AN INTELLIGENT MULTI-MODAL FRAMEWORK TOWARDS ASSESSING HUMAN COGNITION

PhD Dissertation

Presented to the Department of Computer Science
of The University of Texas at Arlington
in Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

ASHISH JAISWAL

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2023

AN INTELLIGENT MULTI-MODAL FRAMEWORK TOWARDS ASSESSING HUMAN COGNITION

ASHISH JAISWAL, Ph.D.

The University of Texas at Arlington 2023

Cognition is the mental process of acquiring knowledge and understanding through thought, experience, and senses. Fatigue is a loss in cognitive or physical performance due to physiological factors such as insufficient sleep, long work hours, stress, and physical exertion. It adversely affects the human body and can slow reaction times, reduce attention, and limit short-term memory. Hence, there is a need to monitor a person's state to avoid extreme fatigue conditions that can result in physiological complications. However, tools to understand and assess fatigue are minimal.

This thesis primarily focuses on building an experimental setup that induces cognitive fatigue (CF) and physical fatigue (PF) through multiple cognitive and physical tasks while simultaneously recording physiological data and visual cues from a person's face. First, we build a prototype sensor shirt embedded with various physiological sensors for easy use during cognitively and physically demanding tasks. Second, participants' self-reported visual analog scores (VAS) are reported after each task to confirm fatigue induction. Finally, an evaluation system is built that utilizes machine learning (ML) models to detect states of CF and PF from multi-modal sensor data, thus providing an objective measure.

This effort is the first step towards building a robust cognitive assessment tool that can collect multi-modal data and be used for industrial applications

to monitor a person’s mental state. For instance, it enables safe human-robot cooperation (HRC) in industrial environments to avoid physical harm when a person’s mental state is not good. Another example can be a personalized assistive robot for individuals with motor impairments to perform a task such as preparing lunch with real-time interventions based on the help required from the user.

THESIS COMMITTEE

Prof. Fillia Makedon

(Supervisor)

Prof. William Beksi

Prof. Shirin Nilizadeh

Prof. Dajiang Zhu

The Late Prof. Ramez Elmasri

To all the people who have been there by my side throughout these years and constantly supported me.

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my supervisor, Dr. Fillia Makedon, for her unwavering belief in me and her continuous and unconditional support throughout my Ph.D. Her guidance has been invaluable, shaping the course of my research.

I would also like to express appreciation to my esteemed committee members: Dr. William Beksi, Dr. Shirin Nilizadeh, Dr. Dajiang Zhu, and the late Dr. Ramez Elmasri. I'm also grateful for the fruitful collaborations with Dr. Morris Bell from Yale University, Dr. Glenn Wylie from the Kessler Foundation, and Dr. Maria Kyrarini from Santa Clara University on various projects. Each of them has provided invaluable insights and feedback that significantly contributed to the completion of this thesis.

To my colleagues from Heracleia Human-Centered Computing Lab, I extend my thanks for the numerous discussions and collaborations that played a unique role in the success of this endeavor. Special thanks are due to Dr. Ashwin Ramesh Babu, Dr. Mohammad Zaki Zadeh, and Dr. Akliesh Rajavenkatanarayanan for their constant encouragement and reality checks, enhancing the quality of my work.

Finally, heartfelt appreciation goes to my parents, family, and friends for being an unwavering source of emotional support and encouragement throughout this challenging journey. Their belief in me has been instrumental in the successful completion of my Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 A Brief Introduction to Human Behavior and Cognition

Behaviorism is a branch of psychology that deals with people's actions based on external environmental influences. In contrast, cognitive psychology is based on the thought process that alters a person's behavior [125]. Human Behavior research focuses on understanding human mental functions, including perception, memory, attention, reasoning, and decision-making. Behavioral and cognitive psychology uses principles of human learning and development and cognitive processing to overcome problem behavior, emotional thinking, and thinking. Human behavior can be expected to be adaptive (i.e., reproduction-maximizing). Hence, a science of human behavior can be based on analyses of the reproductive consequences of human action [113].

Cognition can be defined as a person's ability to understand their surroundings using their cognitive skills to perform a specific task [10]. Several cognitive processes run through our daily activities: Attention, Language, Learning, Memory, Thought, etc. For instance, attention is the process of selecting one piece of information from the environment and focusing on that while ignoring other events going on simultaneously.

Understanding human behavior in the wild remains a far-fetched goal despite the advancement in cognitive assessment technologies. However, the progress in learning behavior and cognition has undoubtedly driven us closer to that goal with the latest techniques and tools.

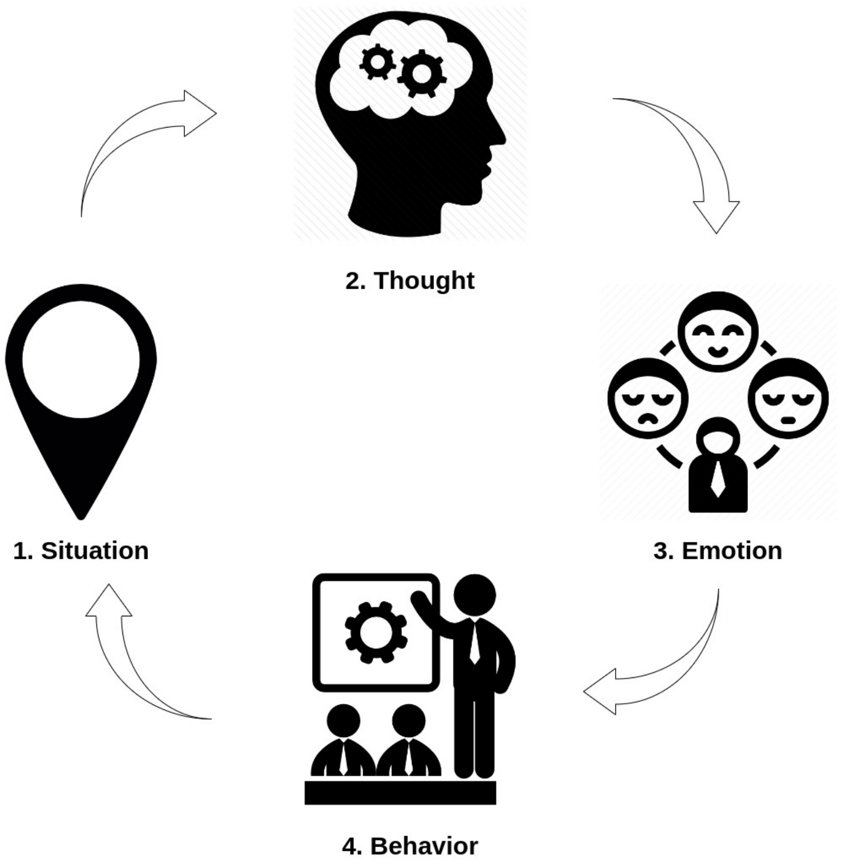


Figure 1.1: Stages of human cognitive development for any given situation. The figure represents the relation between Action, Cognition, and Emotion in Human Behavior. The highly dependent association among these stages makes a human able to perceive the world around them and respond to different situations and tasks.

The focus of this dissertation is to analyze the relation between a person's brain activity (thought process) and their performance in a real-world task. The research focuses on using physiological and vision sensors to monitor a person's cognitive and behavioral cues while operating a robot. These cues and their performance in the task can help in providing specific feedback to the user

1.2 Cognitive Assessments and Monitoring Systems

Cognitive assessments are tests or evaluations that measure various aspects of an individual's cognitive abilities and mental functioning. These assessments aim to gain insight into an individual's strengths and weaknesses in areas such as memory, attention, executive function, processing speed, and language abilities. On the other hand, monitoring systems are computerized systems that use cognitive assessments as part of a comprehensive evaluation to monitor changes in cognitive function over time. These systems typically involve repeated administration of cognitive assessments and can be used in various settings, including clinical research or workplace settings.

The principal components contributing to a specific human behavior are actions, thoughts, and emotions [71]. Understanding complex human behaviors require a system that can simultaneously track multiple interaction signals and monitor parameters that do not always have an apparent correlation [111, 101]. Hence, it encourages researchers to deploy cognitive assessment and rehabilitation tools at home and in the workplace as an urgent necessity.

For instance, Cogbeacon [102] is a computerized game that tests how well people can use their thinking skills, like problem-solving and remembering. The authors measured the subject's brain activity through EEG signals while playing the game and compared their performance on different versions of the game. They concluded that the subjects could learn better if they had played a different version of the same game in an earlier session.

1.3 Understanding Human Factors for Cognitive Assessment

When interacting with the surroundings, many factors may directly or indirectly affect a human's performance in accomplishing a particular task. Here, 'Human Factors' refers to people's different attributes that define their abilities, limitations, and characteristics. According to [121], several factors that may constitute Human Factors in assessing a person's cognition and behavior are as follows:

- Cognitive Functions: attention, working memory, perception, detection, reasoning, and judgment.
- Cognitive Systems: An example is Kahneman's dual process theory, which refers to cognition as a coordinated activity of two independent but connected systems. For instance, it can help in enhancing a person's problem-solving capacity.
- Error Types: Lapse in judgment, reason slips, mistakes, etc.
- Subjective Behaviors and Non-technical Skills: decision making, situation awareness, and teamwork.
- Physical, cognitive, and emotional states: stress, emotion, fatigue, etc., can affect a person's performance.
- Physical characteristics such as speed, strength, balance, accuracy, etc., help in any task performance.

These 'Human Factors' must be understood to study the relationship between the brain and the body. Research in human factors engineering aims to minimize human error, enhance safety, reduce workload, and increase comfort.

In addition, productivity and accuracy can be improved while performing different tasks [45].

1.4 Technology as a Tool for Cognitive Assessment

Digital technologies are a rapidly advancing field that provides a previously unavailable opportunity to alleviate challenges faced by researchers in understanding how human cognition works. As well as monitoring health and cognition, digital technologies that provide adaptive assistance are now emerging due to advances in machine learning [26]. With the recent advancements in AI technology, research communities have focused more on unraveling complex human behavior to understand the connection between the mind and the surrounding environment. Moreover, much work has been carried out to identify factors affecting daily human performance. Several research groups have been working to demystify the behavioral instincts of humans and the reason why a person behaves a certain way in any given situation.

Specifically, advancements in sensory technologies, data acquisition techniques, and analysis have opened new doors to find solutions to these challenging problems [107, 16]. In addition, the number of people diagnosed with severe mental illnesses such as dementia, ADHD, etc., is rising appreciably. Hence the use of technology facilitates repeated and continuous assessment and supports smooth analysis of auxiliary behavioral markers to assess cognitive abilities [75]. However, there's still room for improvement to systematically monitor and break down the brain processes that trigger and support human reactions to different stimuli [89].

Humans can be considered active agents are continuously interacting with their surrounding environment, producing and perceiving countless information at any given moment. Hence, a non-stop process that eventually affects their bodily needs, our reactions, and our mental desires [71] as depicted by Figure 1.1.

Developing neurophysiological models of complex human behavior is essential in exploring human-computer interaction applications that can assist people with cognitive impairment. This thesis focuses on utilizing technologies like sensors and machine learning to understand human behavior and cognition while performing real-world tasks. More specifically, monitoring brain signals while a physically disabled person takes help from an automated robot can give us detailed insights on how they perform real-world tasks.

1.5 Sensor-Based Analysis of Human Behavior and Cognition

Behavior and cognitive analysis using wearable physiological sensors are being used prominently over other survey-based systems such as e-healthcare and life-log analysis systems, especially in the healthcare domain [108]. Moreover, several automated monitoring systems have been built with the help of such sensors for different domain applications like healthcare, security, entertainment, user authentication, 3D games, etc [110, 15, 127, 55, 82, 18]. In this dissertation, we will be using some of these sensors to enhance data collection from human subjects in understanding their behavioral and cognitive skills.

Table 1.1: Common Factors that can have an impact on Human Behavior and Cognition along with their significance

Factors	**Measures**	**Significance**
Physiological	Heart Rate, SpO2, EDA [64], body temperature, sleep quality, respiration rate, etc. [78]	insights on an individual's overall health, stress levels, emotional state, and cognitive function.
Behavioral	eye movements [80], gaze duration [95], posture [69], gait [103], attention & executive function [110], activity levels, etc.	insights on an individual's behavior, physical activity, and cognitive processes.
Environmental	light, noise, temperature, humidity, air quality	insights on the impact of surrounding environmental state on the individual

1.5.1 Physiological Features for Cognitive Assessments

Physiological and vision sensors allow real-time data collection from human subjects while they perform day-to-day tasks. They open the potential of assessing essential traits and components for identifying the cognitive load on the participants [60]. The sensor systems employed nowadays are used to augment and replace task performance measures when they are unavailable.

For instance, physiological sensors such as electroencephalography (EEG) and electrocardiogram (ECG: for heart rate variability) can be used to measure the brain's electrical activity and changes in the heart rate, respectively. These physiological signals can provide valuable information about an individual's level of arousal, attention, and stress, which are essential indicators of cognitive function [109].

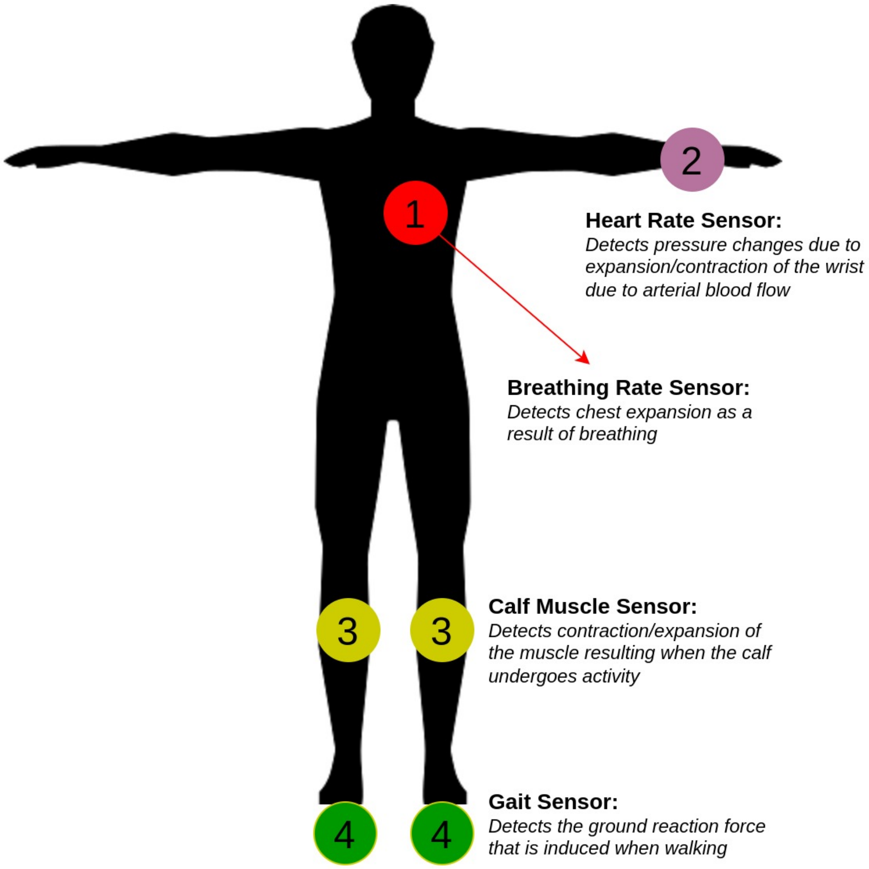

2
1
Heart Rate Sensor:
Detects pressure changes due to expansion/contraction of the wrist due to arterial blood flow
Breathing Rate Sensor:
Detects chest expansion as a result of breathing
3
3
Calf Muscle Sensor:
Detects contraction/expansion of the muscle resulting when the calf undergoes activity
4
4
Gait Sensor:
Detects the ground reaction force that is induced when walking

Figure 1.2: Examples of different wearable sensors that can be used to measure different physiological signals from a human body. [59]

1.5.2 Behavioral Features for Cognitive Assessments

Vision sensors such as eye tracking devices or RGB cameras are widely used to measure a person's eye movements and gaze duration. These observable behaviors can provide insight into the individual's attention levels, visual search patterns, and cognitive processes [80]. We can also use eye tracking to evaluate visual attention, visual processing speed, and the ability to sustain attention over time, which are essential indicators of cognitive function. Earlier work have shown the value of tracking eye movements and changes in pupil size as measures of cognitive load in human subjects [60, 61].

For example, Saccades are rapid eye movements that shift the gaze from one point to another. The frequency, amplitude, and velocity of saccades can provide information about the efficiency of visual attention and the ability to direct attention to relevant information [84]. Moreover, other behavioral features like posture and gait movements of a person can provide more information about their engagement in a particular task [69, 103].

CHAPTER 2

COMPONENTS OF AN INTELLIGENT COGNITION ASSESSMENT FRAMEWORKS

An intelligent assessment system is a computer-based system designed to assess various aspects of human cognitive and behavioral functioning. The components of an intelligent assessment system typically include innovative user interfaces, cognitive assessments, data acquisition/processing, modeling (machine learning models), user feedback, etc. This chapter will look into the significant components critical in building an intelligent assessment system.

2.1 Cognitive Assessments

A cognitive assessment can be a study or a tool that analyzes the brain's functioning. In other words, it monitors and evaluates the processes involved in thoughts inside the brain. Cognitive assessment not only refers to the use of IQ tests but to any ability test designed to identify cognitive processing deficits that influence academic skills. A comprehensive cognitive test is essential in understanding a person's cognitive skills as well as building systems that can make better decisions about interventions whenever required [52].

While measuring cognitive functions, the approaches can either have the testing procedures administered by trained technicians in a laboratory or clinical setting or use mobile technology and sensors that enable the participants to perform the tests in uncontrolled and naturalistic settings [123]. The National Institutes of Health (NIH) Toolbox [44] gives us a set of cognitive and physical measures designed for use in studies of aging, disease, and rehabilitation. Some of the tests for cognitive abilities in the NIH Toolbox include Dimensional Change Card Sort Test (DCCS) to assess executive function, Pattern Comparison Processing Speed Test (PCPS) to assess processing speed and visual attention, etc. Furthermore, there are several cognitive assessment techniques designed by psychologists such as Sequence Learning Task, Wisconsin Card Sorting Task, N-back task, etc. Some of these techniques have been applied in this dissertation proposal.

