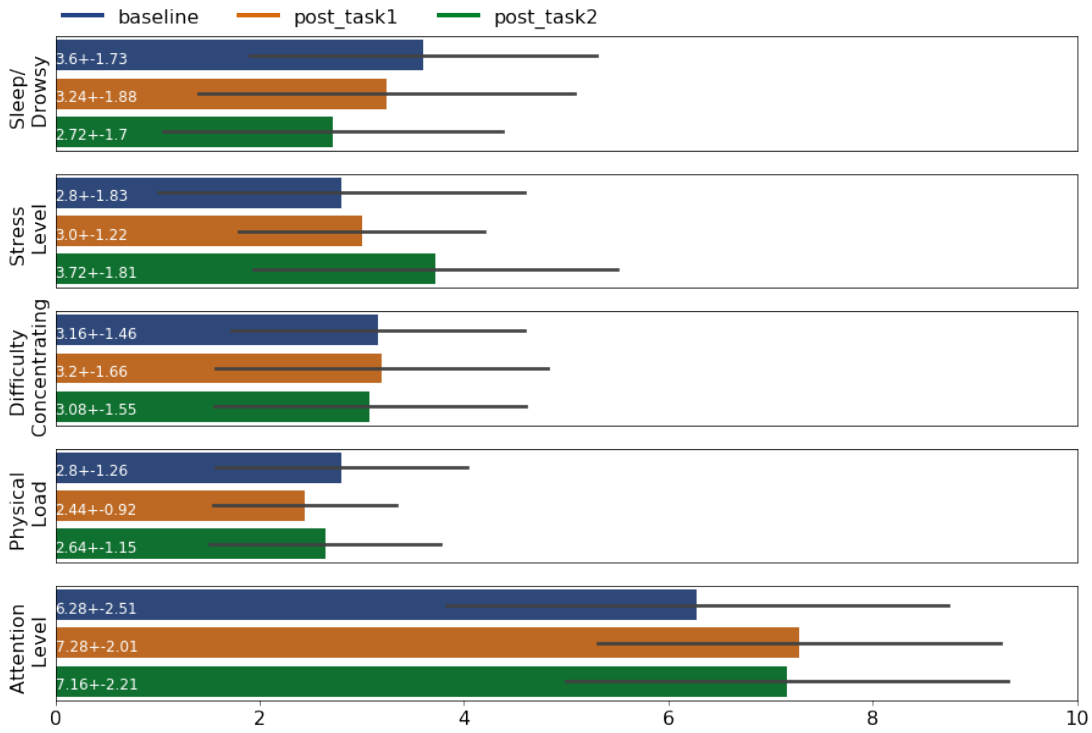
- On a scale of 1 to 10 how sleepy or drowsy do you feel at the moment?
- On a scale of 1 to 10 do you have difficulties concentrating at the moment?
- On a scale of 1 to 10 do you feel physically tired at the moment?
- On a scale of 1 to 10 how distressed do you feel at the moment?
- On a scale of 1 to 10 how attentive do you feel at the moment?

In post_task1 and post_task2 surveys, additional task-specific questions were also included. The questions included in post_task1, and post_task2 were:

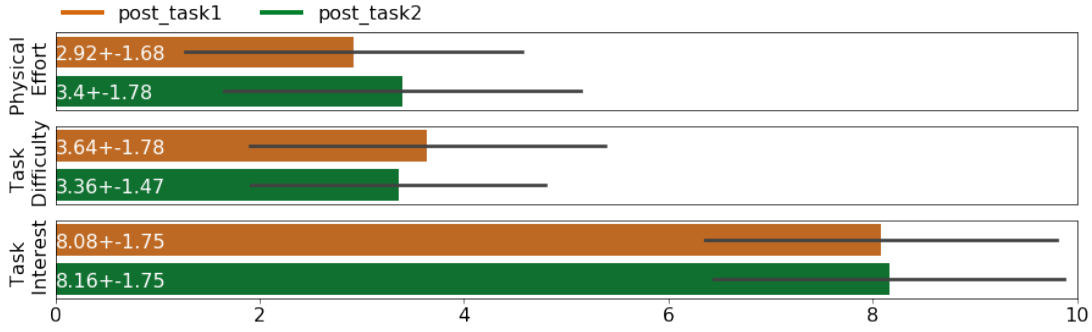
- How much physical effort did you spend on this task?
- How difficult was this task?
- On a scale of 1 to 10 how sleepy or drowsy did you feel during the task?

- On a scale of 1 to 10 did you have difficulties concentrating during the task?
- On a scale of 1 to 10 did you feel physically tired while assembling?
- On a scale of 1 to 10 how stressed did you feel during the task?
- On a scale of 1 to 10 how attentive were you during the task?
- On a scale of 1 to 10 how interested were you in the task?

Figure 5.6 shows a summary of the survey responses for all the questions. Figure 5.6a shows a comparison of the survey questions common for baseline, post_task1, and post_task2. Figure 5.6b shows a comparison of the task-specific survey questions. Some important responses to note in Figure 5.6 are responses to sleepiness, stress, attention, and physical effort. The responses indicate that the the users felt more stress during *S2* than in *S1*. This indicates that the timed RoboAssist setup can induce enough stress in the participant that they can actually feel, thus making this system useful to assess stress-related measures. On the other hand, the participants felt more sleepy or drowsy during *S1* than in *S2*. This indicates that there is a negative correlation between stress and sleep. That is, the participants felt less sleepy as the stress increased while working on a task. Responses to attention and physical effort for the two sessions imply that the participants had to apply more physical effort for a timed session (*S2*) and a relatively high attention level during both sessions. This could indicate that in an assembly task, there is a need for the participant to exert more effort both cognitively and physically. This bolsters the need to implement operator well-being measures in an HRC setup that assess human factors affecting task performance, to provide personalized assistance and alleviate stress and fatigue. However, it is important to note that the values obtained from the user surveys were not significantly different between the two tasks. This is because of the small number of participants in the study, and further experiments are required to validate these findings.



(a)



(b)

Figure 5.6: Subjective feedback from user feedback. (a) A comparison of user response across the three Baseline, post_task1, and post_task2 surveys. (b) A comparison of user response after post_task1, and post_task2 surveys.

5.4.1.2 Machine Learning Results

As mentioned earlier, ECG and EDA data were recorded during RoboAssist to model participants' CL. Out of the data collected from 25 participants, three were dropped due to sensor malfunction. The remaining data was split into two classes; *S1* was considered "low CL" while *S2* was considered "high CL". A total of 44 data points were available that were equally split across the two classes. After preprocessing the signal to reduce noise, 17 time and frequency-domain features were extracted from the ECG signal (Section 5.3.2.1) while 16 statistical, spectral, and energy-domain features are extracted from EDA signals (Section 5.3.2.2). The EDA data comprises two different components, the phasic and the tonic component, as shown in Figure 5.5. For this preliminary analysis, the EDA phasic component was used to extract features for analysis because the analysis focused on modeling the user's cognitive load to the presented stimuli. The Support Vector Machine (SVM) algorithm with a linear kernel is used for this preliminary analysis to see if the data is viable for a classification task. The number of features is also reduced using Principal Component Analysis (PCA) to avoid overfitting and reduce computational complexity. Table 5.3 shows the results

Table 5.3: Preliminary results using Support Vector Machines (SVM) to predict cognitive load using data collected from RoboAssist. Results are presented SVM with and without Principal Component Analysis (PCA) for each signal: ECG, EDA Phasic, and a combination of both signals. Abbreviation: F1 - F1 Score; Acc - Accuracy

	ECG		**EDA_Phasic**		**ECG+EDA**	
	F1	**Acc**	**F1**	**Acc**	**F1**	**Acc**
SVM	0.333	42.85	0.714	71.42	**0.941**	**92.85**
SVM (PCA)	0.667	57.14 (4)	0.833	78.57 (10)	**0.933**	**92.85 (15)**

of both SVM with linear kernel and SVM with PCA using a linear kernel which is evaluated using accuracy and F1-scores. Generic accuracy and F1 scores are used

where 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 [135]. F1-score gives a better understanding of the misclassified cases as it is critical in the framework's design. These results show promising classification performance and viability of the RoboAssist system.

5.4.2 Limitations and Research questions

Despite the high accuracy and F1 score produced by the algorithms discussed in Section 5.4.1.2, more dataset is needed to generalize well. In an industrial setup, multimodal data may not always be feasible. ECG and EDA sensors in the current setup require many cables on the participant, which may not be preferred by an assembly line worker or a robot operator. On the other hand, it is complicated to design a single unobtrusive sensor that comprises different sensors, and an operator may not choose to wear multiple sensors. These limitations in the RoboAssist system gives rise to several questions such as:

- Is it possible to use data collected in a different experimental setup to train a model and predict CL in RoboAssist?
- Can data collected from a different brand of the same sensors be used to build machine learning models for RoboAssist?
- Does the distribution of the data affect the classification problem?
- Which of the several extracted ECG or EDA features are useful?
- In order to increase the number of samples for machine learning techniques, how to augment data for ML models while traditional augmentation techniques tend to induce noise?

In the following sections, different approaches used to answer these research questions are discussed.

5.5 Modeling Cognitive Load from Public Datasets

Training machine learning or deep learning models for modeling CL needs a large dataset to generalize well in real-time. However, most human-centric research studies in academia often face a shortage of participants and usually only manage to collect data from the system using hundreds or fewer participants. Sometimes, data collected in these studies may need to be removed during analysis due to issues with the data collection procedure or too much noise in the data. For instance, physiological sensor data collected using the RoboAssist system consists of 25 participants, of which three were removed during analysis due to sensor malfunction while recording. One solution to address this issue is to use publicly available datasets. However, these datasets also face the same shortcomings. In addition to that, publicly available datasets may not use the same set of sensors we want to use or may have a completely different data collection methodology. This research explores the possibility of using different publicly available datasets collected using different types of sensors and different experimental setups to increase corpus size and build better and generalizable models. More specifically, the 9PM Cognition Dataset [7] and CLAS - Cognitive Load, Affect, and Stress Detection Dataset [6] are explored. These datasets are used for modeling cognitive load and tested on the RoboAssist dataset both individually and by merging them.

5.5.1 9PM Cognition Dataset

The 9PM cognition dataset [7] is a publicly available dataset that comprises data collected from different cognitive tests that cover a wide range of frequently used assessments in research, and clinical practice [136]. It includes cognitive tests such as, the Stroop Test [137], the Wisconsin Card Sorting Test (WCST) [138] and the

NIH Toolbox Picture Sequence Memory Test (PSMT) [139] that was superimposed in the 9 Hole Peg Test (9HPT)[140]. The participants performed the 9HPT based on instructions provided using the cognitive tests. The dataset comprises data from 63 participants that include 56 male and seven female participants. The dataset includes data from physiological sensors such as ECG, EDA, EEG, and an Inertial Measurement Unit (IMU). ECG and EDA data are recorded using a Biosignalsplux Explorer unit [40], EEG data is recorded using an OpenBCI ULTRACORTEX MARK IV sensor [141], and IMU data is recorded using a MetaMotionR sensor [142]. For this research, data from the ECG and EDA sensors recorded at 1000Hz were utilized for modeling machine learning algorithms to predict cognitive load. The ECG sensor is mounted on the participant using a lead II configuration of the standard 3-point bipolar limb leads configuration of the Einthoven's triangle. The EDA sensors were attached to the two fingers of the non-dominant hand.

In the 9PM data collection setup, physiological data were recorded from the participants while performing five different tasks (T1, T2, T3, T4, T5). T1 is a classic 9HPT where the participants were asked to move pegs from the source area to the destination area. T2 is a simple low cognitive load task where participants need to follow the given instructions directly. T3 follows the Stroop Test, where participants are given instructions based on Stroop Test, and appropriate action needs to be taken. Similarly, T4 follows the WCST, and T5 follows PSMT. After each task, a user survey was also collected that provides insight into the subjective user responses regarding the task, such as task difficulty, cognitive load, physical load, and many more. Responses from the user surveys were used to split the dataset into different classes (see Section 5.5.4) for machine learning analysis.

5.5.2 CLAS - Cognitive Load, Affect and Stress Detection Dataset

CLAS [6] is also a publicly available dataset consisting of data collected from various interactive and perceptive tasks. An interactive task is a task that is designed with the primary objective of obtaining a qualitative and quantitative assessment of various cognitive aspects such as cognitive load, attention, and concentration. To this end, Math Test, Logic test, and Stroop test are designed that has several questions to respond to within a very short time limit. The Math test consists of a sequence of 24 mathematical questions that the participants had to respond to within four seconds. The Logic test consists of 20 questions, similar to questions often used in IQ tests [143], that the participants had four seconds to respond. On the other hand, in a perceptive task, the participants are presented with instructions to watch images or videos curated to elicit specific emotions in the four different quadrants of the arousal-valance space as shown in Figure 5.7, using the International Affective Picture System (IAPS) dataset [144]. The different emotions are stress, excitement, boredom, and calmness. The CLAS data consists of data collected from 62 healthy participants that include 45 men and 17 women.

Physiological data from ECG, EDA, and PPG sensors were recorded during the experiment protocol. ECG and PPG data were recorded using the Shimmer3 ECG Unit [145], while the EDA data was recorded using the Shimmer3 GSR+ Unit [146]. The data from these sensors were recorded at a sampling rate of 250Hz. For this research, data from the ECG and EDA sensors recorded during the Math and Logic tests were utilized for modeling machine learning algorithms to predict CL in RoboAssist. In addition to this data, the dataset also consists of a neutral dataset, which is the sensor data collected when the participant is not performing any task in order to relax the participants and restore their emotional state. The neutral sessions

Figure 5.7: Illustration of the distribution of stimulus across the different quadrants of the valance and arousal scale [6], ©2019 IEEE.

were conducted before and after the Math and Logic tests. In the CLAS dataset, ECG was recorded from the Lead I configuration of the Shimmer3 ECG Unit, and EDA was recorded from sensors attached to the two fingers of the non-dominant hand.

5.5.3 Differences Between the CLAS and 9PM Datasets

The CLAS and the 9PM datasets were specifically chosen to explore how datasets with different properties like sensors used, sensor placement, and experimental setup affect the modeling of CL in a real-world scenario. This section will briefly go over some of the datasets' key aspects and how the two datasets differ between themselves and with the RoboAssist system setup.

Data Collection Setup: The 9PM dataset used long cognitive tests like the Stroop Test, WCST, and PSMT superimposed in the 9HPT to induce cognitive load. The CLAS dataset used short standardized tests such as Math and Logic tests to induce cognitive load.

Sensors Used: The 9PM dataset used the biosignalsplux sensor suite that collects data at 1000 Hz, while CLAS uses the Shimmer3 ECG unit for ECG and the Shimmer3 GSR+ unit for EDA data that collects data at a sampling rate of 250 Hz to record data. In the RoboAssist setup, the biosignalsplux sensor suite is used for ECG and EDA data collection at a sampling rate of 1000Hz. *Sensor Placement:* In the 9PM dataset, ECG is collected from a Lead II setup of the Einthoven's triangle similar to the proposed RoboAssist system, while CLAS uses a Lead I configuration of the Shimmer3 unit. On the other hand, for EDA data, both 9PM and CLAS datasets record EDA data from the fingers of the non-dominant hand while the RoboAssist system records EDA data from the right shoulder.

These differences clearly show that the two datasets used are fundamentally different in how they induce CL and the sensors and sensor setup used for data collection.

5.5.4 Machine Learning Analysis to Predict Cognitive Load

An exploratory research was conducted with data from the 9PM dataset, the Math test data, and the Logic test data of the CLAS dataset separately to train different machine learning algorithms that predict CL in the RoboAssist system. The data collected in the preliminary user study (Section 5.4) using the RoboAssist setup was used as the test set. This section briefly explains how data is represented for training from 9PM and CLAS and explains the machine learning pipeline used to predic CL using the RoboAssist data.

5.5.4.1 Data Representation and Feature Extraction

9PM Dataset: Data from the 9PM dataset (Section 5.5.1) is split based on the different tasks the user performed. Each task is assigned a class label based on the

subjective user report for task difficulty. Figure 5.8 shows the average user response to the task difficulty question on the user surveys.

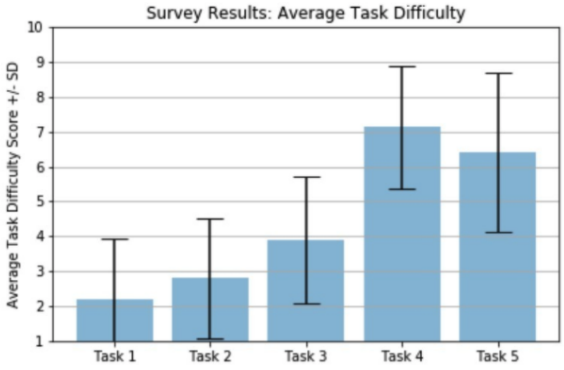


Figure 5.8: A Graph of Task Difficulty from User Survey Data [7].