2.1.1 N-Back Task

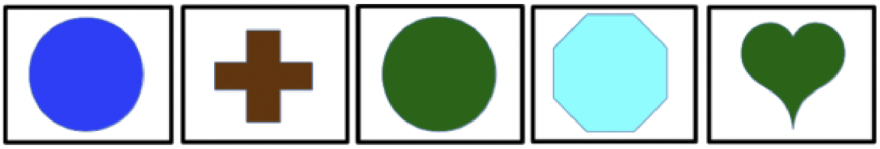


Figure 2.1: An example of the N-back (2-back) Task where a sequence of different shapes is shown to the user as stimuli.

The N-back task is a cognitive task commonly used to assess a person's working memory capacity and attention [67]. It typically involves presenting a series of stimuli (such as letters, numbers, or shapes) and asking the person to indicate when the current stimulus is identical to a stimulus presented ”n” steps earlier in the sequence. For example, in a 2-back task, the participant would need to press a button whenever the current stimulus is the same as the stimulus presented two steps earlier. The N-back task is generally used in cognitive psychology research to study the neural basis of working memory. It has been used to assess the effects of drugs, brain injury, and diseases like ADHD and working memory. It has also been used in the past to induce cognitive load, and fatigue [56].

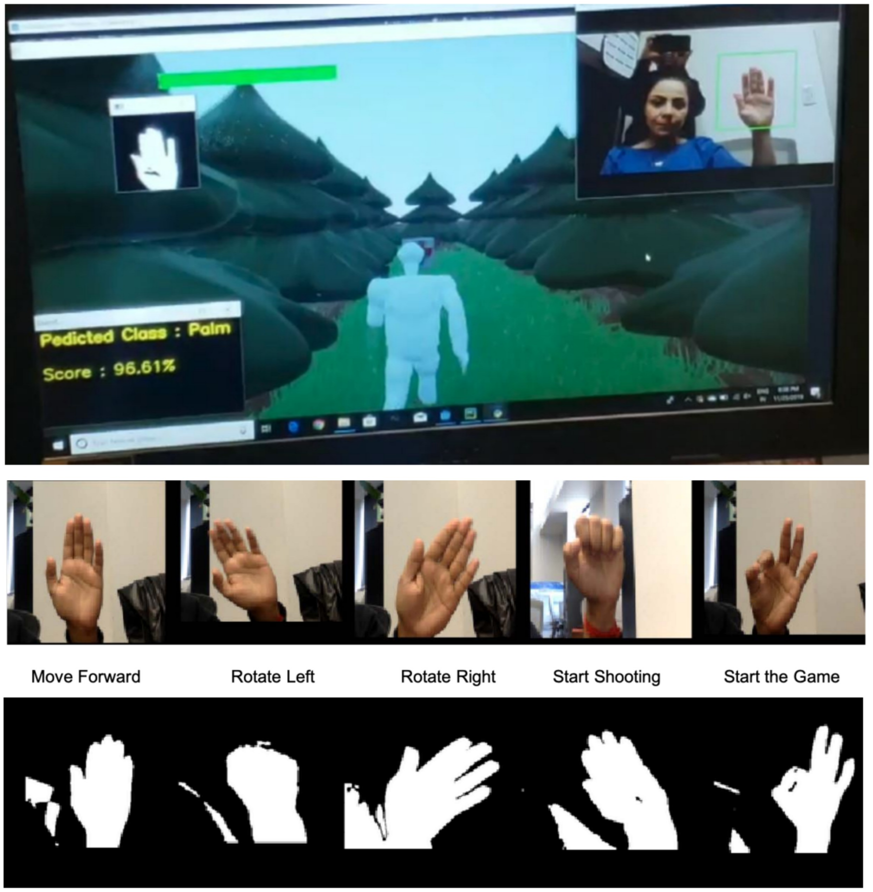
Several versions of the N-back task exist where researchers have used alphabets and numbers for sequences. Furthermore, in some cases, the visual stimuli appears on different screen positions.

2.2 Intelligent Interfaces

Intelligent interfaces are a crucial component in Human-Computer Interaction (HCI). These interfaces are designed to create a more seamless and intuitive experience for users when interacting with computers and digital devices. Furthermore, intelligent interfaces aim to make these interactions as natural as possible so that users can access information and perform tasks without feeling frustrated or overwhelmed.

There are two practical approaches to human-computer interaction: direct manipulation and intelligent agents (also known as delegation) [120]. The first approach relies on input from the user while the computer passively waits, whereas the latter course deals with the computer taking over smartly whenever necessary. Human-Computer Intelligent Interaction (HCII) bolsters the interface with smarts (in the means of sensors and intelligent algorithms), thus improving the overall user experience [105]. Hence, it is essential to design an interface with sensors that can intelligently respond to user inputs and adapt based on the user profile.

In one of our previous works [39], we built a hand-gesture recognition system that acts as an intelligent interface to control an avatar in a game using hand



Pedicted Class : Palm

Score : 96.61%

Figure 2.2: Example of an Intelligent Interface: A hand-gesture based game for wrist rehabilitation. The user controls an avatar in the game with different clinically approved hand gestures for wrist rehabilitation [39]

gestures recorded through a web camera. First, we picked several pre-defined gestures from a pool of clinically approved gestures that can be used for wrist rehabilitation. Then, deep learning models were trained to perform real-time hand gesture recognition, making it an intelligent interface.

2.2.1 Computerized Tests

Some psychological assessment tests utilize computer technology to administer, score, and interpret various measures of cognitive abilities. These tests are designed to measure different aspects of cognitive functioning such as attention, memory, executive function, processing speed, and language abilities. Some common examples of computerized cognitive assessments include the Continuous Performance Test (CPT), Digit Span Test, Stroop Test, and the Wisconsin Card Sorting Test (WCST).

Computerized tests have several advantages over traditional paper-and-pencil tests, including increased standardization, reduced administration time, the ability to present stimuli in a controlled and consistent manner, and improved scoring accuracy and reliability. They are often used in clinical settings to diagnose cognitive and neurological disorders, assess the effect of brain injuries, or monitor the progression of degenerative diseases [102].

2.3 Data Acquisition and Processing

This dissertation follows a data-driven approach involving several user studies that comprise data collection, analysis, and modeling from human subjects. Data collection consists in capturing data from the person being assessed through various inputs, such as touchscreens/keyboards, gestures, voice recognition, etc. Similarly, the data collected during assessments include responses to questions, time taken to complete tasks, and other metrics relevant to the evaluation.

The raw data collected throughout the assessment sessions need to be preprocessed to remove any noise or inconsistencies, correct errors and prepare it for further analysis. This can involve techniques such as data normalization, filtering, and outlier detection. Finally, the processed data is analyzed to determine a person's cognitive abilities and to generate results. It involves statistical methods, machine learning algorithms, or other data analysis techniques. Our user studies collect physiological and behavioral data, which are explained below.

2.3.1 Wearable Sensors

Physiological data is measured using sensors capable of measuring the autonomic nervous system's involuntary response to stimuli [83]. Wearable sensors are devices worn on the body to measure physiological signals, such as heart rate, body temperature, and movement. These sensors use various technologies, such as accelerometers, gyroscopes, and photoplethysmography (PPG), to gather data about the user's physiology. Wearable sensors have several applications in health and wellness, including physical activity monitoring [76], sleep tracking [119], stress management [25], and chronic disease management [50].

This research uses data from EEG, ECG, EDA, EMG, and breathing sensors which are capable of recording user data non-invasively. ECG, EDA, and EMG sensors that we used are part of a wearable sensing kit called Biosignalplux [14]. These sensors are further described in the following subsections.

MUSE EEG Headset

EEG is a method of measuring electrical activity in the brain and in widely used in the field of neuroscience and MUSE EEG headset is a wearable device that measures brain activity using electroencephalography (EEG) technology. It is a non-invasive wearable device widely used for Brain Computer Interface Systems [9].

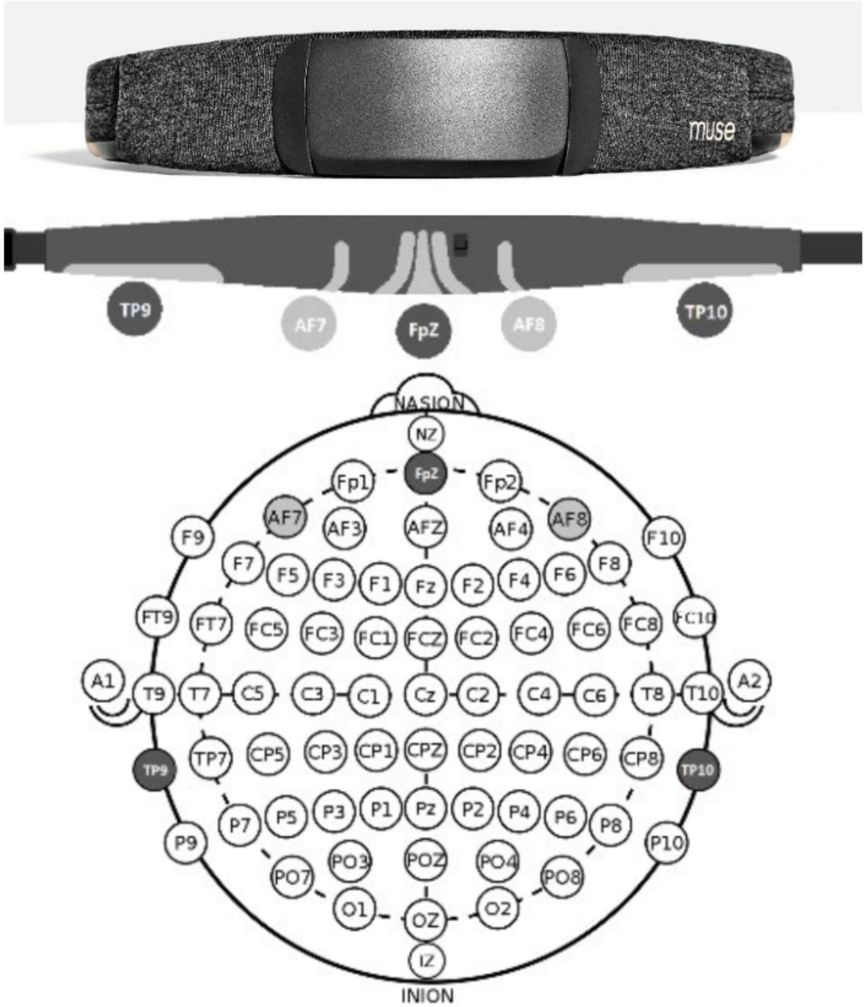


Figure 2.3: Electrode placement comparison between MUSE and the international 10-20 system. Top: Commercial MUSE S Headband. Bottom: 10-20 Electrode Placement System

The MUSE EEG headset consists of a headband with electrodes as shown in Figure 2.3 that are placed on the scalp and forehead to detect electrical signals generated by brain activity. These signals are then processed and analyzed by the headset to provide information about the wearer′s brain activity, including brainwave patterns, attention levels, and meditation levels.

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 its embedded signal processing unit, we can store the features extracted from individual EEG frequency bands namely: gamma (g) 32-100 Hz, beta (b) 13-32 Hz, alpha (a) 8-13 Hz, theta (t) 4-8 Hz, and delta (d) 0.5-4 Hz at a sampling rate of 10 Hz. For each of the four electrodes, we record different types of data streams such as Raw EEG, Absolute Frequency Bands (A), Relative Frequency Bands (R), Session Score (s) for each frequency band, and Signal Quality Indicator (h).

- **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.

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

where x is one of the frequency bands.

- **Session Score for each Frequency Band** *(s)*: It is a value computed by comparing the current value of a band power to its history in sampling frequency of 10 Hz. The 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.
- **Signal Quality Indicator** *(h)*: It is an integer value from 1 (optimal quality) to 3 (the worst quality)

Electrocardiogram (ECG)

ECG sensors measure the electrical activity of the heart and record voltage over time. As shown in Figure 2.4 (a), the ECG graph is divided into three parts. The first part is the P wave, which represents the atria's depolarization; the second is the QRS complex, which means the ventricle's depolarization. Finally, the last part is the T wave representing the repolarization of the ventricles [122].

The most common way to record ECG is using Einthoven's triangle [38]. It dictates three locations on the body that represent a triangle where surface electrodes of the sensor should be attached: Right Arm (RA), Left Arm (LA), and Left Leg (LL), as shown in Figure 2.4 (b). This setup is grounded using a fourth electrode attached to the right leg. The ECG signal helps to monitor the cardiovascular system and is vital when investigating cognitive workload and fatigue [98].

Figure 2.5 shows a sample snapshot of the ECG signal after pre-processing.

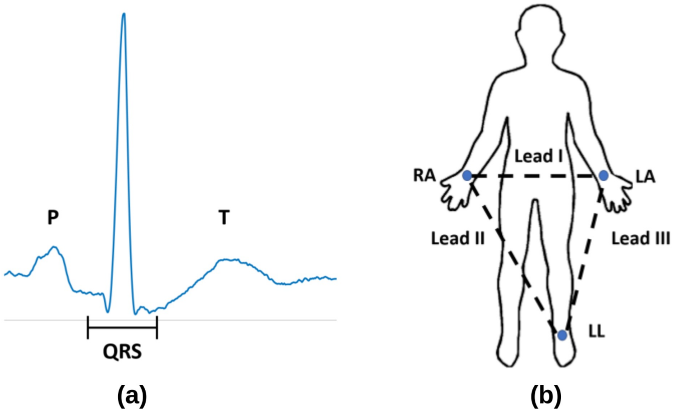


Figure 2.4: **(a)** ECG Waveform. P wave, QRS complex, and T wave. This waveform is repeated throughout the ECG signal. Analysis of this waveform helps us study the physiological changes in the heart. **(b)** Einthoven's Triangle: Right Arm (RA), Left Arm (LA), and Left Leg (LL). This electrode configuration is followed to collect ECG signals.

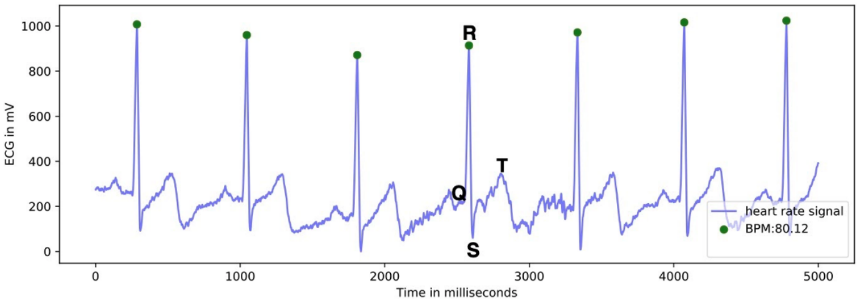


Figure 2.5: Sample ECG signal snapshot

Since ECG is measured in millivolts, it needs a unit conversion as follows:

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

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

where

- $ECG(V)$ is the ECG value in volt (V)
- 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 3V
- G_{ECG} is the sensor gain equal to 1000
- $ECG(mV)$ is the ECG value in millivolts

Electrodermal Activity (EDA) / Galvanic Skin Response (GSR)

EDA sensors measure the skin conductivity of the body. For example, when a person sweats, the skin's conductance improves. This helps to identify episodes of psychological and emotional arousal. In addition, studies have shown the impact of fatigue on the activities of the Sympathetic Nervous System (SNS) [42]. Hence, observing skin conductance helps to study both cognitive and physical fatigue. EDA signals are measured in microsiemens (μS) as shown in a sample snapshot in Figure 2.6.

Electromyogram (EMG)

In EMG, the contraction and relaxation of a muscle is recorded as a change in voltage between two electrodes. The electrodes must be placed at either end on the longitudinal midline of the muscle. EMG has been mostly used in detecting physical fatigue by analyzing reduced muscle activation [29]. It has been studied that the median frequency of EMG signals decrease due to fatigue. We have tried to detect physical fatigue in some of this work, especially using the EMG signals. Figure 2.7 illustrates a sample snapshot of the EMG signal over time.

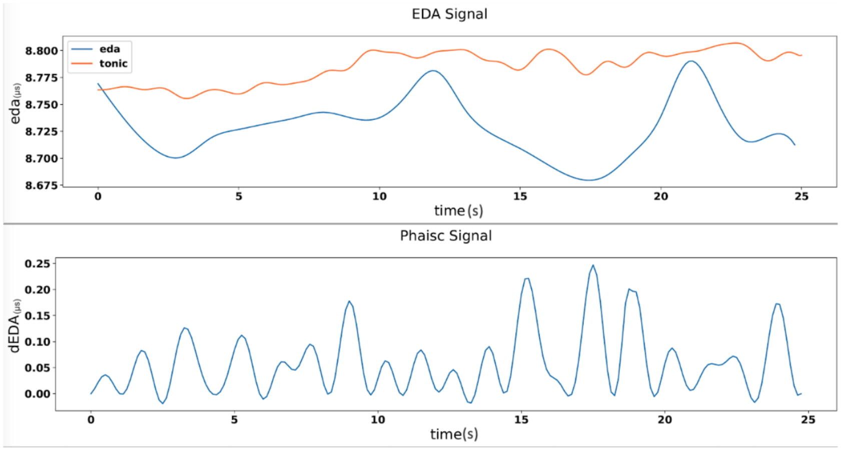


Figure 2.6: EDA signal, Raw EDA and tonic component signal snapshot(top). Phasic component signal snapshot(bottom)

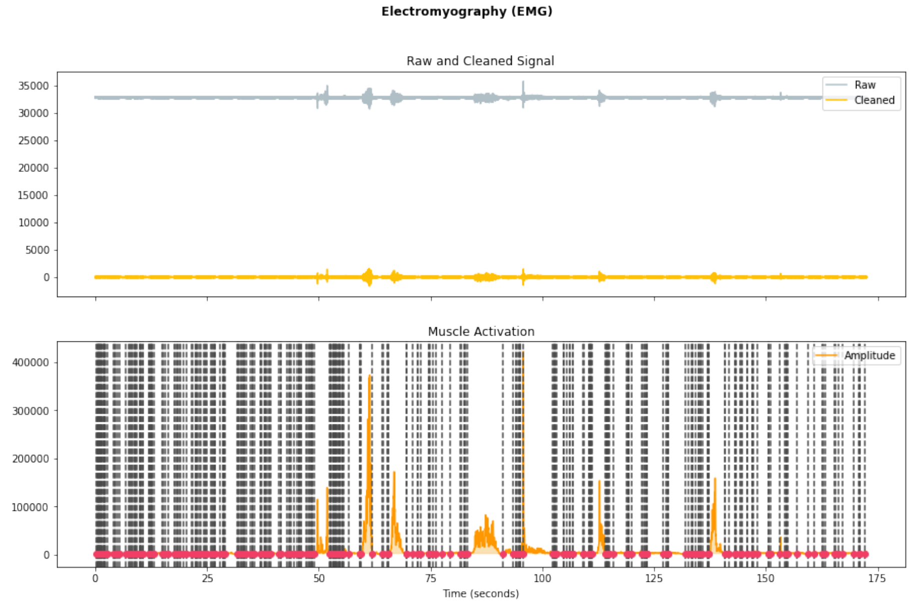
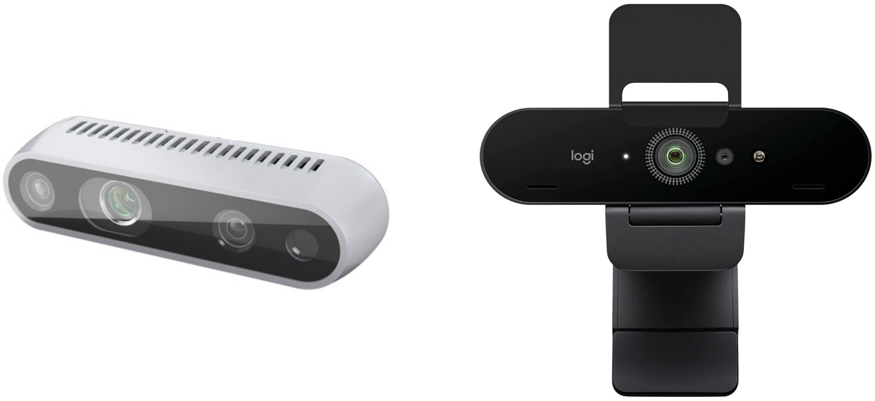


Figure 2.7: EMG Signal. This is a sample segment of the EMG signal. The signal represents the electric potential across a muscle. If the muscle is activated, the amplitude of the signal increases.