For the purpose of this study, the prediction of CL was modeled into a binary classification problem by considering data from T1 as low CL class and T4 as high CL class based on the results of the survey. Out of the data from 63 participants, data from 13 participants were dropped for analysis because of extremely noisy data or sensor malfunction. In T1, the participant performed the task for one round, while in T4, the participant performed four rounds of the same task. Due to this, the Low CL class consists of 50 samples, while the high CL class consists of 200 samples causing an imbalance in the two classes. In machine learning, an imbalanced class can cause the machine learning model to overfit the data and cause the model to predict only the majority class during testing. This issue makes the model very unstable and unusable. A property of the 9PM dataset is utilized to address this issue. As mentioned earlier, the low CL class consists of only one round of data per user, whereas the high CL class consists of four rounds of the same task per participant. Once the features were extracted, the average features of each participant were calculated for the majority

high CL class to make one sample per participant, thus reducing the number of samples in the majority class to 50 samples.

Finally, features were extracted using the techniques discussed in Section 5.3.2 for each data sample of both the classes; 17 features were extracted for ECG data, and 16 features were extracted for EDA data (Section 5.3.2). The phasic component was extracted from the EDA data for analysis and feature extraction as it represents the change in skin conductance response due to stimuli. During analysis, both the Raw signal and the Phasic component were used to determine these individual signals' efficacy in modeling CL.

CLAS Dataset: Data from the CLAS dataset (Section 5.5.2) is divided into two classes based on the recording mode. Physiological data recorded during the Math/Logic test was considered high CL, while data recorded during neutral stimuli recorded at the beginning of each session and in-between the tasks were used as low CL. This dataset consists of two types of tests that elicit different cognitive responses and uses different cognitive abilities. Hence, data from Math and Logic Tests were used as two different corpora for modeling CL. The Math Tests consists of 24 questions, and each data recording lasts eight seconds long that includes 4 seconds for listening to questions, 3 seconds to respond, and 1 second to display the result. The Logic Test consists of 20 questions, and each data recording lasts fifteen seconds long that includes 10 seconds to display a question, 4 seconds to respond, and 1 second to display the correct answer. In both these cases, each question was considered as a sample. Data from the two classes were then preprocessed before extracting features using the techniques discussed in Section 5.3.2. Because of the different configuration of the ECG sensor during data collection, the data was not inverted before feature extraction. Finally, 17 features were extracted for ECG data,

but only 12 time-domain features were extracted for EDA data (Section 5.3.2) because of the short signal length. Similar to the 9PM dataset, the phasic component was extracted from the EDA signal for feature extraction and further analysis. During analysis, both the Raw signal and the Phasic component were used to determine these individual signals' efficacy in modeling CL.

After feature extraction, for analysis using the Math test, the high CL class consisted of 646 samples, and the low CL class consisted of 324 samples. Similarly, while using the Logic test, the high CL class consisted of 1085 samples, and the low CL class consisted of 324 samples. This data split indicates that there is an imbalance in the two classes. Unlike the 9PM dataset, the issue is not straightforward because there are no properties of the dataset itself that can address the issue. A Synthetic Minority Class Oversampling Technique (SMOTE) [147] was performed to increase the number of samples in the minority neutral class. This technique works by picking a random example of the minority class and creating a synthetic example similar to its neighbor. A borderline SMOTE [148] technique using SVM as a classifier is used. This algorithm works by locating the decision boundary defined by the support vectors, and examples in the minority class close to the support vectors became the focus for generating synthetic examples. Based on the recommendation by Nguyen et al. [148], the majority class is first undersampled, and then the minority class is oversampled to avoid creating meaningless examples. After applying SMOTE, the number of samples in the Math corpus was balanced to 'high CL' and 'low CL' with 516 samples, whereas the Logic corpus was balanced to 868 samples in both the classes.

RoboAssist Dataset: As mentioned earlier, the physiological data recorded during the RoboAssist preliminary study was used as the test set to predict CL. In this dataset, 3 participant's data were dropped due to sensor malfunction. Thus, the

dataset comprised 22 samples in the low CL class and the high CL class. The pre-processing and feature extraction steps were the same as explained in Section 5.3.2. However, the number of features extracted for this dataset varied based on the training set used. When the 9PM dataset is used, 17 features were extracted for ECG that includes time and frequency domain features, and 16 features that include spectral, energy, and statistical-based features were extracted for EDA signal. On the other hand, when the CLAS data was used for training, 17 ECG features and only 12 EDA features were extracted. This is because the signal length of the CLAS data samples was less than or equal to 10 seconds which is too little to extract any meaningful frequency domain information.

5.5.4.2 Machine Learning Pipeline

For the classification of CL, several traditional machine learning algorithms that are commonly used for modeling sensor data in HMI studies are used [68]. More specifically, algorithms like Support Vector Machines (SVM), AdaBoost (AB), Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Naive Bayes (NB) are used. Three different analyses were performed using the three combinations of the data groups. They are:

1. ECG and EDA data from the 9PM dataset was used to train machine learning models, while the data from the RoboAssist system was used to predict CL.
2. ECG and EDA data from the Math test of the CLAS dataset was used to train machine learning models, while data from the RoboAssist system was used to predict CL.
3. ECG and EDA data from the Logic test of the CLAS dataset was used to train machine learning models, while data from the RoboAssist system was used to predict CL.

Before training the algorithms, different pre-training steps were performed to ensure better algorithm performance on the testing data. Since the goal is to use these algorithms in real-time, fewer features ensure less computational overhead. Hence, different feature selection techniques were performed in all three analyses to ensure fewer features are found to train on that yield the best results. Feature selection techniques such as the Univariate feature selection using ANOVA (UFA) and dimensionality reduction techniques such as the PCA were explored to see their effects on different datasets. Depending on the type of technique used, the data were first normalized using either robust or standard scaling. That is, for PCA, robust normalization was used whereas, standard scaling was used for the UFA. Standard scaling was used with UFA because the ANOVA test assumes that the data is normally distributed and has an equal variance. On the other hand, robust normalization was used with PCA because visualization of the data samples showed several outliers on the training data.

One of the main goals of this study was to combine different corpora so that multiple data sources can be used to model a specific human factor for HRC studies. To this end, based on the results discussed from these analyses, data from the 9PM and CLAS are combined to predict CL. In the following sections, results from these studies are discussed along with the final results from the combined dataset that predicts CL in the RoboAssist System.

5.5.5 Machine Learning Results

A multimodal and unimodal approach is used for building machine learning algorithms. Different combinations of sensor data such as only ECG (E), only EDA Raw (ER), only EDA Phasic (EP), ECG and EDA Raw (E+ER), and ECG and EDA Phasic (E+EP) are used. As mentioned earlier, it is essential to have a lower

computational load on the system while building machine learning or deep learning algorithms for real-time systems. A different number of features for PCA (C value) and UFA (K value) were experimented with such as, a C/K value of 3 to 15 was used for unimodal analysis, and a C/K value of 3 to 25 was used for multimodal analysis. The minimum C/K value that provided the best weighted average F1 score and accuracy was selected and presented. It is important to note that different C/K values affect the accuracy of different machine learning algorithms differently [149]. Hence, for each algorithm and signal combination, the best C/K value that provides the best precision and recall scores for each class, the weighted average F1 score, and accuracy are reported and highlighted in green.

5.5.5.1 Results from the 9PM dataset

Table 5.4 shows the classification results of the RoboAssist system, which was trained using the 9PM dataset. In this first analysis, PCA is used to reduce the dimensionality of the features, and robust normalization is used to normalize the data. The results reported are the best results achieved for each signal type and lists the best algorithm and the C value that achieves it. From this result, it is evident that EDA Raw data achieves the best results. The poor performance of the ECG signal is because heart activity monitored using ECG may be different across different activities whereas, EDA data is a signal that gets activated based on stimuli. However, it must be noted that the phasic component of the EDA signal also shows poor performance. This is because the phasic component varies based on the stimuli shown. Because the task in the training set is entirely different from the task in the test set, such a result is expected.

Similar results are observed in the second analysis as shown in Table 5.5, where the UFA feature selection was performed on the 9PM dataset, and standard scaling is

Table 5.4: Classification results of predicting cognitive load on the RoboAssist data. For each combination, PCA was performed and the least best C value is shown.