2.3.2 Vision Sensors

Physical reactions to stimuli, such as body postures and facial expressions, are classified as behavioral data. In this work, we have primarily used an RGB webcam and RGB-D Intel RealSense sensors to capture the behavioral activity of the users. Figure 2.8 shows two cameras that have been used in this work.

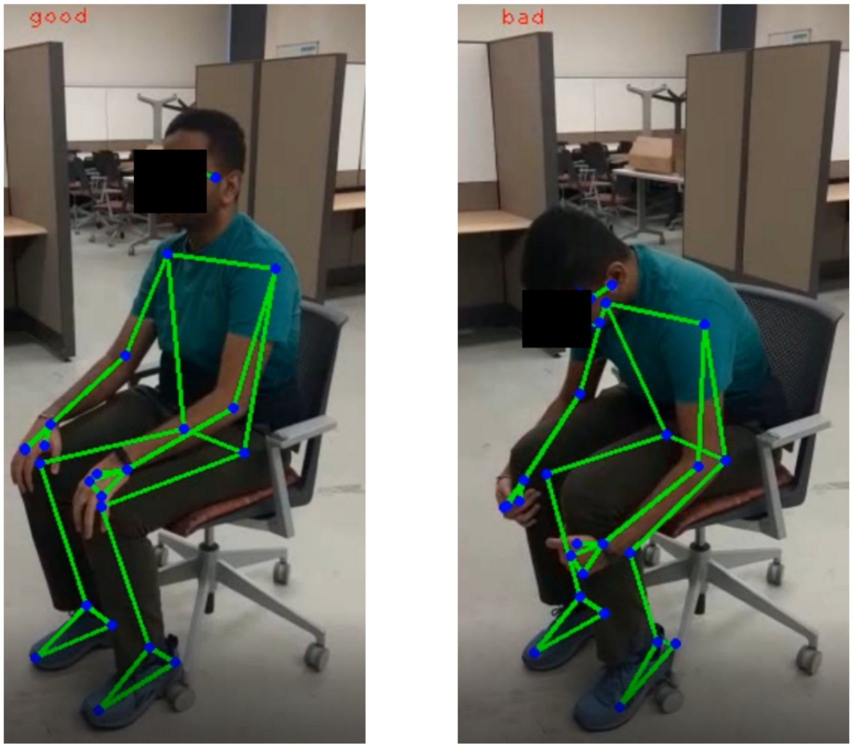
For instance, in one of our works [69], we built a lightweight end-to-end system that monitor's the subject's posture and provides feedback such as fixing their posture or taking a break whenever it is required. The system utilizes an RGB camera to input frames as shown in Figure 2.9 and a machine learning model is trained to distinguish between good and bad posture.



(a) Intel RealSense D435i **(b) Logitech BRIO 4K Webcam**

Figure 2.8: Visual sensors for recording and collecting behavioral data

Lousy posture for prolonged periods can result in numerous health issues such as back pains, moderate discomfort in eyes/neck/head, upper back/shoulders, and elevated stress levels as highlighted in [3]. In addition, bad posture can also represent a symptom of physical fatigue. According to Chavalitsakulchai et al. [24], fatigue consists of unpleasantness as an aversion to work, desire for rest, impatience, and physical, mental, and neuro-sensory feelings of incongruity that workers experience. Furthermore, the authors mention that physical fatigue can impact psychological or cognitive fatigue. Hence it is essential to study posture to get insights into the overall fatigue of the subject.



(a) Posture classified as good by the NN model.

(b) Posture classified as bad by the NN model.

Figure 2.9: A neural network (NN) model classifying good and bad postures using the body pose skeleton information extracted from the input frames.

2.4 Role of Machine Learning in Cognitive Assessment Frameworks

Machine learning is a subset of artificial intelligence that involves the use of algorithms and statistical models to enable a system to ”learn” from data, without being explicitly programmed. In the context of cognitive assessment, machine learning can be used to analyze patterns in data collected from individuals to make predictions about their cognitive abilities, such as memory, attention, problem-solving, and decision-making skills.

Machine learning can be used to analyze physiological and behavioral features collected from wearable sensors and other sources to gain insight into an individual's cognitive abilities and functions. For example, several works have been done to analyze heart rate variability [92], eye movements [80], and posture data [47] to predict an individual's level of stress, attention, and emotional state using machine learning algorithms. This information can be used to support the diagnosis and treatment of cognitive impairments, monitor the progression of cognitive disorders, and evaluate the effectiveness of cognitive rehabilitation programs.

In addition, machine learning can be used to develop personalized models of human cognition based on an individual's physiological and behavioral data. These models can be used to understand how different cognitive processes and behaviors are related to one another and how they vary across individuals. It also allows for the analysis of large amounts of data, the identification of patterns and relationships, and the development of personalized models of human cognition. These capabilities have the potential of to revolutionize our understanding of human cognition and have applications in fields such as healthcare, psychology, and sports performance.

2.5 Multi-modal analysis using Machine Learning

Utilizing multi-modal analysis in machine learning involves leveraging diverse data types (modalities) to forecast specific phenomena. For instance, when conducting a cognitive assessment, various data sources like demographic details, test scores, brain imaging data, and self-reported symptoms can be combined to predict an individual’s cognitive abilities.

As previously mentioned, two crucial attributes for evaluating and comprehending human cognition are Physiological and Behavioral features. Furthermore, environmental factors such as light, noise, temperature, and air quality can influence an individual’s cognitive capacity. Consequently, employing machine learning algorithms to process and scrutinize these multiple data sources can enhance our understanding of the interconnections among physiological, behavioral, and environmental factors and their effects on human cognition and fatigue [102].

This dissertation employs multi-modal methodologies to extract information from diverse modalities, aiming to enhance the efficiency of machine learning models and overall assessment frameworks. Detailed discussions on these methodologies are presented in the subsequent chapters.

CHAPTER 3

TASK-BASED COGNITIVE ASSESSMENT FRAMEWORKS

3.1 Activate Test of Embodied Cognition (ATEC)

With improvements in computer technology, there has been a need to assess cognitive abilities with objective measures. Numerous computerized assessments have been proposed in the recent past to assess various cognitive aspects, mainly executive functions. The inclusion of physical exercises in cognitive training is motivated by research illustrating that physical fitness and activity in children lead to measurable improvements in cognitive skills and academic performance [31]. The Activate Test of Embodied Cognition (ATEC) is a computer-based cognitive assessment tool that measures a person's cognitive abilities through physical movements. It is based on the theory of embodied cognition, which suggests that our cognitive processes are closely linked to our bodily experiences and interactions with the environment.

The ATEC system consists of a series of interactive tasks that require the user to make specific movements in response to visual and auditory stimuli. The tasks are designed to assess a range of cognitive functions, including attention, working memory, executive function, and processing speed [11, 35, 7]. It provides a more ecologically valid measure of cognition that traditional paper-and-pencil tasks. The results of the test are automatically scored and compared to normative data to provide an objective measure of the person's cognitive abilities.

Data Collection

The ATEC setup uses Kinect technology to record videos of children performing different physical tasks. The Kinect sensor captures 3D images of the participant's movements, including RGB videos of the scene and depth information. Therefore, two Microsoft Kinect V2 cameras [124] capture the participants' front and side views. In addition, the recording modules are linked to an Android-based administrative interface that enables the management of the task flows and options to switch tasks. Figure 3.1 highlights the data collection setup. Each session comprises an instructional film and tutorial videos to ensure that the subjects understand the rules of the activity. Furthermore, an administrator is present to ensure that the tasks run smoothly.

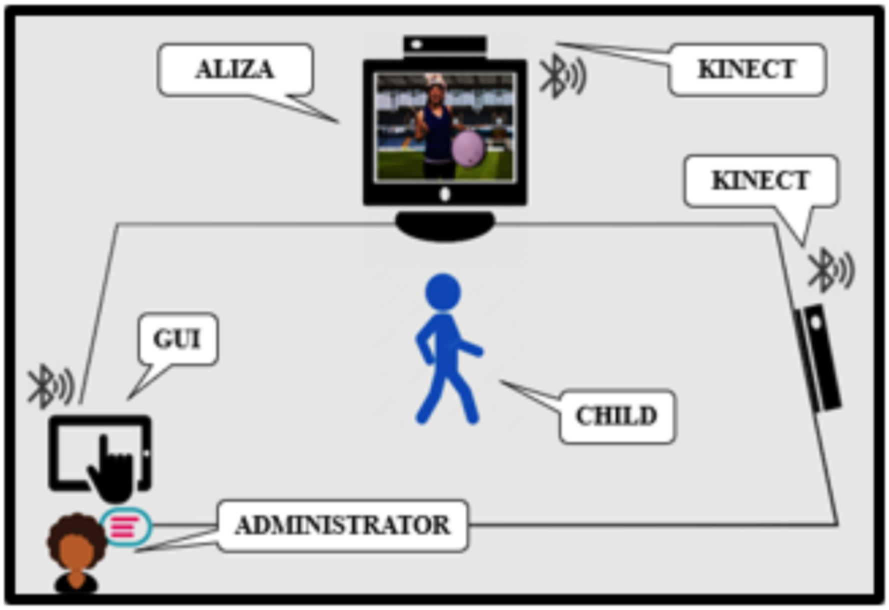


Figure 3.1: **The ATEC Setup**: Participants perform various physical tasks based on the visual and auditory cues provided. It includes two Kinect cameras for recording, a large screen displaying a virtual assistant for instructions, and a tablet interface with a smart GUI for the administrator.

Data was recorded from *N=55* children between the age of 5 - 11 years *(mean=8.04, std=1.36)* in classroom environments. Under procedures approved by the University IRB, the parents supplied written informed consent, and the children offered verbal assent. Although all the children were in regular classes, 9 (16.4%) received additional services through a 504 plan approved by the school. The population was ethnically diverse *(56.4% Caucasian, 58.2% male)* [11].

Pre-screening paperwork is collected to obtain the children's and their family's history. Next, paper-based assessments such as Child Behavior Checklist (CBCL) [1], Social Responsive Scale, Swanson, Nolan, and Pelham questionnaire [6] are carried out. In addition, the participants are requested to complete a couple of standard computerized tests from the NIH Toolbox: Flanker test and Working Memory Test [44]. Finally, the children perform all the tasks from the ATEC program, and data is collected for two trials, each a week apart.

Table 3.1: ATEC tasks to assess various Cognitive Measures

Category	**Test**
Lateral Preference Patterns	**-**
Gross Motor Gait and Balance	Natual Walk, Gait on Toes, Tandem Gait, Stand Arms Outstretched, Stand on One Foot
Synchronous Movements	March Slow, March Fast
Bilateral Coordination and Response Inhibition	Bi-Manual Ball Pass with red, green, and yellow light
Visual Response Inhibition	Sailor Step Slow, Sailor Step Fast
Cross Body Game	Cross your Body (Ears, Shoulders, Hips, Knees)
Finger-Nose Coordination	Hand Eye Coordination
Rapid Sequential Movements	Foot Tap, Foot-Heel, Toe Tap, Hand Pat, Finger Tap, Appose Finger Succession

The ATEC system consists of 17 physical tasks with different variations and difficulty levels that are designed to provide an assessment of executive and motor functions, including sustained attention, self-regulation, working memory, response inhibition, rhythm & coordination, and motor speed & balance as depicted in Table 3.1. This thesis incorporates some of the ATEC tasks that are described below.

3.1.1 Ball Drop To The Beat

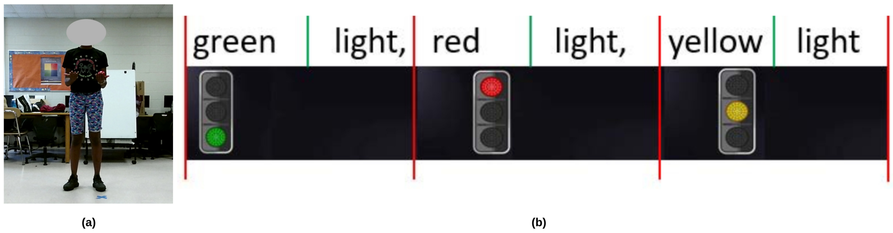


Figure 3.2: (a) A child performing Ball Drop to the Beat Task with ball on one hand at a time (b) Audio-visual stimuli for instructions. Green light: **Pass** the ball, Yellow light: **Raise** the hand, and Red light: **No Pass**. [35]

Ball Drop to the Beat is a core ATEC task devised to assess bilateral coordination, and response inhibition [11, 7]. It evaluates audio and visual cue processing while performing a physical task involving upper body movements. In this task, the participant must pass a ball from one hand to the other by following verbal and visual instructions. According to the rules, a participant must pass the ball from one hand to another when a green light is shown (Pass). If there's a yellow light, they must raise the hand that has the ball (Raise). Finally, they remain still on a red light with no movement at all (No Pass). Figure 3.2 (a) illustrates a child performing the task with a ball in their hand by following the audio-visual stimuli shown in Figure 3.2 (b).

The task is assessed at 60 beats per minute (slow trial) and 100 beats per minute (fast trial) for a total of 16 counts for each trial [11]. In addition to accuracy and response inhibition, this task also measures rhythm. A virtual avatar **Aliza** on the TV screen rhythmically announces the stimuli by saying *Green Light*, *Red Light*, and *Yellow Light* in two beats. The first beat represents the color, and the second beat means the word light. Therefore, the subjects are expected to perform the task in two beats. For instance, in the case of yellow light, the participant raises the ball on the first beat and lowers it on the second. As shown in the visual stimuli in Figure 3.2 (b), each segment (activity) is divided by red lines (two beats), and each segment contains two beats separated by green lines.

Attention and Response Inhibition from the task

The score for Response Inhibition (RI) is calculated as

$$RI = \frac{\text{No. of correct } \textbf{\textit{No Pass}} \text{ / Red Light Actions}}{\text{Total No. of } \textbf{\textit{No Pass}} \text{ / Red Light Commands}} \tag{3.1}$$

Similarly, the score for attention (Attn.) is calculated as:

$$Attn = \frac{\text{No. of correct } \textbf{\textit{Raise}} \text{ \& } \textbf{\textit{Pass}} \text{ actions}}{\text{Total No. of } \textbf{\textit{Raise}} \text{ and } \textbf{\textit{Pass}} \text{ Commands}} \tag{3.2}$$

Body Skeleton Based Approach

Initially, we use a simple body skeleton-based approach to automate the Ball Drop to the Beat Task scoring. During the task, the videos are recorded at 30 frames per second (FPS). Hence, the video is decoded into individual frames, and body keypoints are extracted as features from each edge. 2D/3D human poses acting as trajectories of skeleton joints are one of the most effective representations for characterizing the dynamics of human actions. Each coordinate in the skeleton is referred to as a joint or a keypoint, and a valid connection between any two keypoints is referred to as a limb or a pair.

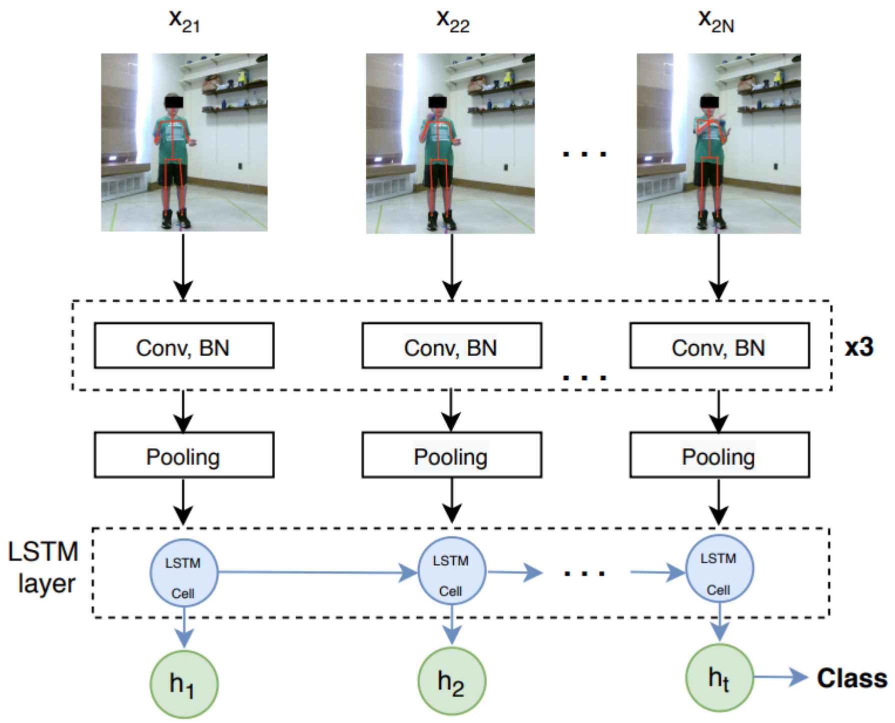


Figure 3.3: Body skeleton keypoints based approach using an Long Short-Term Memory (LSTM) recurrent neural network to predict the movement performed by the participant along with the rhythm [110].