MODEL	BEST_C	SIGNAL	RECALL		PRECISION		AVG_F1	ACC
			Low_CL	High_CL	Low_CL	High_CL		
NB	5	E	0.3636	0.7727	0.6154	0.5484	0.549326	0.568182
SVM	6	ER	0.8636	0.9091	0.9048	0.8696	0.886305	0.886364
SVM	5	EP	0.5455	0.6364	0.6	0.5833	0.590062	0.590909
AB	13	E+ER	0.4091	0.9545	0.9	0.6176	0.65625	0.681818
RF	13	E+EP	0.5909	0.5455	0.5652	0.5714	0.567959	0.568182

used to standardize the data. UFA provides a list of features that achieve these best results. For the EDA raw signal, the SVM algorithm provided the best results using features such as Mean, Min, Range, Energy, 1st Mean, 2nd Mean, 2nd Std, Spectral Centroid, Spectral Entropy (see Section 5.3.2.2). These features are a combination of statistical and spectral features.

Table 5.5: Classification results of predicting cognitive load on the RoboAssist data. For each combination, Univariate feature selection was performed and the least best K value is shown. **Best Features**: Mean, Min, Range, Energy, 1st Mean, 2nd Mean, 2nd Std, Spectral Centroid, Spectral Entropy

MODEL	BEST_K	SIGNAL	RECALL		PRECISION		AVG_F1	ACC
			Low_CL	High_CL	Low_CL	High_CL		
RF	3	E	0.6818	0.6364	0.6522	0.6667	0.658915	0.659091
RF	9	ER	0.8636	0.9091	0.9048	0.8696	0.886305	0.886364
SVM	9	EP	0.6818	0.5	0.5769	0.6111	0.5875	0.590909
AB	17	E+ER	0.5455	0.8636	0.8	0.6552	0.696873	0.704545
AB	18	E+EP	0.2273	0.9091	0.7143	0.5405	0.511397	0.568182

5.5.5.2 Results from the Math Test Data of CLAS dataset

The third analysis uses the Math test data from the CLAS dataset, where physiological sensor data recorded during the Math test are used to train different machine learning algorithms. These algorithms are then tested on the RoboAssist data to predict CL. Table 5.6 shows the results which indicate similar performance

compared to the 9PM results discussed above. The poor performance of ECG signals could be explained because of the difference in the two activities performed. Similarly, the performance of EDA phasic data can be explained because of the different tasks in training and testing data.

Table 5.6: Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from the Math test, PCA was performed and the least best C value is shown.

MODEL	BEST_C	SIGNAL	RECALL		PRECISION		AVG_F1	ACC
			Low_CL	High_CL	Low_CL	High_CL		
RF	3	E	0.4545	0.7273	0.625	0.5714	0.583157	0.590909
AB	6	ER	0.5	1	1	0.6667	0.733333	0.75
RF	9	EP	0.5455	0.6364	0.6	0.5833	0.590062	0.590909
RF	5	E+ER	0.3636	0.9545	0.8889	0.6	0.626485	0.659090
XGBoost	7	E+EP	0.6364	0.5	0.56	0.5789	0.566165	0.568181

Table 5.7 shows the results of the fourth analysis where the UFA feature selection with standard scaler is used on the sensor data recorded during the Math test. The results show much improved performance compared to the third analysis, providing the best prediction performance using the EDA Raw (ER) signal. The prediction accuracy increased by 11% and the F1 score increased by 13% in comparison to the PCA dimensionality reduction discussed in Table 5.6.

Table 5.7: Classification results of predicting cognitive load on the RoboAssist data. For each sensor data combination fom Math Test, Univariate feature selection was performed and the least best K value is shown. **Best Features:** Spectral Centroid, Spectral Rolloff, Spectral Entropy.

MODEL	BEST_K	SIGNAL	RECALL		PRECISION		AVG_F1	ACC
			Low_CL	High_CL	Low_CL	High_CL		
RF	5	E	0.6364	0.6818	0.6667	0.6522	0.658915	0.659091
RF	3	ER	0.7273	1	1	0.7857	0.861052	0.863636
SVM	9	EP	0.3182	0.9091	0.7778	0.5714	0.576684	0.613636
NB	18	E+ER	0.4091	0.8182	0.6923	0.5806	0.596765	0.613636
AB	12	E+EP	0.3182	0.8636	0.7	0.5588	0.558036	0.590909

5.5.5.3 Results from the Logic Test Data of CLAS dataset

In this fifth analysis, physiological data recorded during the Logic test was used to train machine learning algorithms. The trained models were tested on the RoboAssist dataset to predict CL. PCA is used to reduce the dimensionality of the features, and robust normalization is used to normalize the data. The results shown in Table 5.8 indicate that the EDA Raw (ER) signals perform the best. However, the ECG data (E) and the multimodal data combination of ECG and EDA Raw (E+ER) also improves its prediction. This could be because the Logic test emulates more similar physiological response in heart rate in comparison to the RoboAssist data.

Table 5.8: Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from Logic test, PCA was performed and the least best C value is shown.

MODEL	**BEST_C**	**SIGNAL**	**RECALL**		**PRECISION**		**AVG_F1**	**ACC**
			Low_CL	**High_CL**	**Low_CL**	**High_CL**		
SVM	9	E	0.7273	0.6818	0.6957	0.7143	0.704393	0.704545
SVM	9	ER	0.5455	0.9545	0.9231	0.6774	0.739084	0.75
RF	7	EP	1	0.0909	0.5238	1	0.427083	0.545455
XGBoost	22	E+ER	0.6818	0.7727	0.75	0.7083	0.726708	0.727273
XGBoost	14	E+EP	0.8182	0.3636	0.5625	0.6667	0.568627	0.590909

In the sixth and final analysis, the UFA feature selection with standard scaling predicts CL in the RoboAssist dataset. The machine learning algorithms are trained using the physiological data recorded during the Logic test. The results shown in Table 5.9 indicate the best performance in all the analyses performed so far. The EDA Raw (ER) signal once again provides a high prediction accuracy that is increased from the PCA analysis discussed in Table 5.8 approximately by 13% F1 score and 11% accuracy. However, in this case, the multimodal data combination of ECG and EDA raw signal (E+ER) provides the best results that improve from the PCA analysis

by 18% both in terms of F1 score and accuracy. It is important to note that the best K value is also higher in the multimodal scenario, which could mean a higher computational load. Hence, a researcher must consider the tradeoff between higher accuracy and better performance.

Table 5.9: Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from Logic test, Univariate feature selection was performed and the least best K value is shown. **ER Best Features:** Max, Range, Standard deviation, Variance, Zero Crossing Rate **E+ER Best Features:** BPM, IBI, medianNN, rangeRR, SDRR, SDSD, SDHR, CVRR, RMSSD, MAD_RR, Mean, Min, Spectral_Centroid

MODEL	BEST_K	SIGNAL	RECALL		PRECISION		AVG_F1	ACC
			Low_CL	High_CL	Low_CL	High_CL		
RF	4	E	0.7727	0.5455	0.6296	0.7059	0.654631	0.659091
RF	6	ER	0.8181	0.9091	0.9	0.8333	0.863354	0.863636
XGBoost	11	EP	0.4545	0.8182	0.7143	0.6	0.623932	0.636364
AB	13	E+ER	0.9091	0.9091	0.9091	0.9091	0.909091	0.909091
AB	7	E+EP	0.4545	0.8636	0.7692	0.6129	0.644205	0.659091

5.5.5.4 Results from the Combined Data

The original goal of this study is to combine different datasets to increase the data available for training a machine learning algorithm. A high number of samples in a training set improves the robustness of the model's performance. To this end, based on the results of the analysis performed so far, different inferences were made from the results. The inferences are:

- The UFA feature selection performs better across different corpora.
- Raw EDA signals are best when working across different setups.
- A multimodal setup seems to work best in the data from the Logic test that indicates that heart activity may be similar in Logic test and RoboAssist.

- The best performing algorithms include Support Vector Machines, Random Forest, and Adaboost, which explains that the data may be non-linear.

Based on these findings, the data from the 9PM dataset and the data from the Logic test of the CLAS dataset are merged to address the lack of data available for HRC studies to predict CL. Two different signals, the EDA Raw (ER) and ECG with EDA Raw (E+ER) are used to predict CL in the RoboAssist dataset from this merged dataset. The UFA feature selection is applied on a dataset that is standardized using the standard scaling technique to train different machine learning algorithms like SVM, RF, and AB. The results shown in Table 5.10 show that when different corpora

Table 5.10: Classification results of predicting cognitive load on the RoboAssist data. For each combination of sensor data from the combined dataset, Univariate feature selection was performed and the least best K value is shown. **Best Features:** EDA Raw - RF - Mean, Min, Max, Energy, 2nd Std, Spectral_Centroid, Spectral_Rolloff, Spectral_Entropy, Entropy_of_Energy **Best Features:** EDA Raw - SVM - Mean, Min, Energy, Spectral_Centroid, Spectral_Rolloff, Spectral_Entropy, Entropy_of_Energy

	SVM			**RF**			**AB**		
	Best_K	**Avg_F1**	**Acc**	**Best_K**	**Avg_F1**	**Acc**	**Best_K**	**Avg_F1**	**Acc**
ER	7	0.840166	0.840909	9	0.863636	0.863636	9	0.746729	0.75
E+ER	3	0.608582	0.613636	13	0.576684	0.613636	13	0.705357	0.727273

of data are combined, EDA raw signal seems to provide the most valuable information for the algorithm to model cognitive load. Thus confirming our previous analysis. The feature combination that provides the best results are, Mean, Min, Max, Energy, 2nd Std, Spectral_Centroid, Spectral_Rolloff, Spectral_Entropy, Entropy_of_Energy. While other signals provide better than random results, further research is needed to understand the interaction between the features as the chosen feature extraction techniques do not address the non-linearity across the features.

5.6 Conclusion and Discussion

In this chapter, an intelligent HRC framework that monitors cognitive human factors to enhance interaction, CogniSmart, is presented. For this framework, a robot-assisted assembly task called RoboAssist is developed. This task simulates a real-world scenario where a robot and a human operator work together to achieve a common goal. To assist the human operator and enable a safe and productive workflow, unobtrusive and wearable physiological data from sensors like ECG and EDA are monitored to predict the user's cognitive load. This information can be used further in the CogniSmart framework to change or adapt the robot's behavior like speed and other parameters to make the collaboration productive.

Initial findings from the user survey (Section 5.4.1.1) show that the participant's answer to their attention level during the task indicated an increase in cognitive load exertion, indicating that the RoboAssist system can be used to induce cognitive load. The responses also showed that the participants felt less sleepy as the task progressed, which also indicated increased attention to the task. The responses also show increased physical effort and attention exertion that bolsters the need to build well-being measures into the HRC system. This is because a continuously increased physical and cognitive exertion may lead to cognitive fatigue that may lead to severe physiological and mental problems if left unchecked.

Different limitations of an HRC system are discussed in Section 5.4.2 which gives rise to the need for using public datasets despite the restriction of various sensors or experimental setup used. Two different datasets, the 9PM cognition dataset and the CLAS dataset, are explored in Section 5.5. As discussed in Section 5.5.5, an exploratory research to find the performance of these datasets in predicting CL indicates that data from EDA Raw signals performs better than ECG and EDA Phasic

signals. These results suggest few crucial facts about the underlying physiology of the participants. The EDA raw signal seems to provide general information regarding the user's affective response to the given task, which helps model the generic cognitive behavior. However, the EDA phasic component may be too specific to the stimuli presented for a task and hence does not provide any helpful information to learn for the machine learning algorithm. Additionally, the tasks used for data collection in the 9PM and CLAS datasets are entirely different from RoboAssist. Different data collection setups elicit different physiological responses from the ECG sensor and provide poor prediction performance. Data from the Logic test of the CLAS dataset shows the best results for EDA Raw and ECG signals and a combination of both. This indicates that the Logic test elicits a similar physiological response when compared to the RoboAssist data. Based on these results, data from the 9PM dataset and the Logic Test of the CLAS dataset are combined to increase the number of samples in the training data. The results discussed in Section 5.5.5.4 show the best performance using EDA Raw signal and the best features are Mean, Min, Max, Energy, 2nd Std, Spectral_Centroid, Spectral_Rolloff, Spectral_Entropy, Entropy_of_Energy.

Research using the Cognismart framework is only getting started because, in this study, only one cognitive aspect is explored. However, the results discussed in Section 5.5.5 provide a promising lead for using different corpora for modeling different cognitive and physical conditions. For instance, similar research [150] has used facial expressions to detect fatigue in a robot-assisted setup. By using several such corpora, a database of different cognitive and physical conditions can be created for a holistic HRC system that takes in to account the robot operator's well-being.

CHAPTER 6

Concluding Remarks and Future Directions

6.1 Concluding Remarks

This dissertation focuses on cognitive human factors monitoring for HRC in the context of an industrial assembly system that uses a collaborative robot. To this end, a two-pronged research approach is followed that first explored building intelligent cognitive assessment systems and then exploit the domain expertise gained to build a cognitive assessment system in an HRC framework. More specifically, intelligent cognitive assessment systems are developed that used physiological data to model a cognitive ability for assessment.

Chapter 3 discusses a novel fatigue assessment framework, CogBeacon, that uses the participant's real-time feedback as a label to model cognitive fatigue that the participant might feel during the WCST cognitive assessment task. This chapter also discussed a traditional machine learning approach to model physiological EEG data that is collected using a wearable EEG headset, MUSE (see Figure 2.4). Following this study, in Chapter 4, the EEG data is used in combination with facial expressions and body posture to build a holistic system to assess user performance using task engagement and emotions as factors. In this study, the Sequence Learning task is utilized to assess engagement and predict user performance. This study also introduces a novel expert user interface that can be used by the person administering the study. This user interface design follows a user-centric design approach where a user study (see Section 2.2.4) was conducted to find the appropriate visualization of information to maximize ease of use and effectiveness. This chapter also discusses

the other indirect factors such as sleep (see Section 4.4) that affect cognitive performance. The results indicate a direct correlation between lack of proper sleep and deterioration in performance.

Personalization based on user behavior and emotions to ensure the health and safety of human robot-operator is a popular area of research in recent times. Despite the state-of-the-art results achieved by these assessment systems, these assessment systems can't be directly implemented in the real world. Two major drawbacks that add to this issue are, lack of a large baseline dataset and sensors that are not practical in the real world. For instance, despite being very minimal and wearable, the MUSE EEG headset can't be used in an industrial assembly setup because the robot operator may wear protective headgear and goggles for safety. Hence there is a need to explore alternate wearable and discreet options for cognitive modeling. To this end, an intelligent cognitive human factor monitoring and assessment framework, CogniSmart, is proposed in Chapter 5 to enhance HRC. This framework addresses the need for a unified framework that can assess cognitive abilities and use the information to predict the next robot action for interaction. In this framework, physiological sensors such as ECG and EDA are explored for cognitive modeling.

On the other hand, limited datasets are available for modeling cognitive abilities for HRC. This field of research is still in its infancy, and further research needs to be done to produce more stable datasets. There is also a limitation on the type of dataset to choose based on the sensors used and the data collection methodology followed. To address this issue, this dissertation performed an extensive study on two different corpora as explained in Section 5.5. The results indicate that due to the difference in the tasks used for data collection, not all datasets can model ECG signals as they tend to differ based on the activity performed. Similarly, the phasic component of the EDA signal was too specific to the task's stimuli that it did not

provide better prediction results. However, the results showed that the EDA Raw signal was able to capture the generic cognitive model of the user. Based on these results, the two corpora were merged for training a machine learning model for the prediction of CL. The results provide a promising start to vast research in cognitive modeling for HRC that can use public datasets despite the differences in sensors used and data collection methodology. It is also clear that an EDA sensor is sufficient to build an effective system for modeling CL.

6.2 Future Directions

This research proposed an intelligent human factors monitoring system for HRC that uses physiological signals for modeling CL. A novel HRC task that simulates real-world assembly called RoboAssist is developed to implement and analyze the proposed framework. Despite promising results in predicting CL, there are several avenues for improvement and development. In section 5.5.4, a machine learning approach is used to model CL using specific hand-crafted features. However, there is a lack of understanding of the inter-feature relation and the non-linearity of the data. There is a need for research on representation learning for sensor data which can help understand the signals better. Several representation techniques, such as the deep embedded classification, can be used for representation learning that uses an encoder-decoder network to learn features that maximize the classification results. To ensure that the classification yields the best results, there is a need to explore expert labeling and sampling techniques. Online learning methodologies that use real-time expert labeling for the sensor on the fly can better understand the signal and yield better classification results. This dissertation is the first step towards building a holistic system that can understand general human models for integrating personalized feedback for better engagement and satisfaction with the system. Novel machine learning tech-

niques such as reinforcement learning and interactive machine learning can be used for active learning and expert labeling for adaptation and personalization. Some research in this domain is already underway. Still, progress needs to be made for a real-time system to ensure safe HRC and improve the human partner's well-being.