We use the convolutional neural network (CNN) based frameworks OpenPose [19], and VIBE [73] to extract the body key points. These feed-forward networks predict 2D confidence maps of the body's joints' locations and a set of 2D vector fields of part affinity fields which is the degree of association between the parts. It is a top-down method that first detects humans in the scene and performs pose estimation for each detected subject. Furthermore, noise such as incorrect detection of keypoints or missing keypoints is rejected before using the algorithm in Figure 3.3.

In our approach, as shown in Figure 3.3, for a video segment containing *n* frames, 18 keypoints are extracted from each frame that represent various body joint positions, including facial keypoints such as eyes, ears, and nose. Since the task only focuses on the movements of the upper body parts, we selected only 9 out of the 18 keypoints that include the upper body joints. Each keypoint is represented as a 3D coordinate (*z, y, v*) on the image plane. For instance, for a give frame *P* at time *t* is represented by the coordinates of the *nine* keypoints as follows:

$$P_t = [(z_{1,t}, y_{1,t}, v_{1,t}), (z_{2,t}, y_{2,t}, v_{2,t}), ..., (z_{9,t}, y_{9,t}, v_{9,t})] \tag{3.3}$$

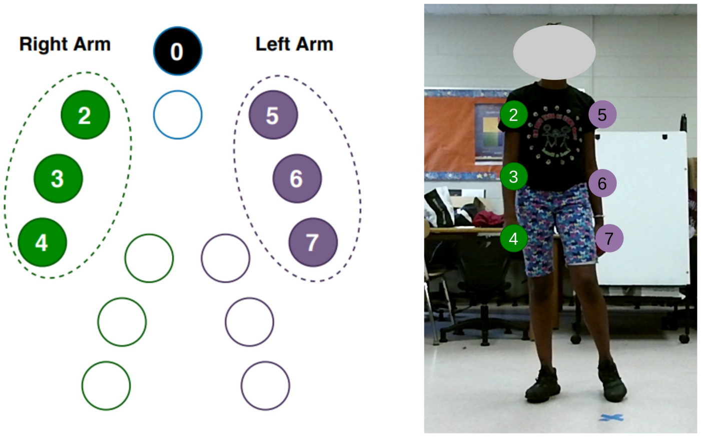
where *z* represents the coordinate extending from left to right (horizontal axis), *y* extending from top to bottom (vertical axis), and *v* represents the depth for each keypoint.

The proposed subnet to extract spatial and temporal features from skeletal points is comprised of a series of 1D convolutional layers and batch normalization followed by a pooling layer as shown in Figure 3.3. A single layered Long Short-Term Memory (LSTM) unit with only one hidden state (*h*) of dimension *32* is used to capture the temporal relation among the frames in the sequence. The network is trained with a softmax layer at the end, providing a 3-fold test accuracy of **74.7 %** for the 3-class classification problem. Furthermore, the rhythm accuracy obtained using this model is **62.8 %** [110].

Multi-modal Deep Learning Approach

Fusing multiple modalities and features from a visual scene have been proven to improve human activity recognition (HAR). For instance, Franco et al. [41]

fused skeleton-based features with video-based features such as histogram of oriented gradients (HOG) to improve the performance of HAR tasks. Similarly, Kapidis et al. found out that combining hand and object detection to recognize human actions from an egocentric view camera enhances the accuracy of a HAR system [68]. In this work, we combine three different modalities for our HAR task (Ball Drop to the Beat with three activity classes): **human pose** (skeleton-based), **optical flow** (video-based), and **object detection** (detection of the ball in hand).



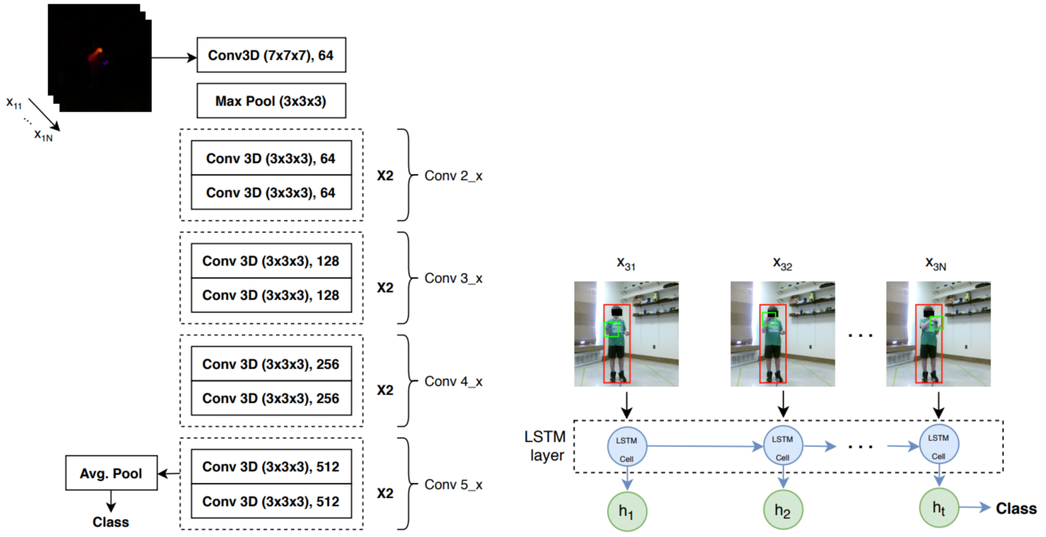
(a) Body Keypoints from OpenPose

(b) Body Keypoints selected for multi-modal approach

Figure 3.4: Body keypoints considered in the multi-modal approach for action and rhythm detection in the Ball Drop to the Beat task when using the body skeleton-based approach [110]

Since the task involves movement of the upper body parts only, nine keypoints, as shown in Figure 3.4 in addition to some facial keypoints such as eyes and nose, are considered for multi-modal fusion. During the fusion process, features h_t, which is the last hidden state of the last LSTM block, are extracted and used.

Optical flow has been rigorously used as input features for HAR tasks as it captures the motion information between consecutive frames [117]. First, opti-



(a) HAR based on Optical Flow features

(b) HAR based on detected ball coordinates

Figure 3.5: Human Activity Recognition (HAR) based on individual modalities: (a) Optical flow (b) Coordinates of the ball in hand using object detection

cal flow is computed from the recorded videos of the participants performing the task using the off-the-shelf implementation of [17] from the OpenCV toolbox. Next, a deep neural network-based architecture inspired from [53] is used to extract meaningful information from the optical flow segment as shown in Figure 3.5 (a). The dotted blocks in Figure 3.5 (a) represent residual blocks, and a batch normalization operation follows every convolution operation to reduce the internal covariate shift.

Finally, the system could benefit from the positional information about the ball in one of the hands. The ball is first recognized in the scene at coordinate $o_i = \{l_i, s_i\}$ comprising of a bounding box l_i and its category $s_i \in S$, where S is the set of all possible object categories (e.g., ball, person) being encoded in the form of Binary Presence Vector (BPV) and i ranges from 0 to k with k representing the total number of objects detected in the scene. A popular object detection algorithm YOLO V3 [112], is used to detect objects. Any missing objects in a given frame t are fixed with information from the previous frame $t - 1$.

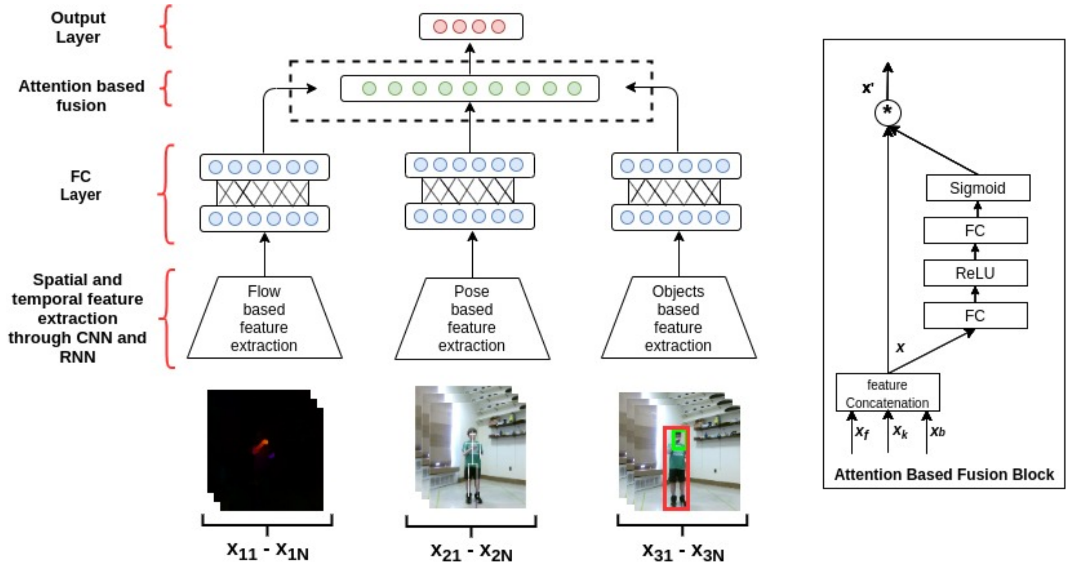

Output Layer
Attention based fusion
FC Layer
Spatial and temporal feature extraction through CNN and RNN
Flow based feature extraction
Pose based feature extraction
Objects based feature extraction
x_{11} - x_{1N}
x_{21} - x_{2N}
x_{31} - x_{3N}
x'
*
Sigmoid
FC
ReLU
FC
x
feature Concatenation
x_f
x_k
x_b
Attention Based Fusion Block

Figure 3.6: Self-Attention based Multi-modal Fusion Algorithm to combine information from optical flow, human pose, and object detection for Human Activity Recognition (HAR)

As explained earlier in section 2.5, not all modalities and features contribute equally towards the predicted result in a multi-modal fusion algorithm. Hence, identifying the highest contributing modalities and prioritizing them while training a machine learning model can improve the overall performance of a system. Therefore, a self-attention-based fusion approach inspired from [57] is proposed as shown in Figure 3.6. In this approach, every feature within a modality is associated with a corresponding weight vector learned during the training process based on its impact on the results. To calculate the weights of features from each modality, first, all the features are concatenated into a single vector as follows:

$$x = [x_f, x_k, x_b] \tag{3.4}$$

where $x_f \in R^{C_f}$ is the feature vector obtained from the optical flow subnet (Figure 3.5 (a)), $x_k \in R^{C_k}$ is the feature vector from the pose subnet (Figure 3.4), $x_b \in R^{C_b}$ is the feature vector from objects position based subnet (Figure 3.5 (b)), and $x \in R^C(C = C_f + C_k + C_b)$ comprises all features from all modalities.

Next, F_w is introduced as represented in equation 3.5 to calculate the attention weights for features of x. For F_w to fully capture feature-wise dependencies, it should meet two criteria. First, it must be capable of learning nonlinear interaction between features. Second, it must retain a non-mutually-exclusive relationship that ensures multiple features can be emphasized. Hence, a gating mechanism with a sigmoid activation is employed.

$$\alpha = F_w(x, W) = \sigma(g(x, W)) = \sigma(W_2\delta(W_1x)) \tag{3.5}$$

where δ refers to the ReLU activation function [97], $W_1 \in R^{\frac{C}{r}\times C}$ and $W_2 \in R^{\frac{C}{r}\times C}$. The gating mechanism is parameterized by forming a bottleneck with two fully-connected (FC) layers ($W_1\&W_2$) around the non-linearity, i.e., a dimensionality reduction layer with reduction ratio r, a ReLU, and a dimensionality increasing layer returning to the original feature dimension of X. The final output is obtained by the element-wise product of combined feature vector X and calculated attention weights vector α:

$$x' = F_a(x, \alpha) = \alpha x \tag{3.6}$$

where x' represents the output of the attention block with the features from the modalities combined and weighted, which is succeeded by a softmax layer for the final prediction.

Results and Discussion

Several methodologies have proposed a multi-modal approach to achieve state-of-the-art results on existing popular action recognition datasets, as shown in Table 3.2. For the 3D convolution-based approach, ResNet with variable depth sizes (18, 34, 51) and inception model was trained. Although, it was observed that as the depth of the model increased, the model started to overfit. Hence, the results shown are only for resNet-18. For the two Stream I3D, an inception-based model was trained for RGB-based and Optical flow-based sequences. During testing, the outcome of both models was combined for the final prediction. The hyper-parameters for initial training were used as recommended by the authors of the papers, followed by fine-tuning in the later trials.

Table 3.2: Comparison of **our proposed method** v/s **existing state-of-the-art** approaches. The results are averaged over 5-folds. **KP** - Key points, **flow** - Dense optical flow, **RGB** - RGB image frames, **Object Pose** - Objects in the scene

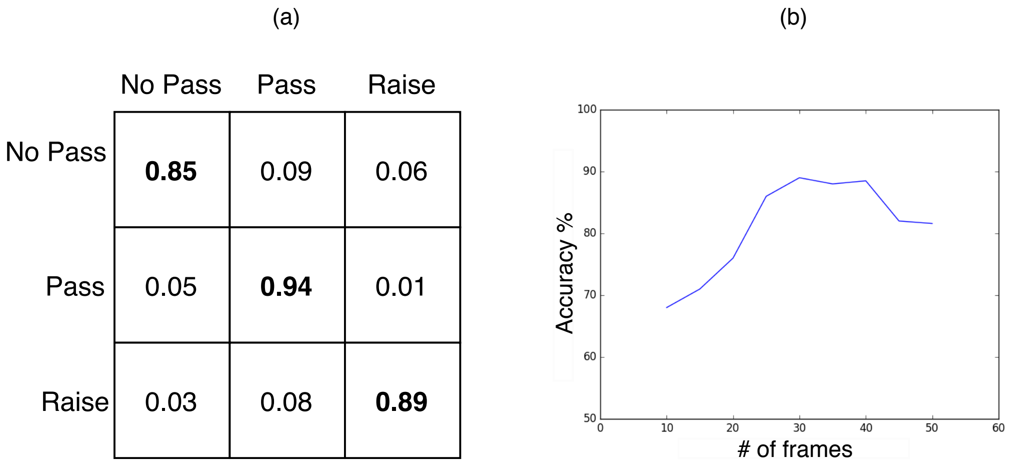
Method	**Test. Acc.**	**Features**
3D CNN [53]	0.730	RGB
Two Stream I3D [20]	0.825	RGB+flow
CNN + RNN(LSTM) [43]	0.690	RGB
DeepGRU [88]	0.610	KP
Dillhoff et. al. [35]	0.780	KP
Attnsense [87]	0.810	RGB
Proposed approach	**0.898**	KP+Object Pose+flow

For training, the dataset was split into training, validation, and testing sets based on the subjects. It was done to ensure that video segments of the same subjects were not present in both training and validation/test sets to influence the results. The validation was performed after every epoch of training to identify the right epoch to stop the training and avoid overfitting. The model was evaluated on the test set at the end of training. Stochastic Gradient Descent based optimization with momentum was used during the training. Since the dataset is comparatively smaller than the other publically available datasets, extensive temporal and spatial augmentation was performed during the training. A video clip of size t is generated with a randomly selected temporal position as the starting frame. If the video is shorter than *t* frames, it's looped through until it matches the size *t*. For spatial augmentation, a spot is randomly chosen between four corners and the center of the image, and multi-scale cropping is performed, after which the images are spatially resized. The cross-entropy loss was used during training with starting learning rate set to 0.0001 and divided by ten every time the validation loss saturates, with a weight decay of 0.001 and 0.9 for momentum. To train the models, four NVIDIA GTX 1080 Ti GPUs were used, whereas, for testing, only one GPU was used.

Table 3.3: Ablation Study: Experimental results for multi-modal approach [110]. All results are averaged over 5-folds.

Method	**Test.**	**Time(Sec.)**
Optical Flow (Flow)	72.0%	0.229
Body Keypoints(KP)	76.0%	0.106
Objects Trajectories (Obj_Pos)	68.0%	0.103
Flow+KP(natural-concat.)	82.0%	0.236
Flow+KP(balanced-concat.)	83.9%	0.239
Flow+KP(Self-Attn.)	84.6%	0.240
Flow+Obj_Pos (natural-concat.)	84.1%	0.232
Flow+Obj_Pos (balanced-concat.)	83.9%	0.236
Flow+Obj_Pos (Self-Attn.)	84.0%	0.241
KP+Obj_Pos (natural-concat.)	79.0%	0.118
KP+Obj_Pos (balanced-concat.)	76.3%	0.123
KP+Obj_Pos (Self-Attn.)	79.5%	0.139
KP+Obj_Pos+flow (natural-concat.)	89.0%	0.254
KP+Obj_Pos+flow (balanced-concat.)	87.5%	0.259
KP+Obj_Pos+flow (Self-Attn.)	**89.8%**	**0.260**

Extensive training was performed using different combinations and different fusion approaches to identify the effective combination of the modalities



	No Pass	Pass	Raise
No Pass	**0.85**	0.09	0.06
Pass	0.05	**0.94**	0.01
Raise	0.03	0.08	**0.89**

Figure 3.7: (a) Normalized confusion matrix of our proposed method [110]. (b) Graph representing model accuracy as a function of the number of frames.

used in our approach. Table 3.3 an ablation study and the results of the experiments. All results were averaged over 5-folds. Similarly, to fuse the features from individual modalities, in addition to the approach depicted in Figure 3.6, other approaches were also attempted. The natural concatenation (natural-concat) is a vanilla approach where output features of different modalities are directly concatenated, followed by a softmax layer to classify the actions. On the other hand, balanced concatenation (balanced-concat) aims to convert the feature vectors from different modalities into the same dimensional size, followed by concatenation and a softmax layer. As the goal was to deploy the proposed system for future data collection, it was essential to measure the execution time of the model, which is also presented in Table 3.3, primarily when multi-modal approaches are used. Figure 3.7 (a) illustrates the normalized confusion matrix of the proposed method on the test data to predict actions.