6.3 Publicly Available Datasets and Implementations

Several open-source contributions were made as a direct result of this research. Data collected from the studies conducted as part of this dissertation are released for public use to further the research in this area. The following are the list of datasets and a data collection platform that is freely available for research use:

1. CogBeacon - A Multi-Modal Dataset for Fatigue Prediction. Available Online: `https://github.com/MikeMpapa/CogBeacon-MultiModal_Dataset_for_Cognitive_Fatigue`.
2. The CogBeacon Data Collection Platform to Collect Multimodal data using the WCST task. This task supports three variations of the WCST cognitive test (see Section 3) and three variations of stimuli such as visual, audio, and text. Available Online: `https://github.com/MikeMpapa/CogBeacon-WCST_interface`
3. An EEG, EDA, and ECG dataset to Model Cognitive Fatigue and Cognitive Load using the N-back Task. Available Online: `https://github.com/akileshrajan/N-back_Cognition`.
4. The 9PM Cognition Datast - A Multimodal Dataset of Behavioral, Performance, and Physiological Data to predict cognitive factors. Available Online: `DatasetforBehavioral,Performance,andPhysiologicalData`.

REFERENCES

[1] 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.

[2] Muse eeg headset. [Online]. Available: https://choosemuse.com/shop/

[3] A. Rajavenkatanarayanan, A. R. Babu, K. Tsiakas, and F. Makedon, "Monitoring task engagement using facial expressions and body postures," in *Proceedings of the 3rd International Workshop on Interactive and Spatial Computing*, 2018, pp. 103–108.

[4] A. R. Babu, A. Rajavenkatanarayanan, J. R. Brady, and F. Makedon, "Multimodal approach for cognitive task performance prediction from body postures, facial expressions and eeg signal," in *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, 2018, pp. 1–7.

[5] A. Rajavenkatanarayanan, H. R. Nambiappan, M. Kyrarini, and F. Makedon, "Towards a real-time cognitive load assessment system for industrial human-robot cooperation," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 698–705.

[6] V. Markova, T. Ganchev, and K. Kalinkov, "Clas: a database for cognitive load, affect and stress recognition," in *2019 International Conference on Biomedical Innovations and Applications (BIA)*. IEEE, 2019, pp. 1–4.

[7] M. Abujelala, V. Kanal, A. Rajavenkatanarayanan, and F. Makedon, "9pm: A novel interactive 9-peg board for cognitive and physical assessment," in *Pro-*

ceedings of the 14th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2021.

[8] 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.

[9] S. Vaidya, P. Ambad, and S. Bhosle, "Industry 4.0 – a glimpse," *Procedia Manufacturing*, vol. 20, pp. 233–238, 2018, 2nd International Conference on Materials, Manufacturing and Design Engineering (iCMMD2017), 11-12 December 2017, MIT Aurangabad, Maharashtra, INDIA. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978918300672

[10] S. Nahavandi, "Industry 5.0—a human-centric solution," *Sustainability*, vol. 11, no. 16, 2019. [Online]. Available: https://www.mdpi.com/2071-1050/11/16/4371

[11] V. Özdemir and N. Hekim, "Birth of industry 5.0: Making sense of big data with artificial intelligence,"the internet of things" and next-generation technology policy," *Omics: a journal of integrative biology*, vol. 22, no. 1, pp. 65–76, 2018.

[12] Shorrock, Steven , Humanistic Systems - Understanding and Improving Human Work, "Four Kinds of 'Human Factors': 2. Factors of Humans," https://humanisticsystems.com/2017/08/12/four-kinds-of-human-factors-2-factors-of-humans/, 2017, online; posted 08-December-2017.

[13] E. MS and M. Ej, "Human factors engineering design," *National Defense Industry Press*, 1992.

[14] C. D. Wickens, S. E. Gordon, Y. Liu, *et al.*, "An introduction to human factors engineering," 1998.

[15] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and trends in human-computer interaction*, vol. 1, no. 3, pp. 203–275, 2007.

[16] T. Bayne, D. Brainard, R. W. Byrne, L. Chittka, N. Clayton, C. Heyes, J. Mather, B. Ölveczky, M. Shadlen, T. Suddendorf, *et al.*, "What is cognition?" *Current Biology*, vol. 29, no. 13, pp. R608–R615, 2019.

[17] What are cognitive abilities and skills, and how can we boost them? [Online]. Available: https://sharpbrains.com/what-are-cognitive-abilities/

[18] Trotto, Sarah , Safety and Health Magazine, "Fatigue and worker safety:Experts say employers play a role in tackling the issue," https://www.safetyandhealthmagazine.com/articles/15271-fatigue-and-worker-safety, 2017, online; posted 26-February-2017.

[19] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[20] T. N. Tombaugh and N. J. McIntyre, "The mini-mental state examination: a comprehensive review," *Journal of the American Geriatrics Society*, vol. 40, no. 9, pp. 922–935, 1992.

[21] R. C. Gershon, D. Cella, N. A. Fox, R. J. Havlik, H. C. Hendrie, and M. V. Wagster, "Assessment of neurological and behavioural function: the nih toolbox." *The Lancet Neurology*, 2010.

[22] "Intro to nih," https://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox.

[23] A. Shafti, A. Ataka, B. U. Lazpita, A. Shiva, H. A. Wurdemann, and K. Althoefer, "Real-time robot-assisted ergonomics," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1975–1981.

[24] J. A. Adams, "Critical considerations for human-robot interface development," in *Proceedings of 2002 AAAI Fall Symposium*, 2002, pp. 1–8.

[25] H. Oliff, Y. Liu, M. Kumar, and M. Williams, "A framework of integrating knowledge of human factors to facilitate hmi and collaboration in intelligent manufacturing," *Procedia CIRP*, vol. 72, pp. 135–140, 2018.

[26] B. Busch, M. Toussaint, and M. Lopes, "Planning ergonomic sequences of actions in human-robot interaction," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1916–1923.

[27] S. C. Huijbregts, N. F. Kalkers, L. M. de Sonneville, V. de Groot, and C. H. Polman, "Cognitive impairment and decline in different ms subtypes," *Journal of the neurological sciences*, vol. 245, no. 1-2, pp. 187–194, 2006.

[28] D. Aarsland, K. Andersen, J. P. Larsen, R. Perry, T. Wentzel-Larsen, A. Lolk, and P. Kragh-Sørensen, "The rate of cognitive decline in parkinson disease," *Archives of neurology*, vol. 61, no. 12, pp. 1906–1911, 2004.

[29] Sarah, Trotto, Safety & Health magazine, published by the National Safety Council, "Fatigue and worker safety — Experts say employers play a role in tackling the issue," https://www.safetyandhealthmagazine.com/articles/15271-fatigue-and-worker-safety, 2017, online; posted 26-February-2017.

[30] L. E. Humes and S. S. Floyd, "Measures of working memory, sequence learning, and speech recognition in the elderly," *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 1, pp. 224–235, 2005.

[31] A. Khalid, P. Kirisci, Z. Ghrairi, J. Pannek, and K.-D. Thoben, "Safety requirements in collaborative human–robot cyber-physical system," in *Dynamics in Logistics*. Springer, 2017, pp. 41–51.

[32] S. E. Gathercole and A. D. Baddeley, *Working memory and language*. Psychology Press, 2014.

[33] S. Grossberg and L. R. Pearson, "Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: toward a unified theory of how the cerebral cortex works." *Psychological review*, vol. 115, no. 3, p. 677, 2008.

[34] F. Lange, C. Brückner, A. Knebel, C. Seer, and B. Kopp, "Executive dysfunction in parkinson's disease: a meta-analysis on the wisconsin card sorting test literature," *Neuroscience & Biobehavioral Reviews*, 2018.

[35] G. Stoet, "Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[36] M. J. Kane, A. R. Conway, T. K. Miura, and G. J. Colflesh, "Working memory, attention control, and the n-back task: a question of construct validity." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 33, no. 3, p. 615, 2007.

[37] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back task—quantified in the prefrontal cortex using fnirs," *Frontiers in human neuroscience*, vol. 7, p. 935, 2014.

[38] T. Lin, M. Omata, W. Hu, and A. Imamiya, "Do physiological data relate to traditional usability indexes?" in *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, 2005, pp. 1–10.