From Table 3.3, it can be observed that the body keypoint-based model achieves the highest accuracy as a single modality. However, the accuracy is lower than required for an assessment system. Although using three modalities produced satisfactory results compared to the previous works for action recognition, extensive tests were necessary with different modalities and fusion strategies to find the optimal solution. No other variety of modalities and fusion methods seemed to outperform our proposed method. Adding object detection as an additional modality improved the accuracy by 5.2% for attention-based fusion. Moreover, the combination of optical flow and object position and optical flow and body keypoints provide similar accuracy, and it is higher than the combination of keypoints with object position.

These results verify the importance of using optical flow as an additional modality to any multi-modal algorithm. Furthermore, irrespective of the combination of the available modalities, the attention-based fusion seems to work better almost every time, as shown in Table 3.3. As expected, combining all three modalities yields the highest execution time of 0.2603 seconds, although producing the best prediction results. Since the assessment system does not require real-time processing of the image frames, this approach works best in this scenario as the execution time is acceptable. In Figure 3.7 (b), it is clear that as the number of frames increases, the model's performance drops linearly. However, it stagnates after a certain number of frames is reached, possibly due to overfitting of the model, as the frames are repeated and duplicated in the loop during our proposed temporal augmentation method.

3.1.2 Tandem Gait Forward

The tandem gait test requires the individual to walk along a straight line heel to toe, with one foot in front of the other, placing the heel of one foot in front of the toes of the other foot. The person performing the task must maintain balance and walk a certain distance in a straight line without falling or stepping off the line. The tandem gait primarily assesses motor coordination and balance, which the cerebellum controls. In addition, it requires the individual to have good executive functions to complete the test [11].

Several machine learning methods have been proposed for gait assessment through the estimation of spatio-temporal parameters on pathological populations impacted by Huntington's disease and post-stroke subjects as well as healthy elderly controls [91]. Furthermore, wearable sensor technologies have been employed to monitor parameters that characterize mobility impairments such as gait speed outside of the clinic [93].

In our proposed task [128, 11], the participants are asked to walk in a straight line and complete eight valid steps based on the abovementioned criteria. A dataset containing 27 children performing the test has been created. In addition, a computer vision assessment system that only requires one camera to monitor the gait movements has been devised.

Data Collection

This task is one of the 17 physical tasks from the ATEC system. Hence, the data collection setup is as shown in Figure 3.1. The side view of the participants is recorded while they perform the task, and a total of 27 children (aged 5-11 years) were involved. In each session, the child is asked to perform eight valid steps of the Tandem gait task. A step is considered valid only if one foot's heel touches another foot's toe.

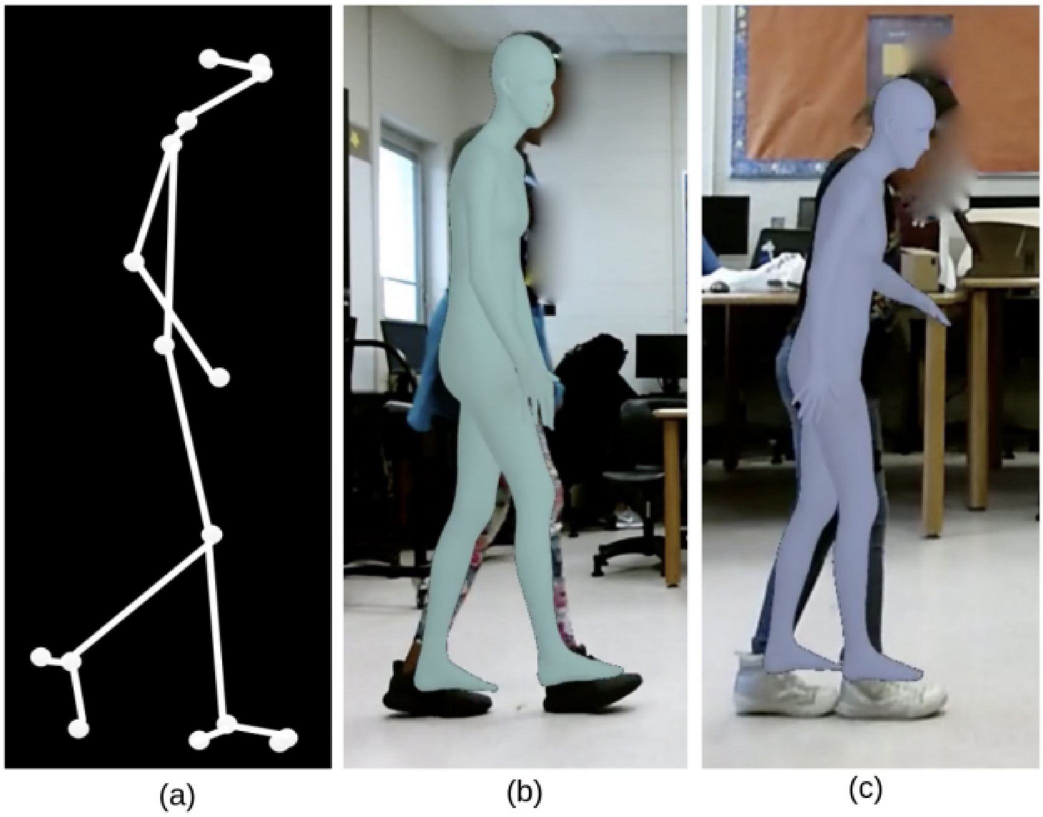


Figure 3.8: A participant performing the Tandem Gait Forward task: (a) Extracted body pose keypoints, (b) An invalid tandem gait step, and (c) A valid tandem gait step [129].

The subject's 3D pose is extracted using the VIBE [73] pose estimation algorithm that predicts the parameters of a skinned multi-person linear (SMPL) body model [85] for each frame of an input video. From the keypoints extracted using VIBE, 17 are selected, including head, hands, hip, feet, and toes. Finally, the extracted body poses from a single video are divided into eight equal segments (with overlap), each corresponding to one step taken by the participant. Examples of valid and invalid steps are shown in Figure 3.8. In the figures, the children's bodies are covered by their estimated SMPL body mesh to see the VIBE pose estimation in action and protect the subject's identity. Each session is manually scored and annotated by the assistants, which act as the ground truth for our algorithms.

Proposed Method and Results

The 3D body keypoints extracted from an input video is first divided into 8 equal segments with each containing a step performed. Each segment ($X \in R^{32x51}$) includes 32 samples with 51 features each. The features are (x, y, z) coordinates for each of the 17 body keypoints rasterized into one vector.

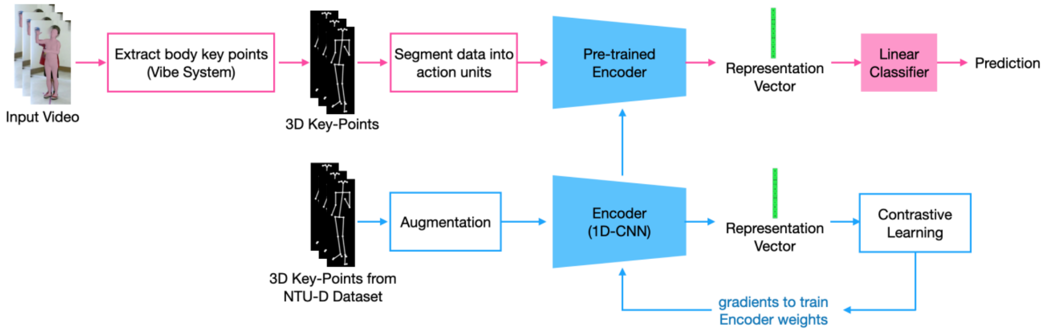


Figure 3.9: Proposed Classification Architecture. **Top (pink)**: Supervised classification and **Bottom (blue)**: Self-supervised pre-training [129].

The input X is fed into an encoder network as shown in Figure 3.9 to obtain a compact latent representation $z \in R^{256}$. It is then fed into a linear classifier to classify between valid and invalid segments. The encoder is comprised of a 4-layer 1D convolutional neural network (CNN) [77]. To evaluate the performance of the proposed method, the dataset is split into three scenarios. First, 80% of the total is used for training and the remaining 20% for testing. In the second scenario, 50% is used for training and 50% for testing. Finally, 10% is training and 90% for testing. An average classification accuracy is calculated using multi-fold cross validation and is presented in Table 3.4.

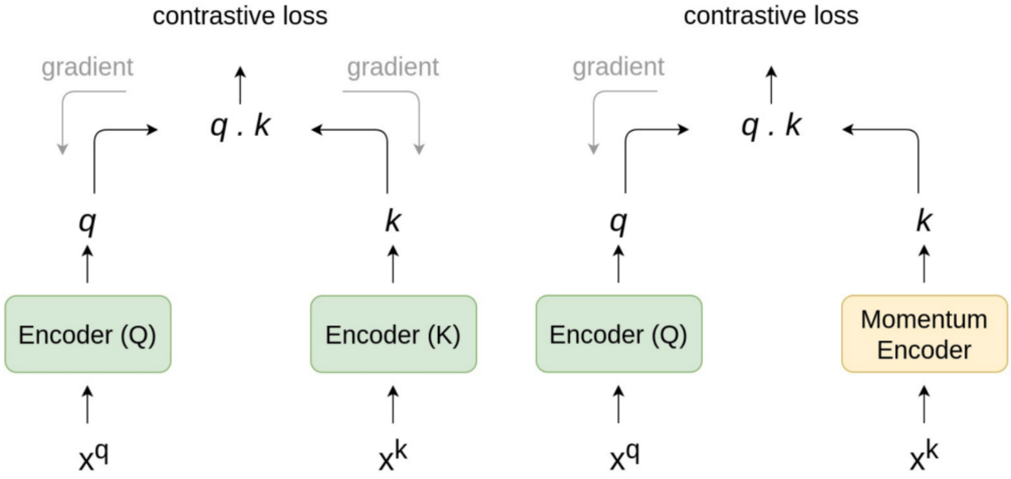


Figure 3.10: Contrastive Learning Algorithms. **Left**: End to End (E2E) training of encoders. **Right**: Using a momentum encoder as a dynamic dictionary lookup (MoCo) [54]

It is evident from Table 3.4 that the baseline supervised algorithm performs worse as the training set size decreases. The goal is to propose an algorithm that performs well despite the dataset's scarcity of labeled data. Therefore, a self-supervised learning technique called contrastive learning is used [62]. To improve the performance of the baseline model, the encoder is pre-trained on a large public dataset NTU-RGB+D 120 [118] using self-supervised learning [62]. Contrastive learning (CL) tries to group similar samples closer and diverse ones far from each other in its latent representation space.

Table 3.4: Top-1 Classification Accuracy when different train/test split sets are used. First, 80% trainset and 20% testset. Second, 50% trainset and 50% testset. Third, 10% trainset and 90% testset [129].

Method	**80% Trainset**	**50% Trainset**	**10% Trainset**
Supervised	72.39%	63.33 %	52.13%
Contrastive Learning (E2E)	76.61%	72.44%	70.90%
Contrastive Learning (using MoCo)	76.61%	74.03%	72.46%

One sample (query x^q) from the training set is taken during training. An augmented (transformed) version (or another view in the NTU dataset) of the sample is considered a positive sample (positive key x^{k_+}), and the rest of the samples in the training batch are considered negative samples (negative key x^{k_-}). The pre-training encourages the encoder network to differentiate the positive samples from the negative ones.

In our work, we use two contrastive learning strategies for pre-training: End-to-End (E2E) learning and Momentum Contrast (MoCo) [54] as shown in Figure 3.10 to improve the performance of our system even when there are fewer annotations available. In E2E learning, a large batch size is used, and all the samples except for the query are considered negative samples. However, large batch sizes inversely affect the optimization during training; one possible solution is to maintain a separate dictionary known as a memory bank that contains all the negative keys and gets updated every epoch. In case of MoCo, the momentum encoder (θ_k) shares the same parameters are the query encoder (θ_q) and gradually gets updated as follows: $\theta_k = m\theta_k + (1 - m)\theta_q, m \in [0, 1)$, where ($m$) is the momentum coefficient.

3.1.3 Stand on One Foot

The ”Stand on One Foot” or ”Balancing on One Foot” is another physical task for measuring gross motor, gait, and balance. The participant is expected to stand on one foot for 10 seconds in this task. Participants are scored based on their capability to sustain their position for the given period. In the first round, the subjects stand on their left foot, and in the second round, they stand on their right foot. An example of a participant performing the task with their left foot is shown in Figure 3.11. The data collection is similar to the ”Tandem Gait” task as explained in section 3.1.2 with the participant’s front and side view recorded. In this work [104], we only use the front view as it is sufficient for our algorithms.

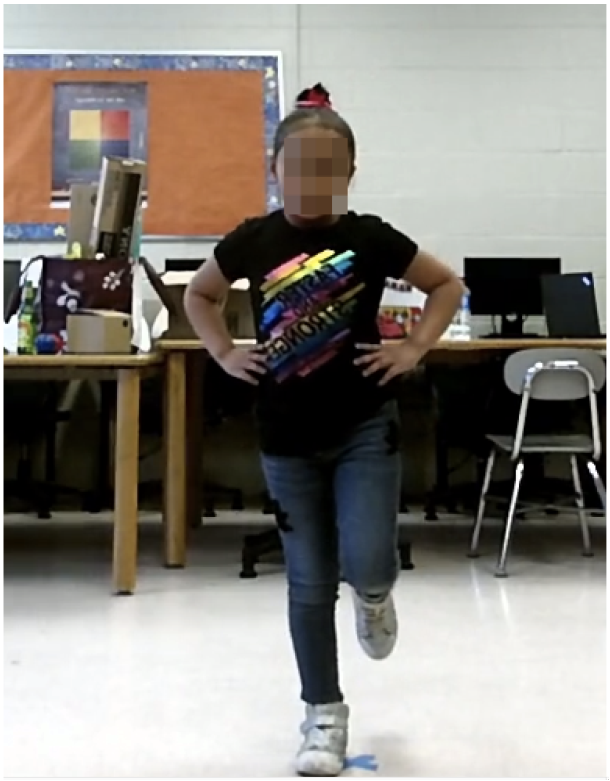


Figure 3.11: An example of a participant performing the second round of the ”Stand on One Foot” task standing on their right foot [104].

Proposed Method

Our proposed method works in two stages. First, it analyzes the total time the foot is balanced correctly out of the entire ten seconds. Next, we present an ergonomic score on a scale of 1 to 3 to indicate the participant’s posture, with 1 being poor, 2 being average, and 3 being perfect. Finally, we use body keypoints extracted using MoveNet [46] to analyze the pose of the subjects from the recorded videos. The idea is to build a lightweight system that can be deployed on smartphones and tablets without requiring specialized hardware and sensors. As shown in Figure 3.12, the algorithm first identifies whether the current pose is a good standing pose. Next, it scores the validity of the pose based on proper posture and ergonomics.

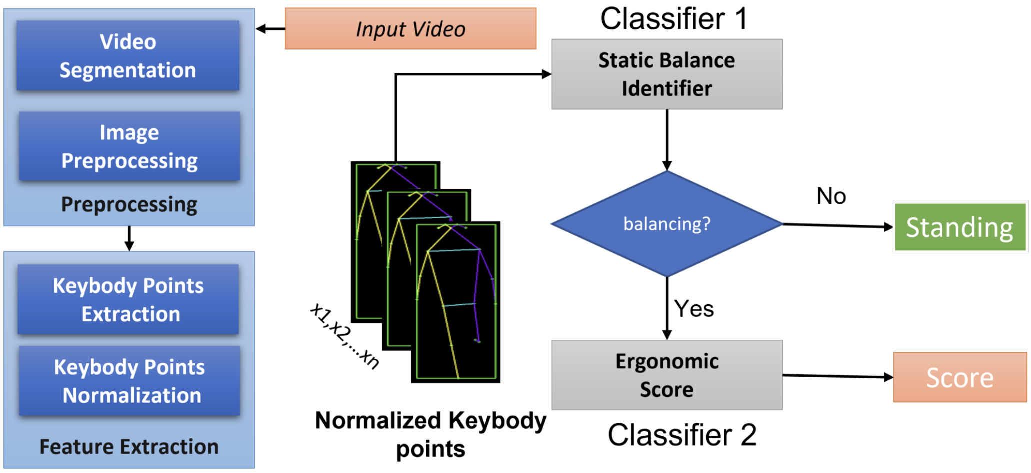


Figure 3.12: Complete system architecture to classify valid standing pose and score it based on their ergonomics (a) Static Balance Identifier: classifier a valid stand on one foot, (b) Scores the standing based on the pose [104].

Static Balance Identifier

First, a more straightforward approach is taken to identify whether the subject is ”Standing” on both feet or ”Balancing” on one foot. Next, the angle between the line joining the knee and the ankle and the perpendicular line to the floor is calculated as shown in Figure 3.13. Based on detailed observations by the experts, if the angles of both the legs are less than ten degrees with the perpendicular (P), the frame is classified as **Standing**.

- Classify the frame as **”Standing”** if the angle between the line segment joining the ankle and knee and P is between 0 and 10 degrees for both legs

$(\theta_1 \& \theta_2)$.

- Classify the frame as **"Balancing"** if the angle between the line segment joining the ankle and knee and P is more than 10 degrees for any one of the legs $(\theta_1 or \theta_2)$.

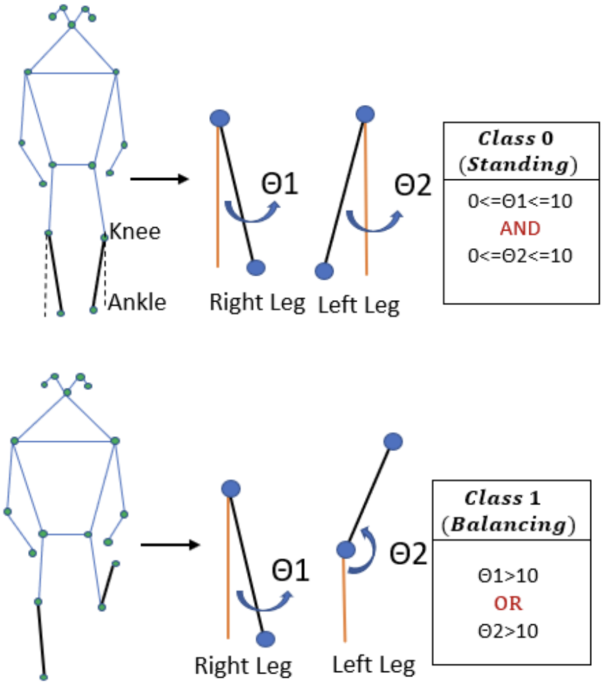


Figure 3.13: Angle range-based classification. **Top**: Standing on both feet (when the angles between the perpendicular and both the legs are less than 10 degrees, **Bottom**: Balancing on one foot (when one of the angles surpasses 10 degrees) [104].

Another approach to classifying between "Standing" and "Balancing" is to use a deep neural network using the body keypoints as the input features. For each image frame extracted from the videos, the feature matrix is of shape 17×3 (17 body keypoints with coordinates (x, y, c)), where (x, y) is the 2D coordinate, and c is the confidence score for the correctness of the keypoint position. The features are first normalized and flattened into a vector of shape 51×1 before training the neural network. The neural network is a 2-layer fully connected network with one output logit predicting one of the two classes.

Ergonomics Scoring

The extracted body keypoints from MoveNet [46] are used to evaluate the body posture and provide an ergonomics score. The first method used is called **Weighted Matching**, where the weighted score for each test image is calculated by comparing its extracted keypoints to the training images that act as the standard images for three different scores from 1 to 3. The weighted distance between the test image and three standard images representing scores 1, 2, and 3 is calculated as:

$$D(F, G) = \frac{1}{\sum_{k=1}^{17} F_{C_k}} \times F_{C_k} \| F_{xy_k} - G_{xy_k} \| \tag{3.7}$$

where F and G are the two L2-normalized pose vectors of the test and train images, F_{xy_k} and G_{xy_k} are the (x, y) coordinates of the k^{th} keypoint for each vector, F_{C_k} represents the confidence score of the k^{th} keypoint of F.

Once the weighted distance between the test images and all standard training images is calculated, we sort them in ascending order and pick the top-k images. Then, the test image is scored based on the highest frequency of the images with the same score.

An alternative method is to use the above-calculated angles (θ_1 & θ_2). For example, suppose only one leg is pulled above the ground, and its angle to the perpendicular P is between 0 to 20 degrees. In that case, we say that the subject is not balancing well enough and assign an ergonomic score of 1. Similarly, if the angle is between 20 and 40 degrees, we set an average ergonomic score of 2. Finally, if the angle exceeds 40 degrees, it is given a perfect score of 3.

Results and Discussion

During training, an 80-20 train-test split is done on the dataset where 20% of the set is used for testing and evaluation. The performance of the models is evaluated using a 5-fold cross-validation method. An SVM classifier is used as a baseline model to compare our proposed methods [104].

Table 3.5: Performance of our proposed methods [104] for Static Balancing classification and Ergonomic Scoring

Method	**Accuracy**	
	Static Balancing	**Ergonomics Scoring**
Weighted Ergonomic	-	**77.24%**
SVM	95.48%	80.1%
Angle (Range) Based	87.15%	61.5%
Neural Network	**97%**	**86.5%**

The experiments are performed in a system with an intel core i7-8750 quad-core CPU, 16GB of RAM, NVIDIA GTX 1060 GPU with 120 Cuda cores, and 14GB of graphics memory. The neural network is trained for 100 epochs with ADAM optimizer [72]. The empirical learning rate and batch sizes selected are 0.001 and 48, respectively.

For the Range-based classification (angles approach), a threshold of 10 degrees is selected to identify if the subject is **Balancing** or **Standing** as explained earlier. It is clear from Table 3.5 that the neural network outperforms other methods and is more suitable for the assessment task with a 97% classification accuracy.

Finally, the classification results and the calculated ergonomics scores are used to derive an ATEC score that helps measure the subjects' balance and attention. Since the videos are processed as 30 frames per second, the total time subject balances on one foot can be calculated as follows:

$$T_B = \frac{1}{30} \times F_B \tag{3.8}$$

where T_B is the total balance time by the subject and F_B is the number of frames classified by the algorithm as balanced in the video. The following ATEC scores are assigned based on the above equation: **0** if $T_B < 5$, **1** if $5 <= T_B <= 8$, and **2** if $9 <= T_B <= 10$. Psychology experts use these scores to diagnose symptoms of different cognitive functions addressed in the ATEC system.

3.2 Cognitive Fatigue in TBI Subjects with fMRI

Functional magnetic resonance imaging (fMRI) measures slight changes in blood flow that occur with activity in different brain regions. This imaging technique is completely safe and non-intrusive to the human brain. It is used to identify parts of the brain that handle critical functions and evaluate the effects of conditions such as stroke and other diseases. Some abnormalities can only be found with fMRI scans as they provide detailed access to human brain activity patterns.

Traumatic Brain Injury (TBI) is one of the most prevalent causes of neurological disorders in the US [40]. It is a condition that has been shown to affect working memory [28], and induce cognitive fatigue [74]. In this work, we focus on understanding cognitive fatigue that results from performing standardized cognitive tasks, as it is one of the primary indicators of moderate-to-severe TBI.

Cognitive Fatigue (CF) is a subjective lack of mental energy perceived by an individual which interferes with everyday activities [34]. It is a common condition among people suffering from moderate to severe brain injury. Many researchers have tried to use different approaches to assess CF through various cognitive tasks and assessment tests by using objective measures such as response time (RT) and error rate (ER) [34]. However, these measures have certain limitations and do not correlate well with the self-reported scores during the tasks [126]. The inability to relate objective measures to self-reported cognitive fatigue led us to study the blood-oxygen-level-dependent (BOLD) signal associated with neural activation changes. The increased BOLD activation in TBI subjects signifies excessive cognitive work compared to healthy subjects [126].

Raw fMRI scans are full of artifacts and noise due to several issues like central point artifacts, data clipping, data error artifacts, etc. These artifacts can differ based on the scanner used, and the settings applied during the scan. Thus, addressing and removing the unwanted noise is essential before analyzing the images. However, if a model can work directly on the raw data, it eliminates the painful process of pre-processing the images and saves time and effort. Hence, we prioritized training our models on raw data directly and compared the performance with models learned from pre-processed data. With the self-supervised approach, our model outperformed supervised methods trained on images without any artifacts or noise.

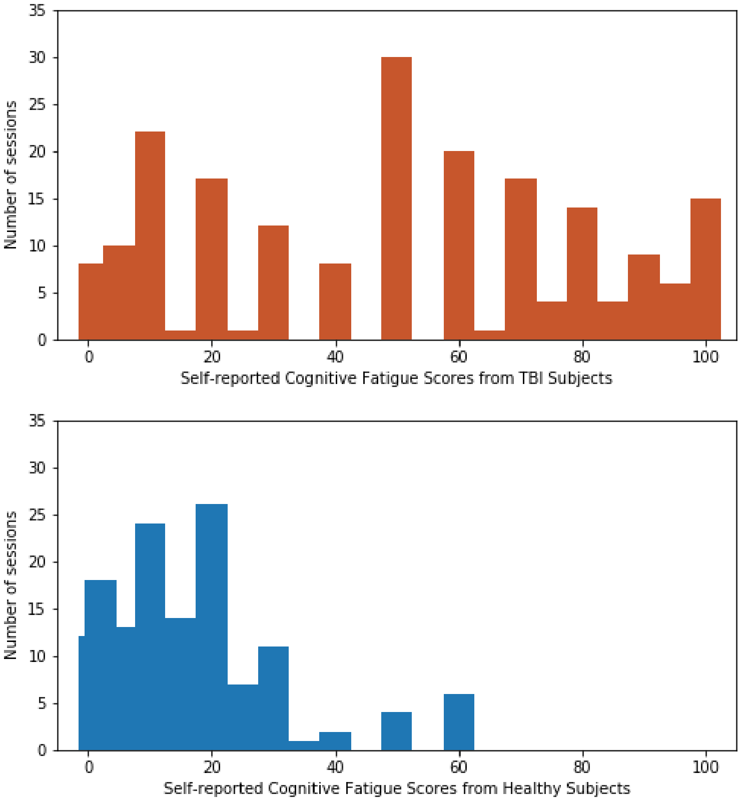


Figure 3.14: Distribution of self-reported cognitive fatigue scores after every N-back session from **TBI subjects** (top) and **Healthy Controls** (bottom). The score is a difference between the reported session score and the resting-state fatigue score recorded at the beginning of the first session.

Data Collection and Processing

For data collection, fMRI scans of the brain were recorded where each subject was asked to perform a series of cognitive N-back tasks, as shown in Figure 3.15. The data was collected from thirty participants with moderate-severe TBI and 24 healthy controls (HCs). The average age of the subjects was 41 years (SD=12.7). Each participant performed four rounds of both 0-back and 2-back tasks. A baseline fatigue score was reported initially, followed by scores being reported after each round. Functional images were collected in 32 contiguous slices during eight blocks (four at each of two difficulty levels), resulting in 140 acquisitions per block (echo time = 30 ms; repetition time = 2000 ms; field of view = 22 cm; flip angle = 80°; slice thickness=4 mm, matrix = 64 × 64, in-plane resolution = 3.438 mm^2). Using the Visual Analog Scale of Fatigue (VAS-F), the subjects were asked to rate the fatigue they experienced (in the range of 0-100) after each round of the N-back task. The self-reported scores were mapped to six classes to make it a multi-class classification problem, as represented in table 3.6.

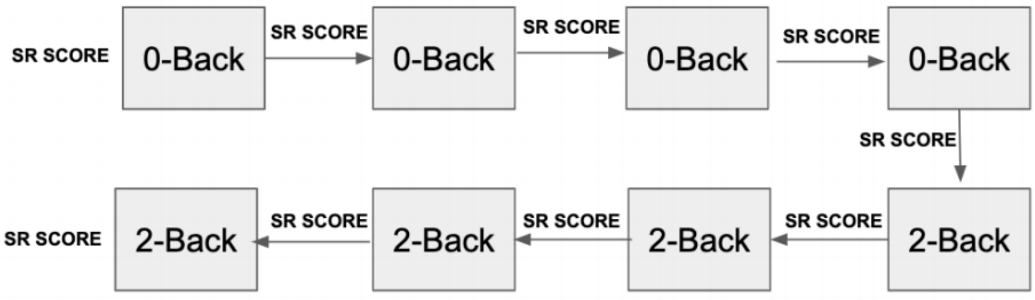


Figure 3.15: A flow diagram of a series of N-back tasks (some performed the 2-Back tasks first) performed during data collection (VAS-F Score: SR Score)

On empirical inspection of the distribution (in Figure 3.14), we find that six categories strike a good balance/compromise between adequately describing the distribution of VAS-F scores in a limited set of categories while also maintaining sufficient complexity in the VAS-F data to allow for accurate modeling. Reducing the number of categories to five or increasing it to seven does not materially affect the model's performance. Additionally, the cognitive fatigue levels shown in table 3.6 are for reference only in order to quantify different levels of fatigue corresponding to the class label. The final 4D tensor acquired in NIfTI format was 140 x 32 x 64 x 64. The raw fMRI images are preprocessed using Analysis of Functional NeuroImages (AFNI) [30] and other standard techniques as discussed in previous works [126], shown in Figure 3.15.

Table 3.6: Mapping self-reported (SR) Cognitive Fatigue scores to respective class labels. The fatigue levels are for reference only and are not of any clinical significance.

Fatigue Score (SR)	Fatigue Level (Reference)	Class
0-10	No Fatigue	0
10-20	Very Low Fatigue	1
20-40	Mild Fatigue	2
40-60	Fatigue	3
60-80	High Fatigue	4
80-100	Extreme Fatigue	5

Proposed Method

fMRI scans are 4D in shape and are represented as (t, x, y, z), where 't' represents the timesteps of individual 3D brain volumes. The other three dimensions represent the intensity of voxels in the brain. The temporal relation between the scans recorded at different time steps is captured using a Recurrent Neural Network (RNN) based architecture. We combine a CNN architecture with an LSTM [48] network for the encoder as shown in Figure 3.16. We use three layers of 2D convolution and batch normalization to learn the images' spatial (structural) features, whereas the LSTM network understands the temporal relation between the timesteps.

We use a specific self-supervised learning approach known as contrastive learning [62] to learn data representations from unlabelled samples. The encoder shown in Figure 3.16 is pre-trained on a public dataset called BOLD5000 [21] as well as our custom dataset using contrastive learning. Many researchers have opted for the BOLD5000 dataset as it is a large-scale, slow event-related human fMRI study incorporating 5,000 real-world images as stimuli. It also accounts for image diversity, overlapping with standard computer vision datasets,

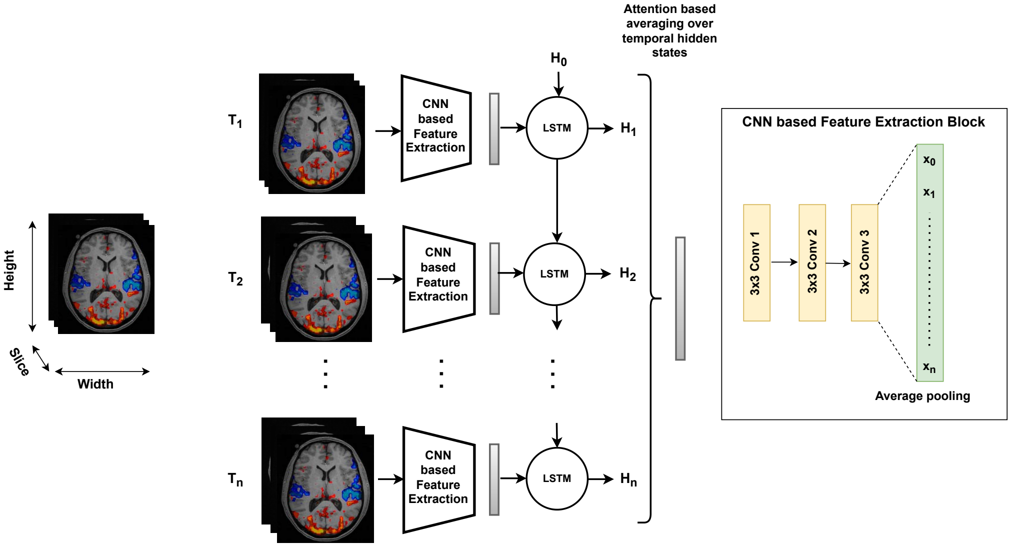

Attention based averaging over temporal hidden states
H0
T1
T2
Tn
CNN based Feature Extraction
LSTM
H1
H2
Hn
Height
Width
Slice
CNN based Feature Extraction Block
3x3 Conv 1
3x3 Conv 2
3x3 Conv 3
x0
x1
xn
Average pooling

Figure 3.16: Spatio-temporal Model Architecture: CNN layers in the Encoder extract spatial features while LSTM layers model the temporal relation of the fMRI images followed by attention-based averaging over time [63].

making it ideal for transfer learning tasks.

In contrastive learning, two augmented versions are generated for every image in a batch containing N samples, resulting in a total of $2N$. Every sample's augmented version is considered the positive candidate, and their similarity is encouraged to be maximum. In contrast, the model tries to minimize the positive and negative pair similarity. This condition is represented in Figure 3.17 with green and red double-headed arrows. We use cosine similarity to measure the closeness between two samples in a batch. In addition, we apply extensive spatial and temporal augmentation during training. As part of spatial transformation, methods such as random affine, z-normalization, and re-scale intensity were used. One arbitrary transformation is also used among random blur, gamma, random motion, and random noise. Similarly, a random starting time t is selected for temporal augmentation, and n consecutive scans are extracted. Finally, the loss is calculated using a variant of the Noise Contrastive Estima-

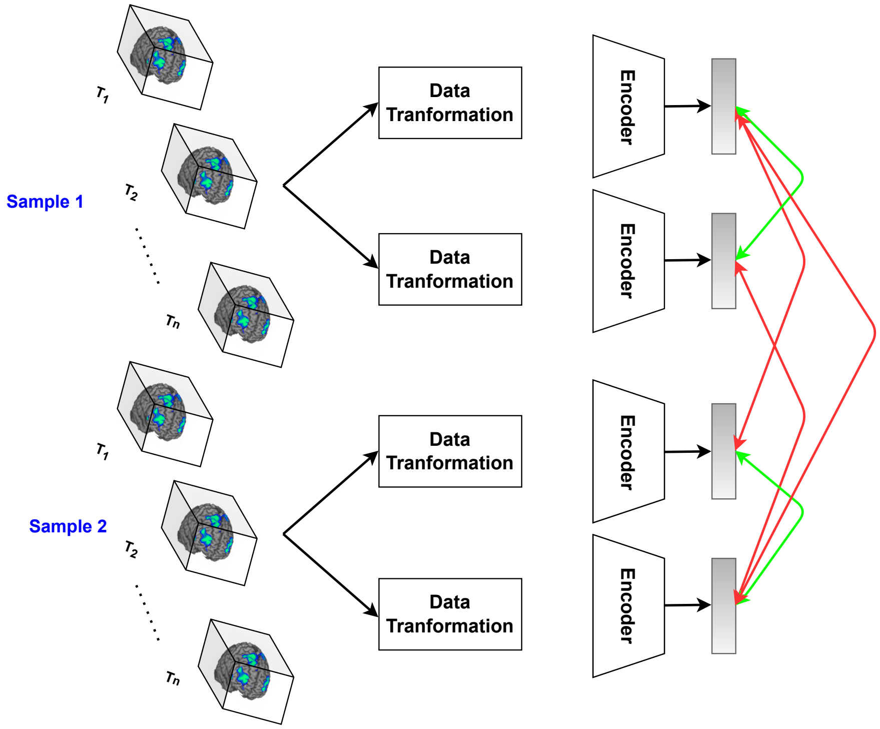


Figure 3.17: Self-supervised Pre-training Framework: MoCo algorithm for pre-training on BOLD5000 dataset. The green arrows represent positive pairs and red arrows represent negative pairs [63].

tion function (NCE) called InfoNCE, which is used when there is more than one negative sample present during the learning process and is defined by equation 3.9.

$$L_{infoNCE} = -log \frac{exp(sim(q,k_+)/\tau)}{exp(sim(q,k_+)/\tau)+\sum_{i=0}^{K} exp(sim(q,k_i)/\tau)} \tag{3.9}$$

where q represents the current sample, k_+ represents the positive sample (augmented version of q), and k_i represents the negative samples (other samples in the batch). τ represents the temperature coefficient, and sim represents the cosine similarity between two samples.