[39] P. Bashivan, I. Rish, and S. Heisig, "Mental state recognition via wearable eeg," *arXiv preprint arXiv:1602.00985*, 2016.

[40] biosignalsplux explorer. [Online]. Available: https://plux.info/kits/215-biosignals-explorer-820201001.html

[41] R. E. Klabunde. Electrocardiogram standard limb leads (bipolar). [Online]. Available: https://www.cvphysiology.com/Arrhythmias/A013a

[42] 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.

[43] 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.

[44] Depth from stereo. [Online]. Available: https://github.com/IntelRealSense/librealsense/blob/master/doc/depth-from-stereo.md

[45] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò, "Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy," *JMIR Serious Games*, vol. 4, no. 2, p. e11, Jul 2016. [Online]. Available: https://doi.org/10.2196/games.5888

[46] A. R. Babu, A. Rajavenkatanarayanan, M. Abujelala, and F. Makedon, "Votre: A vocational training and evaluation system to compare training approaches for the workplace," in *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 2017, pp. 203–214.

[47] M. Papakostas, K. Tsiakas, M. Abujelala, M. Bell, and F. Makedon, "v-cat: A cyberlearning framework for personalized cognitive skill assessment and train-

ing," in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 2018, pp. 570–574.

[48] K. Tsiakas, M. Abujelala, A. Rajavenkatanarayanan, and F. Makedon, "User skill assessment using informative interfaces for personalized robot-assisted training," in *International Conference on Learning and Collaboration Technologies*. Springer, 2018, pp. 88–98.

[49] K. Tsiakas, M. Dagioglou, V. Karkaletsis, and F. Makedon, "Adaptive robot assisted therapy using interactive reinforcement learning," in *International Conference on Social Robotics*. Springer, 2016, pp. 11–21.

[50] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.* IEEE, 2005, pp. 465–468.

[51] S. E. Lerman, E. Eskin, D. J. Flower, E. C. George, B. Gerson, N. Hartenbaum, S. R. Hursh, M. Moore-Ede, *et al.*, "Fatigue risk management in the workplace," *Journal of Occupational and Environmental Medicine*, vol. 54, no. 2, pp. 231–258, 2012.

[52] L. B. Krupp, "Fatigue in multiple sclerosis," *CNS drugs*, vol. 17, no. 4, pp. 225–234, 2003.

[53] A. Belmont, N. Agar, C. Hugeron, B. Gallais, and P. Azouvi, "Fatigue and traumatic brain injury," in *Annales de réadaptation et de médecine physique*, vol. 49, no. 6. Elsevier, 2006, pp. 370–374.

[54] P. Hagell and L. Brundin, "Towards an understanding of fatigue in parkinson disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 80, no. 5, pp. 489–492, 2009.

[55] G. BorragÃ¡n, H. Slama, A. Destrebecqz, and P. Peigneux, "Cognitive fatigue facilitates procedural sequence learning," *Frontiers in Human Neuroscience*,

vol. 10, p. 86, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2016.00086

[56] A. Sehle, M. Vieten, S. Sailer, A. Mündermann, and C. Dettmers, "Objective assessment of motor fatigue in multiple sclerosis: the fatigue index kliniken schmieder (fks)," *Journal of Neurology*, vol. 261, no. 9, pp. 1752–1762, Sep 2014.

[57] L. B. Krupp, N. G. LaRocca, J. Muir-Nash, and A. D. Steinberg, "The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus," *Archives of neurology*, vol. 46, no. 10, pp. 1121–1123, 1989.

[58] K. R. Chaudhuri, D. G. Healy, and A. H. Schapira, "Non-motor symptoms of parkinson's disease: diagnosis and management," *The Lancet Neurology*, vol. 5, no. 3, pp. 235–245, 2006.

[59] A. Qaseem, D. Kansagara, M. A. Forciea, M. Cooke, and T. D. Denberg, "Management of chronic insomnia disorder in adults: a clinical practice guideline from the american college of physicians," *Annals of internal medicine*, vol. 165, no. 2, pp. 125–133, 2016.

[60] P. Vos, Y. Alekseenko, L. Battistin, E. Ehler, F. Gerstenbrand, D. Muresanu, A. Potapov, C. Stepan, P. Traubner, L. Vécsei, *et al.*, "Mild traumatic brain injury," *European journal of neurology*, vol. 19, no. 2, pp. 191–198, 2012.

[61] S. R. Hursh, T. J. Balkin, J. C. Miller, and D. R. Eddy, "The fatigue avoidance scheduling tool: Modeling to minimize the effects of fatigue on cognitive performance," *SAE transactions*, pp. 111–119, 2004.

[62] K. A. Donovan, B. J. Small, M. A. Andrykowski, P. Munster, and P. B. Jacobsen, "Utility of a cognitive-behavioral model to predict fatigue following breast cancer treatment." *Health Psychology*, vol. 26, no. 4, p. 464, 2007.

[63] C. Gonzalez, B. Best, A. F. Healy, J. A. Kole, and L. E. Bourne Jr, "A cognitive modeling account of simultaneous learning and fatigue effects," *Cognitive Systems Research*, vol. 12, no. 1, pp. 19–32, 2011.

[64] J. R. Anderson, C. Lebiere, M. Lovett, and L. Reder, "Act-r: A higher-level account of processing capacity," *Behavioral and Brain Sciences*, vol. 21, no. 6, pp. 831–832, 1998.

[65] D. Golan, G. M. Doniger, K. Wissemann, M. Zarif, B. Bumstead, M. Buhse, L. Fafard, I. Lavi, J. Wilken, and M. Gudesblatt, "The impact of subjective cognitive fatigue and depression on cognitive function in patients with multiple sclerosis," *Multiple Sclerosis Journal*, vol. 24, no. 2, pp. 196–204, 2018.

[66] A. R. Hassan, S. Siuly, and Y. Zhang, "Epileptic seizure detection in eeg signals using tunable-q factor wavelet transform and bootstrap aggregating," *Computer methods and programs in biomedicine*, vol. 137, pp. 247–259, 2016.

[67] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup, "Emd-based temporal and spectral features for the classification of eeg signals using supervised learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 1, pp. 28–35, 2015.

[68] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.

[69] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[70] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.

[71] Better Health Channel, "Fatigue," https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/fatigue, online; last accessed 02-May-2020.

[72] L. S. Aaronson, L. Pallikkathayil, and F. Crighton, "A qualitative investigation of fatigue among healthy working adults," *Western Journal of Nursing Research*, vol. 25, no. 4, pp. 419–433, 2003.

[73] A. Vourvopoulos, A. L. Faria, K. Ponnam, and S. Bermudez i Badia, "Rehabcity: design and validation of a cognitive assessment and rehabilitation tool through gamified simulations of activities of daily living," in *Proceedings of the 11th conference on advances in computer entertainment technology*, 2014, pp. 1–8.

[74] A. Sharifara, A. R. Babu, A. Rajavenkatanarayanan, C. Collander, and F. Makedon, "A robot-based cognitive assessment model based on visual working memory and attention level," in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2018, pp. 583–597.

[75] A. Khawaji, J. Zhou, F. Chen, and N. Marcus, "Using galvanic skin response (gsr) to measure trust and cognitive load in the text-chat environment," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 2015, pp. 1989–1994.

[76] A. A. Blank, J. A. French, A. U. Pehlivan, and M. K. O'Malley, "Current trends in robot-assisted upper-limb stroke rehabilitation: promoting patient engagement in therapy," *Current physical medicine and rehabilitation reports*, vol. 2, no. 3, pp. 184–195, 2014.

[77] K. Tsiakas, M. Abujelala, and F. Makedon, "Task engagement as personalization feedback for socially-assistive robots and cognitive training," *Technologies*, vol. 6, no. 2, p. 49, 2018.

[78] T. McMahan, I. Parberry, and T. D. Parsons, "Evaluating electroencephalography engagement indices during video game play." in *FDG*, 2015.

[79] P. Demosthenous, N. Nicolaou, and J. Georgiou, "A hardware-efficient low-pass filter design for biomedical applications," in *2010 Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2010, pp. 130–133.

[80] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[81] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification," *arXiv preprint arXiv:1710.07557*, 2017.

[82] P.-L. Carrier and A. C. W. Research. The facial expression recognition 2013 (fer-2013) dataset. [Online]. Available: https://datarepository.wolframcloud.com/resources/FER-2013

[83] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," *arXiv preprint arXiv:1611.08050*, 2016.

[84] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4724–4732.