MoCo [54] is a contrastive learning technique that maintains a dictionary queue of negative samples that is updated gradually during the training phase.

Two encoders with the same architectural configuration are used; the main encoder Q (Query Encoder) is trained end-to-end on the sample pairs. The second encoder (Momentum Encoder) shares the same parameters as Q. The momentum encoder generates a dictionary as a queue of encoded keys with the current mini-batch enqueued and the oldest mini-batch dequeued. It gets updated based on the parameters of the query encoder using an update parameter called momentum coefficient as represented by equation 3.10. In equation 3.10, $m \in [0, 1)$ is the momentum coefficient. Only the parameters θ_q are updated by back-propagation.

$$\theta_k \leftarrow m\theta_k + (1 - m)\theta_q \tag{3.10}$$

Region of Interest Analysis and Cognitive Fatigue Interactions

Based on the studies in [22, 36, 126, 37, 27], the striatum of the basal-ganglia also known as caudate, the medial prefrontal cortex (mPFC), the anterior insula, and the middle frontal gyrus (MFG) has been found to play a critical role in functional connectivity of the fatigue network in the brain. Therefore, these brain areas need to be analyzed thoroughly to understand activation regions in the brain during cognitive fatigue.

Data analysis occurs in two steps. A whole-brain study is conducted first, followed by a fatigue-interaction (FI) analysis where cerebral activity in the brain is investigated in different regions of interest (ROIs):

- We train ML models for CF detection using the whole brain scan.

- We apply several masks one at a time to the brain scans corresponding to different selected ROIs before training the same ML models.
- We compare the performance of the ML models for each region of interest against the whole brain scan.

Results and Discussion

Most publicly available datasets are preprocessed with a standard pipeline for fMRI images. However, we used two different data versions to train the models: one using the **raw (unprocessed)** version and the other using **preprocessed** normalized version as obtained from the preprocessing pipeline in Figure 3.15. Furthermore, we used data from all four subjects in the publicly available **BOLD5000** dataset for self-supervised pre-training of the Encoder model as represented in Figure 3.17. In this case, we trained the encoder using MoCo [54] algorithm and Adam optimizer on the public dataset. The pre-training was carried out for a total of 200 epochs. The starting learning rate was set to 0.03 with a weight decay factor of 10^{-4} and a momentum parameter of 0.9. The learning rate was decayed by ten at 120 and 160 epochs, respectively.

Table 3.7: Performance results for different models on cognitive fatigue classification task. Accuracies are calculated with 3-fold cross-validation. The encoder model used is CNN+LSTM and is the same for all three approaches. For the supervised approach, we add a linear layer at the end for classification.

Approach	**Data Format**	**Dataset Used**	**Accuracy**		
			HC only	**TBI only**	**Overall**
Supervised (Encoder + Linear)	Raw	Ours	71.72 ± 0.82	78.44 ± 1.71	74.35 ± 1.27
Supervised (Encoder + Linear)	Pre-processed	Ours	80.87 ± 0.63	84.91 ± 1.44	82.79 ± 0.73
Self-supervised + Fine-tuning	Raw	BOLD5000 + Ours	82.58 ± 0.53	92.39 ± 1.26	**86.84** ± **1.13**

We split our supervised labeled dataset into train, validation, and test sets to train the deep learning models. The train set contained 70% of the dataset, while the validation and the test datasets consisted of 15% each. The test set included a mix of TBI and HC subjects and constituted more than 300 reported instances during the N-back tasks. On the other hand, scans from all four subjects in the BOLD5000 dataset were used to pre-train the model using the self-supervised approach, as mentioned earlier. The primary encoder was initially trained on our collected dataset separately using a supervised approach for benchmarking. We used raw and preprocessed data for the supervised method to train two different models, as shown in table 3.7. Finally, once the encoder was pre-trained on the BOLD5000 dataset using the self-supervised algorithm, it was fine-tuned on our dataset. The performance of the different models is presented in table 3.8. The results show that the model pre-trained on the public dataset (BOLD5000) and later fine-tuned on our dataset outperformed other supervised methods.

One of the main objectives of analyzing different brain regions is to understand and quantify the activity in those regions when a subject exerts effort in the brain and fatigue increases. The higher the activation in an area, the more significant its contribution towards the induction of CF. It can be made more evident by testing the ML models on each selected brain region separately. To achieve this, 3D binary masks of the same size as the original scan were generated that correspond to each brain region respectively. Next, they were applied to the input scans using multiplication to prepare them for training. In this way, each of the four areas mentioned above in the brain was used to train our ML models and evaluated based on its sole ability to detect CF.

Table 3.8 highlights the performance of different models on detecting cog-

Table 3.8: CF Detection using Different Regions of Interest (ROIs) to identify areas with most brain activity. **Supervised** model refers to (Encoder + Linear) combination of layers trained on labeled data from our dataset where as **SSL** refers to Self-supervised Model initially trained on BOLD5000 dataset and later finetuned on our dataset.

Mask Used	**Data Format**	**Model**	**Accuracy**
Caudate	Pre-processed	Supervised	69.21%
	Raw	Supervised	62.33%
		SSL	**76.87**%
Insula	Pre-processed	Supervised	64.77%
	Raw	Supervised	62.12%
		SSL	67.94%
MedialPFC	Pre-processed	Supervised	73.78%
	Raw	Supervised	70.94%
		SSL	**78.92**%
MFG	Pre-processed	Supervised	75.29%
	Raw	Supervised	70.98%
		SSL	**79.13**%
NONE	Pre-processed	Supervised	82.79%
	Raw	Supervised	74.35%
		SSL	**86.84**%

nitive fatigue when trained using scans from various regions in the brain. It is prominent that the models perform better when the whole brain scan is used than using only a part of the brain. However, it is interesting to note that some regions in the brain provide more information than others when detecting cognitive fatigue. In this case, the medial prefrontal cortex (mPFC) and the middle frontal gyrus (MFG) seem to contribute way higher than the insula and slightly more than the caudate. It indicates a higher functional activity in the brain's frontal portion during fatigue induction.

Since our dataset contains a mix of TBI and HC subjects, it is essential to understand the difference between the cerebral activity in the brain that induces cognitive fatigue in both groups. Therefore, we compare the performance of our ML models independently on data from each group. As shown in table 3.7, when testing TBI and HC subjects separately, the models seem to perform better on the TBI data. It could mean that the enhanced brain activations in the TBI subjects made it easier for the model to predict cognitive fatigue compared to the scans from healthy subjects. However, the difference in the performance is negligible, and the model seems to perform comparatively well on data from both subjects, which makes it robust for all cases. Also, based on the score distribution of TBI and HC subjects in Figure 3.14, TBI subjects seem to induce more fatigue than healthy subjects.

CHAPTER 4

ASSESSMENT OF COGNITIVE FATIGUE WITH MULTI-MODAL SENSORS

In this chapter, we look into the current work that is being carried out and the outline of the final dissertation.

4.1 Introduction

Fatigue is a state of weariness that develops over time and reduces an individual's energy, motivation, and concentration. Fatigue can be classified into three types: Acute fatigue is caused by excessive physical or mental exertion and is alleviated by rest. Changes in circadian rhythm and daily activities influence Normative fatigue. In contrast, Chronic fatigue is primarily caused by stress or tension in the body and is less likely to be relieved by rest alone. While various factors influence human fatigue in the real world, factors affecting sleep and the circadian system have a high potential to contribute to fatigue [66].

Severe or chronic fatigue is usually a symptom of a disease rather than the result of daily activities. Some conditions, such as Multiple Sclerosis (MS) [79], Traumatic Brain Injury (TBI) [12], and Parkinson's Disease (PD) [51], have fatigue as a significant symptom. Physical and cognitive fatigue are the two types of fatigue. Physical fatigue (PF) is most commonly caused by excessive physical activity and is usually associated with a muscle group or a general feeling of fatigue in the body [23]. Cognitive fatigue (CF), on the other hand, can occur as a result of intense mental activity, resulting in a loss of cognition with decreased attention and high-level information processing [94]. In the real world, however, there is no clear distinction between what causes both types of fatigue. Workers with heavy machinery, for example, may require both cognitive skills and physical labor to complete a task that may induce both PF and CF simultaneously.

Researchers have previously attempted to assess both types of fatigue separately by approaching them differently. One of the most common methods of studying fatigue is to analyze the participants' subjective experience by having them fill out surveys rating their current state of fatigue. Although these methods have successfully quantified human fatigue, they are frequently prone to human bias and poor data collection methods. For example, an aviation study [13] discovered that 70-80 percent of pilots misrepresented their fatigue level. As a result, relying solely on a subjective measure from the participants may raise safety concerns. It is where physiological sensors, which provide objective measures of fatigue, come into play. To study fatigue, data collected from sensors such as electrocardiograms (ECG) [58], electroencephalograms (EEG)[65], electrodermal activity/galvanic skin response (EDA/GSR) [32], and electromyograms (EMG) [29] have been commonly used.

4.2 Multi-modal Physiological Sensory System

In a multi-modal sensory system, it is essential to figure out what variety and combination of sensors contribute the most towards assessing cognitive fatigue. A review of wearable and vision sensors [2] uncovers that motion (MOT), electroencephalogram (EEG), photoplethysmogram (PPG), electrocardiogram

(ECG), galvanic skin response (GSR), electromyogram (EMG), skin temperature (Tsk), eye movement (EYE), and respiratory (RES) sensors are the most effective for monitoring fatigue.

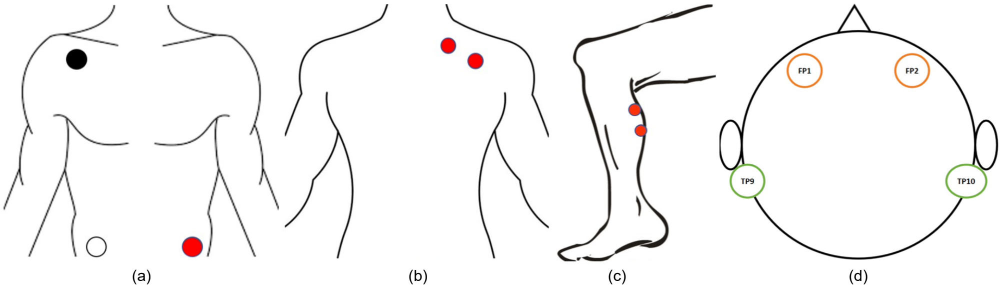


Figure 4.1: Sensor placements on the human body (a) **ECG**: right shoulder, the left and the right hip forming Einthoven's triangle [38] (b) **EDA/GSR** electrodes on the left shoulder to record the skin conductivity, (c) **EMG** electrodes recording muscle twitches from the right calf, (d) **EEG** sensor positions in the 10-10 electrode system used by MUSE. It records data from the TP9, AF7, AF8, and TP10 positions in the system.

In our preliminary work, we use a combination of four wearable sensors (ECG, EDA, EMG, and EEG) as shown in Figure 4.1. A MUSE S headset (section 2.3.1, Figure 2.3) is used to monitor the EEG signals of the participants. On the other hand, Biosignal Plux kit [14] is integrated into a wearable shirt (Figure 4.2) to simultaneously collect ECG, EDA, and EMG signals from the participants. Fatigue can harm the cardiovascular system, endocrine system, and brain. Therefore, these multi-modal signals help keep track of the subject's physical state and can provide quality information on whether the person is feeling fatigued.

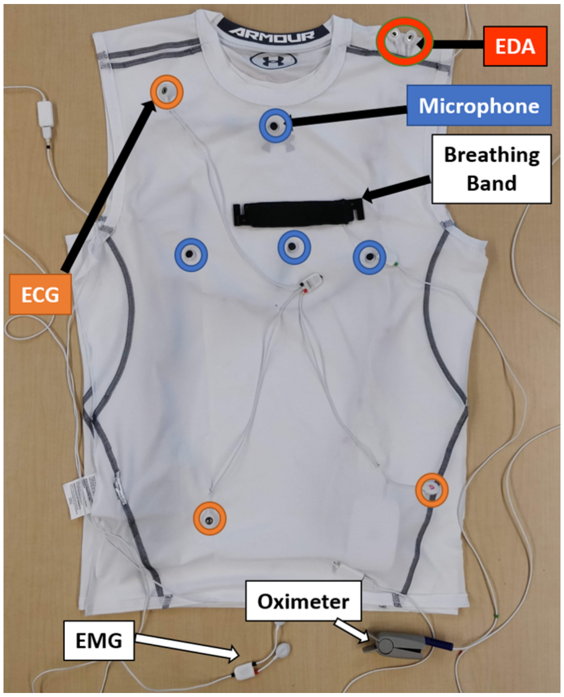


Figure 4.2: A prototype of the sensor shirt (inside-out view): Physiological sensors are embedded in the shirt and remain in contact with the subject's torso during data collection using adhesive tapes. Multiple sensory signals are collected using different combinations of the sensors present. **In this work, we do not use the microphone, oximeter, and breathing band signals.**

4.3 Experimental Setup

We build an experimental setup (Figure. 4.3) around our custom-built wearable t-shirt (presented in Figure. 4.2) to record physiological data using the attached sensors and a MUSE S headband (as shown in Figure. 2.3). In Figure. 4.2, although additional sensors are connected to the shirt, such as Microphone, Breathing Band, and Oximeter, we ignore these signals for fatigue detection as they are not significant for cognitive assessment. The suit uses a stretchable Under Armour shirt [5]. The advantage of using a shirt with embedded sensors is its ease of use during data collection and practical applications during day-to-day use. In addition, the sensors are hidden behind a detachable covering, so they can easily be removed while washing the shirt. In two separate study sessions, we collected data from 32 healthy people (18-33 years old, average age 24 years, 28/72% female/male). In addition, brain EEG signals are recorded using a MUSE S headband sensor. Participants are asked to wear the t-shirt and the MUSE headband throughout the experiment.

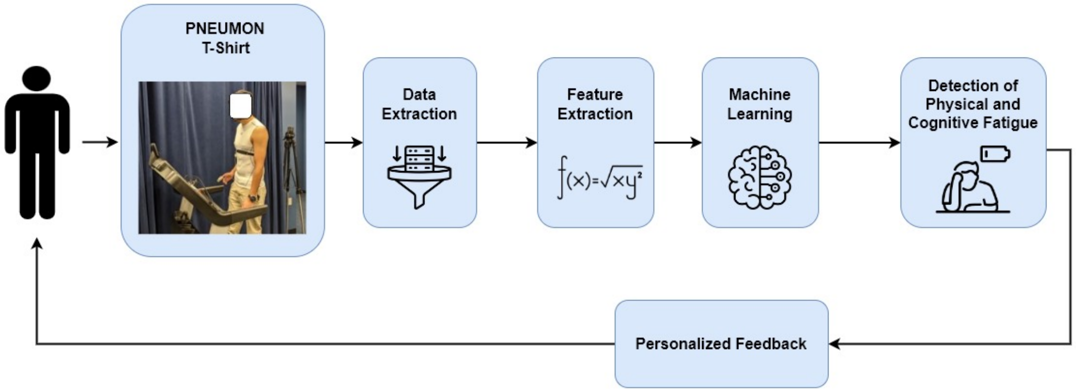


Figure 4.3: System Flow Diagram for CF Detection: Data collection using the sensors attached to the t-shirt (named PNEUMON) and MUSE S worn by a participant while performing the tasks presented in Figure. 4.4. Features extracted from the recorded signals are used to train ML models to detect the state of CF [64].

The depicted architecture in Figure 4.3 represents a real-time machine learning system designed for the detection of both physical and cognitive fatigue. The data extraction component retrieves sensor data from the PNEUMON T-shirt, encompassing parameters like heart rate, respiratory rate, and body temperature. Subsequently, relevant features for fatigue detection are extracted from this sensor data. For instance, it might include metrics such as the average heart rate over the past minute or heart rate variability.

A machine learning algorithm is then deployed to predict fatigue based on the extracted features, trained on data from individuals known to be either fatigued or not fatigued. Following the prediction, personalized feedback is

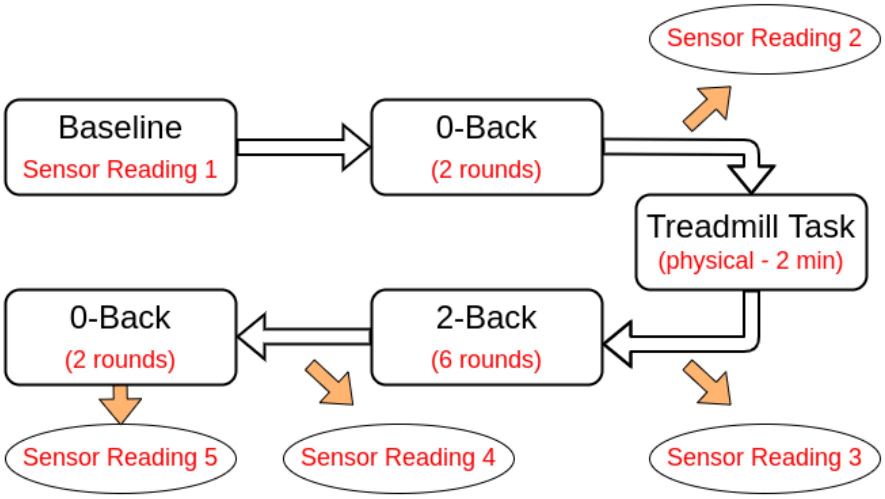


Figure 4.4: Flow diagram of the tasks performed by a participant.

provided to the user in the form of a fatigue score, enabling them to monitor changes in their fatigue levels over time. This versatile system finds applicability in various settings such as workplaces, schools, and sports facilities. Its utility lies in aiding individuals to recognize and manage their fatigue levels, consequently enhancing safety, performance, and overall well-being.