[85] N. Fourati and C. Pelachaud, "Emilya: Emotional body expression in daily actions database." in *LREC*, 2014, pp. 3486–3493.

[86] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[87] N. Bianchi-Berthouze, P. Cairns, A. Cox, C. Jennett, and W. W. Kim, "On posture as a modality for expressing and recognizing emotions," in *Emotion and HCI workshop at BCS HCI London*, 2006.

[88] H. G. Wallbott, "Bodily expression of emotion," *European journal of social psychology*, vol. 28, no. 6, pp. 879–896, 1998.

[89] M. M. Gross, E. A. Crane, and B. L. Fredrickson, "Methodology for assessing bodily expression of emotion," *Journal of Nonverbal Behavior*, vol. 34, no. 4, pp. 223–248, 2010.

[90] M. Papakostas, K. Tsiakas, T. Giannakopoulos, and F. Makedon, "Towards predicting task performance from eeg signals," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 4423–4425.

[91] N. Covassin, M. de Zambotti, M. Sarlo, G. D. M. Tona, S. Sarasso, and L. Stegagno, "Cognitive performance and cardiovascular markers of hyperarousal in primary insomnia," *International journal of psychophysiology*, vol. 80, no. 1, pp. 79–86, 2011.

[92] Fitbit. Fitbit smartwatch. [Online]. Available: https://www.fitbit.com/global/us/home

[93] R. B. D'agostino, A. Belanger, and R. B. D'Agostino Jr, "A suggestion for using powerful and informative tests of normality," *The American Statistician*, vol. 44, no. 4, pp. 316–321, 1990.

[94] D. Kosecki. (2018, dec) Rem, light, deep: How much of each stage of sleep are you getting? [Online]. Available: https://blog.fitbit.com/sleep-stages-explained/

[95] 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.

[96] 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.

[97] 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.

[98] 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.

[99] 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.

[100] 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.

[101] 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.

[102] 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.

[103] F. N. Biondi, A. Cacanindin, C. Douglas, and J. Cort, "Overloaded and at work: Investigating the effect of cognitive workload on assembly task performance," *Human Factors*, p. 0018720820929928, 2020.

[104] 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.

[105] 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.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] S. Profanter, A. Perzylo, N. Somani, 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.

[111] 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.

[112] 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.

[113] 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.

[114] 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.

[115] 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.

[116] 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.

[117] V. Villani, L. Sabattini, 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.

[118] C. T. Landi, V. Villani, F. Ferraguti, L. Sabattini, C. Secchi, and C. Fantuzzi, "Relieving operators' workload: Towards affective robotics in industrial scenarios," *Mechatronics*, vol. 54, pp. 144–154, 2018.

[119] Sawyer black edition. [Online]. Available: https://www.rethinkrobotics.com/sawyer

[120] 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.

[121] OSHA. Occupational safety and health administration - workers' rights. [Online]. Available: https://www.osha.gov/sites/default/files/publications/osha3021.pdf

[122] N. Isakadze and S. S. Martin, "How useful is the smartwatch ecg?" *Trends in cardiovascular medicine*, vol. 30, no. 7, pp. 442–448, 2020.

[123] 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.

[124] 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.

[125] 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.

[126] 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.

[127] H. Critchley and Y. Nagai, *Electrodermal Activity (EDA)*. New York, NY: Springer New York, 2013, pp. 666–669. [Online]. Available: https://doi.org/10.1007/978-1-4419-1005-9_13

[128] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 2012, pp. 420–423.

[129] M.-Z. Poh, N. C. Swenson, and R. W. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity," *IEEE transactions on Biomedical engineering*, vol. 57, no. 5, pp. 1243–1252, 2010.

[130] 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.

[131] 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.

[132] 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.

[133] T. Giannakopoulos, "pyaudioanalysis: An open-source python library for audio signal analysis," *PloS one*, vol. 10, no. 12, 2015.

[134] M. G. Falleti, P. Maruff, A. Collie, and D. G. Darby, "Practice effects associated with the repeated assessment of cognitive function using the cogstate battery

at 10-minute, one week and one month test-retest intervals," *Journal of clinical and experimental neuropsychology*, vol. 28, no. 7, pp. 1095–1112, 2006.

[135] Y. Sasaki, "The truth oh the f–measure," *Manchester: School of Computer Science, University of Manchester*, 2007.

[136] J. Taylor, "Heritability of wisconsin card sorting test (wcst) and stroop color-word test performance in normal individuals: implications for the search for endophenotypes," *Twin Research and Human Genetics*, vol. 10, no. 6, pp. 829–834, 2007.

[137] J. R. Stroop, "Studies of interference in serial verbal reactions." *Journal of experimental psychology*, vol. 18, no. 6, p. 643, 1935.

[138] D. A. Grant and E. Berg, "A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem." *Journal of experimental psychology*, vol. 38, no. 4, p. 404, 1948.

[139] S. S. Dikmen, P. J. Bauer, S. Weintraub, D. Mungas, J. Slotkin, J. L. Beaumont, R. Gershon, N. R. Temkin, and R. K. Heaton, "Measuring episodic memory across the lifespan: Nih toolbox picture sequence memory test," *Journal of the International Neuropsychological Society*, vol. 20, no. 6, pp. 611–619, 2014.

[140] P. Feys, I. Lamers, G. Francis, R. Benedict, G. Phillips, N. LaRocca, L. D. Hudson, R. Rudick, and M. S. O. A. Consortium, "The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis," *Multiple Sclerosis Journal*, vol. 23, no. 5, pp. 711–720, 2017.

[141] "Ultracortex Mark IV — OpenBCI Documentation." [Online]. Available: http://docs.openbci.com/Headware/01-Ultracortex-Mark-IV

[142] "MetaMotionR - MbientLab." [Online]. Available: https://mbientlab.com/metamotionr/

[143] K. Richardson, "What iq tests test," *Theory & Psychology*, vol. 12, no. 3, pp. 283–314, 2002.

[144] P. J. Lang, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," *Technical report*, 2005.

[145] "Shimmer 3 ECG Unit." [Online]. Available: https://www.shimmersensing.com/products/shimmer3-ecg-sensor

[146] "Shimmer 3+ GSR Unit." [Online]. Available: https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

[147] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[148] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 3, no. 1, pp. 4–21, 2011.

[149] 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.

[150] A. R. Babu, J. Cloud, M. Theofanidis, and F. Makedon, "Facial expressions as a modality for fatigue detection in robot based rehabilitation," in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 2018, pp. 112–113.

BIOGRAPHICAL STATEMENT

Akilesh Rajavenkatanarayanan was born in Salem, India. In 2012, he received his Bachelors in Technology from the Department of Computer Science and Engineering at the Amrita University, Coimbatore, India. During his undergraduate education, he was involved in several organizational and volunteering activities and worked on several human-centered application development projects. He has worked as a Programmer Analyst at Cognizant Technology Solutions, India, until May 2014, where he focused on developing several web applications and production support.

In August 2014, he started his Master of Science degree at the University of Texas at Arlington, USA. As a Masters student, he was very curious about Human-centric research and development. He built and developed a mobile phone app that captures hand gestures while driving to control a mobile phone for calls and music. In August 2016, he joined the HERACLEIA - Human-Centered Computed Lab to pursue his research interest in human-computer interaction. He participated as a research associate in NSF-funded projects to develop state-of-the-art human factors monitoring and assessment system under the supervision of Dr. Fillia Makedon. He was the lead researcher of the NSF-funded project "PFI:BIC: iWork, a Modular Multi-Sensing Adaptive Robot-Based Service for Vocational Assessment, Personalized Worker Training and Rehabilitation."

He also served as a member of the organizing committee of the international conference on The PErvasive Technologies Related to Assistive Environments (PETRA) in the years 2018 and 2019. Following this, he was the organizing committee chair for PETRA 2020 and PETRA 2021. In addition to this, he has served as a teaching assis-

tant for courses such as Mobile Systems Engineering, Discrete Structures, Distributed Systems, and Advanced Topics in Human-Computer Interaction. In April 2020, he was awarded by the Department of Computer Science for the outstanding contributions to the CSE department. He was also chosen as the outstanding performer at the Heracleia lab for two consecutive years in 2020 and 2021. In Summer 2020, he worked as a summer research intern in the UT Arlinton Research Institute, where he built and assembled a multi-robot system for rehabilitation and assistance projects. Akilesh defended his dissertation in Spring 2021. His research interest revolves around user experience research and artificial intelligence, focusing on signal processing and machine learning for human behavior analysis and monitoring for Human-Machine Interaction.