4.4 Data Collection

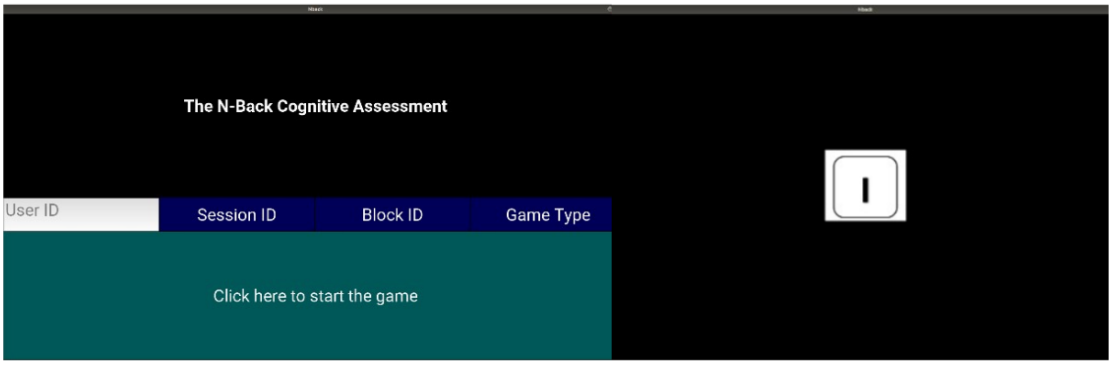


Figure 4.5: Graphical User Interface built for the N-Back tasks with an example image of a letter during a game round on the right.

As shown in Figure 4.4, we first collect baseline readings from the sensors while the subject stands still for a minute. Then, since the experiment is focused on inducing cognitive fatigue (CF), the participants are asked to perform multiple sets of N-back tasks (section 2.1.1). Next, a GUI for the N-back game is developed, as shown in Figure 4.5. In the N-back tasks, the subject is shown a series of letters, one after the other. The goal is to determine whether the current letter matches the letter presented N steps back. If it does, the subject must perform the specified action (pressing the space bar on the keyboard). Furthermore, they are also asked to run for 2 minutes on a treadmill (speed - 5mph, incline - 10%) to induce physical fatigue (PF). Again, it is done to study the effects of physical activity on CF. The signals from the sensors are recorded after each block of activity presented in Figure 4.4.

The study is divided into two sessions on separate days for each participant. Each subject is asked to come in the morning for one session and in the evening for another. It eliminates the effect on the data caused by the time of the day. The tasks performed in both sessions are identical, the only difference being the order in which they are done. The first session follows the flow depicted in Figure. 4.4, whereas the cognitively challenging 2-Back game is performed before the physical task in the second session. It dismisses PF's reliance on CF and allows us to collect more robust data for analysis. Each round of 0-back and 2-back tasks lasted between 80.4 seconds and 144.3 seconds, respectively (on average) during the data collection process.

Participants are asked to complete a brief survey indicating their current physical and cognitive fatigue levels following each task. In addition, they reported visual analog scale (VAS) scores ranging from 1 to 10 for the following questions:

1. Describe your overall tiredness on a scale of 1-10.
2. How physically fatigued/tired do you feel on a scale of 1-10?
3. How cognitively fatigued do you feel on a scale of 1-10?
4. How sleepy or drowsy do you feel on a scale of 1-10?

4.5 Data Processing

4.5.1 EEG

We employed the MUSE S headset to capture EEG signals while subjects engaged in various tasks throughout the experiment. The headset features four electrodes (AF7, AF8, TP9, TP10) placed on different regions of the head, as illustrated in Figure. 4.1(d). EEG signals serve as a measure of electrical activity in the brain, commonly decomposed into five frequency bands: alpha, beta, delta, gamma, and theta, as depicted in Figure. 4.6. Each band corresponds to a distinct brain state; for instance, delta waves (0.5 Hz to 4 Hz) emerge during deep sleep, while beta waves (13 Hz to 30 Hz) are indicative of active thinking. Similarly, alpha waves (8-12 Hz) are associated with normal awake conditions, gamma waves (30-80 Hz) with sensory perception integration, and theta waves (4-7 Hz) with drowsiness and early stages of sleep. To prevent power line interference on the signals, frequencies in the range of 50-60 Hz were preprocessed.

The MUSE SDK facilitated the streaming of various EEG bands and raw signals through a UDP server, with data recorded from all four electrodes into a

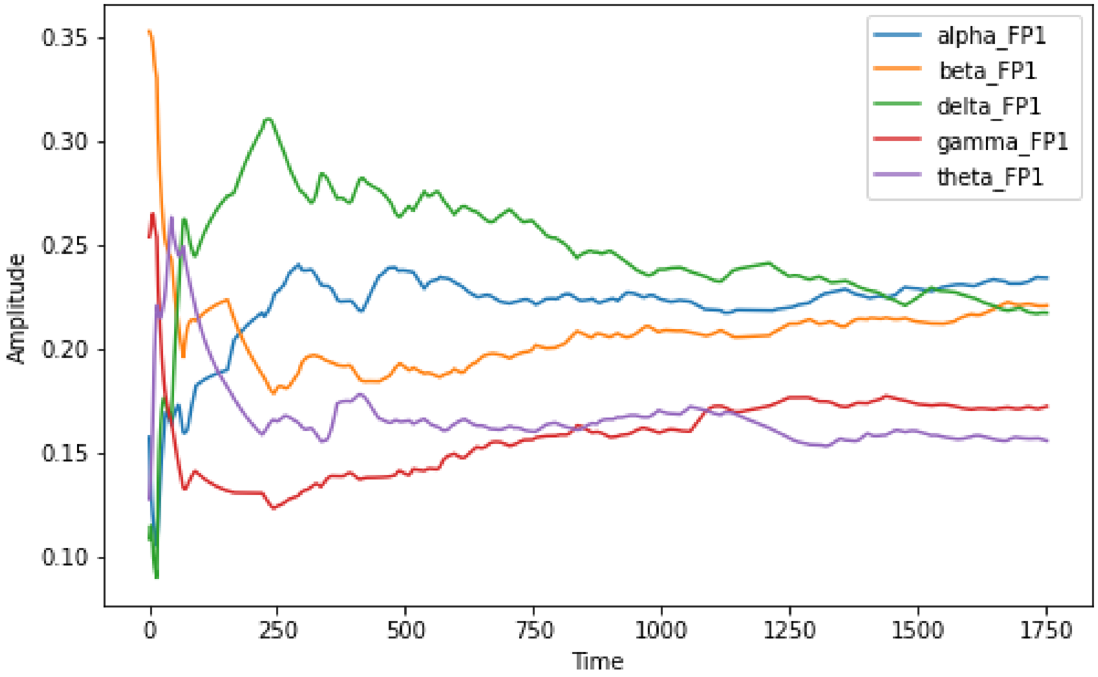


Figure 4.6: Raw Amplitude plot of Frequency bands extracted from the electrode at AF7 position (on MUSE) from a sample raw EEG signal from one of the subjects. The readings were collected during one of the 2-Back tasks undergoing for a little under 3 minutes.

CSV file. For feature extraction, we applied the sliding window technique to calculate statistics such as mean, standard deviation, minimum, maximum, and median for each band from the corresponding electrodes.

4.5.2 ECG, EDA, and EMG

In the wearable T-shirt, illustrated in Figure. 4.2, physiological sensors were integrated to concurrently capture ECG, EDA, and EMG signals throughout the experiment. Given that fatigue can impact the cardiovascular system, endocrine system, and brain, these multi-modal signals serve to monitor the subject’s physical state and offer valuable insights into their fatigue levels.

ECG signals reflect changes in the cardiovascular system by capturing the heart's electrical activity. Characterized by fiducial points labeled P, Q, R, S, and T, ECG provides essential information about cardiac pathologies [114]. Studies indicate that fatigue influences the cardiovascular response [98]. Einthoven's triangle approach, utilizing three limb leads in an imaginary triangle, was employed for ECG signal recording [38], aiding in the identification of incorrect sensor lead placement.

In contrast, EDA (also known as galvanic skin response or GSR) gauges sympathetic nervous system activity, dependent on physiological and emotional activation, by measuring skin conductivity. EDA signals offer insights into emotional states, allowing the identification of psychological or emotional arousal episodes. EMG signals, capturing voltage changes during muscle contraction and relaxation, reveal muscle activity alterations influenced by fatigue [116].

To mitigate unwanted noise in ECG signals, the Pan and Tompkins QRS detection algorithm was employed [100]. Signal cleaning involved a high-pass Butterworth filter with a fixed cutoff frequency of 0.5 Hz, followed by a notch filter to eliminate components at 50 Hz and prevent power line interference. RR intervals were extracted from the signals, outliers were removed, and missing values were interpolated linearly. Subsequently, 113 time-domain and frequency-domain features, including heart rate variability (HRV) metrics, were extracted for training machine learning models.

EDA signals underwent a low-pass Butterworth filter with a 3 Hz cutoff frequency. The phasic component, representing faster-changing elements due to stimuli, was extracted for analysis. Skin Conductance Response (SCR) peaks were identified as features from the cleaned EDA signals. Time-domain and frequency-domain features were similarly extracted from EMG signals. The Neurokit2 package [90] facilitated most of the feature extraction for all three signals.

4.6 Data Analysis and Results

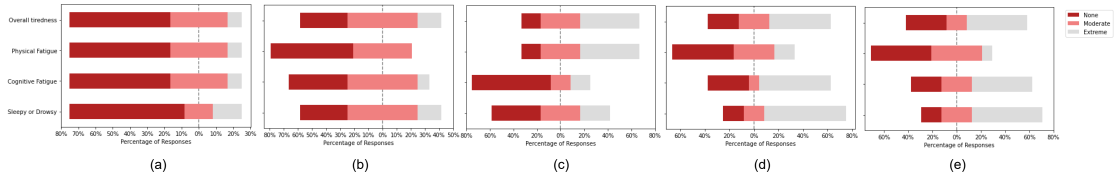


Figure 4.7: Likert plots for survey responses from subjects after each block of tasks (before taking the sensor readings) as shown in Figure. 4.4. (a) Initial survey at the start of the session showing that less to no fatigue was found in most of the participants. (b) Response after the first 0-Back task and before sensor reading 2 showed a slight increase in CF for some participants. (c) Response after the treadmill task verifying that PF is induced in most of the participants. (d) Response after the first 2-Back task indicating a prominent increase in CF (e) Shows that CF continues to be persistent even during the easier 0-Back task [64].

Figure. 4.7 visualizes the distribution of self-reported VAS scores after each task. We divide the scores from 1 to 10 into three categories for simplified visualization: None (<4), Moderate (≥ 4 and =7), and Extreme (>7). We found that most participants gradually began to feel cognitively fatigued as they kept performing each task block. In fact, after the fourth block, CF appears to be induced in more than 80% of the subjects (Figure. 4.7 (d)), supporting the hypothesis on which the experimental setup is based. Similarly, as shown in Figure. 4.7(c), the physical task manages to induce at least moderate PF in more than 90% of the subjects (c). Thus, data collected during the fourth and fifth blocks of the study are considered a state for CF in the participants, whereas data collected immediately following the physical task is regarded as a PF condition. We expect to have some biases in the subjective scores collected. However, the participants' overall performance indicates the experiment stands a success.

We derived 100 statistical features from EEG signals and a combined set of 169 features from ECG, EDA, and EMG to train the machine learning models. Responding to participant feedback (as depicted in Figure. 4.7), data obtained from sensor readings 1, 2, and 3 (prior to the 2-Back tasks illustrated in Figure. 4.4) were categorized as the ″No CF″ condition. Conversely, data recorded during the final two readings (4 and 5, i.e., post the 2-Back rounds) were labeled as ″CF″ conditions. Similarly, data acquired immediately after the physical task (sensor reading 3) was designated as a ″PF″ condition when subjects were stationary. Notably, readings 4 and 5 were excluded from PF analysis.

For training purposes, rather than processing the entire signal block for a task as a singular input, we segmented the time signal into multiple slices based on different window sizes (5 seconds, 10 seconds, and 20 seconds). Each signal slice retained the original signal's label, and features were extracted, augmenting the volume of input data for ML model training. Nonetheless, the models were also evaluated using entire signal blocks as inputs. Likewise, during inference, the input signal was subdivided into smaller slices, adhering to the window size chosen during training. Each slice was individually classified by the model, and the entire signal block was ultimately classified based on the predominant class among the classified slices. This technique enhances the model's resilience to noise or outliers in the signals, as the impact of noise in certain slices may not significantly influence the final classification outcome.

The complete dataset was randomly partitioned into training (70%, 22 subjects), validation (15%, five subjects), and test (15%, five subjects) sets. Stratified

Table 4.1: Detection of Cognitive Fatigue (CF) with EEG Features only

Model	Accuracy (Window Size)				Avg. Recall
	5s	10s	20s	Full Block	
Log Reg.	69.7%	72.6%	71.3%	62.3%	0.76
SVM	73.1%	73.3%	71.7%	69.4%	0.81
RF	72.3%	**81.9%**	79.1%	76.3%	**0.89**
LSTM	69.8%	71.8%	73.8%	**81.9%**	0.82

Table 4.2: Detection of Cognitive Fatigue (CF) with ECG + EDA + EMG Features

Model	Accuracy (Window Size)				Avg. Recall
	5s	10s	20s	Full Block	
Log Reg.	69.8%	70.1%	67.2%	65.3%	0.69
SVM	71.2%	71.7%	70.8%	70.1%	**0.73**
RF	74.8%	**76.3%**	72.1%	70.9%	0.71
LSTM	62.2%	63.7%	68.9%	70.1%	0.69

Table 4.3: Detection of Physical Fatigue (PF) with ECG + EDA + EMG Features

Model	Accuracy (Window Size)				Avg. Recall
	5s	10s	20s	Full Block	
Log Reg.	72.2%	72.2%	68.2%	62.9%	0.74
SVM	76.1%	79.6%	75.2%	73.1%	0.86
RF	79.9%	**80.5%**	77.6%	77.2%	**0.88**
LSTM	64.2%	64.8%	62.7%	68.9%	0.79

Table 4.4: Detection of Cognitive Fatigue (CF) with EEG + ECG + EDA + EMG Features

Model	Accuracy (Window Size)				Avg. Recall
	5s	10s	20s	Full Block	
Log Reg.	64.0%	66.9%	66.1%	60.4%	0.69
SVM	70.3%	74.6%	74.5%	70.3%	0.79
RF	67.9%	77.2%	76.8%	74.5%	0.81
LSTM	71.3%	74.2%	74.8%	**84.1%**	**0.90**

sampling was employed during the dataset split to address any potential imbalance. Additionally, a 5-fold cross-validation was conducted for each of the four machine learning models: Logistic Regression (Log Reg.), Support Vector Machines (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) recurrent neural network. Various combinations of features extracted from the signals were utilized for physical and cognitive fatigue prediction.

Given that EEG signals capture brain activity linked to cognitive functions, the initial focus involved using features exclusively extracted from EEG data for predicting cognitive fatigue (CF), as outlined in Table 4.1. Similarly, models predicting both cognitive and physical fatigue (PF) were trained using features derived from physiological sensors (ECG, EDA, and EMG), as detailed in Tables 4.2 and 4.3. Ultimately, features from all data modalities were combined and normalized for CF detection, as depicted in Table 4.4. Principal Component Analysis (PCA) [115] was applied to reduce dimensions to 189 features for optimal ML model performance in the results presented in Table 4.4.

The initial three models (Log Reg., SVM, and RF) were trained using features extracted from the signals. In contrast, the LSTM model (with 256 hidden layers) was trained directly on raw signals, as it is proficient in processing time-series data. The LSTM models were trained using a window-based approach, with the input size for EEG signals set at t x 20 x 1 (five frequency bands from each of the four electrodes). Conversely, ECG, EDA, and EMG signals were combined to form t x 3 x 1 inputs. Finally, the LSTM was trained on t x 23 x 1 inputs for all combined signals, where ”t” represents the number of timesteps in the signal, varying based on the chosen window size.

The average recall (Avg. Recall) featured in all four tables signifies the average recall for the ”Fatigue” condition (either cognitive or physical) attained through 5-fold cross-validation for each model. The optimal value achieved for each model across various window sizes was selected. Notably, in Table 4.1, RF demonstrates superior performance in CF prediction, with an accuracy of 81.9%

Table 4.5: Comparison of different models with the state-of-the-art algorithms

Fatigue Type	Model	Accuracy	Avg. Recall	Ref.
Physical	RF	71.85	0.72	[86]
Physical	cCNN + RF	71.40	0.73	[86]
Physical	**RF (Ours)**	**80.50**	**0.88**	Table 4.3
Cognitive	RF	64.69	0.65	[86]
Cognitive	RF	66.20	0.66	[86]
Cognitive	**LSTM (Ours)**	**84.1**	**0.90**	Table 4.4

and correctly identifying CF conditions 89% of the time. The emphasis on recall is crucial, given our primary objective of accurately detecting actual fatigue conditions in subjects and minimizing false negatives.

Furthermore, RF exhibits a notable edge when trained with features from ECG, EDA, and EMG to identify both CF and PF, achieving respective accuracies of 76.3% (recall=0.71) and 80.5% (recall=0.88). Notably, the 10s window size proves optimal for feature-based models, while the LSTM model excels when provided with the entire signal block, as evident in Table 4.4. The amalgamation of all four modalities processed directly with an LSTM network eliminates the need for feature engineering and yields the best results for CF detection.

In a comparative context, Luo et al.'s study [86] is the only one addressing fatigue detection in human subjects using wearable sensors. Their pilot study explores various digital data sources related to physical activity, vital signs, and other physiological parameters, examining their correlation with subject-reported non-pathological physical and mental fatigue in real-world scenarios. A performance evaluation against their methodology, as presented in Table 4.5, reveals that our approaches outperform theirs in detecting both physical and cognitive fatigue, with RandomForest (RF) and LSTM models, respectively.

4.7 EEG-based Cognitive Fatigue Detection

In one of our studies [70], we delved into EEG data to identify cognitive fatigue in participants. The investigation focused on the correlation between self-reported, non-pathological cognitive fatigue and physiological multi-variate time-series data obtained from a MUSE EEG sensor device worn by healthy subjects. Various machine learning and deep learning approaches, such as Support Vector Machines (SVM) and Convolutional Neural Networks (CNNs), were employed for time series classification.

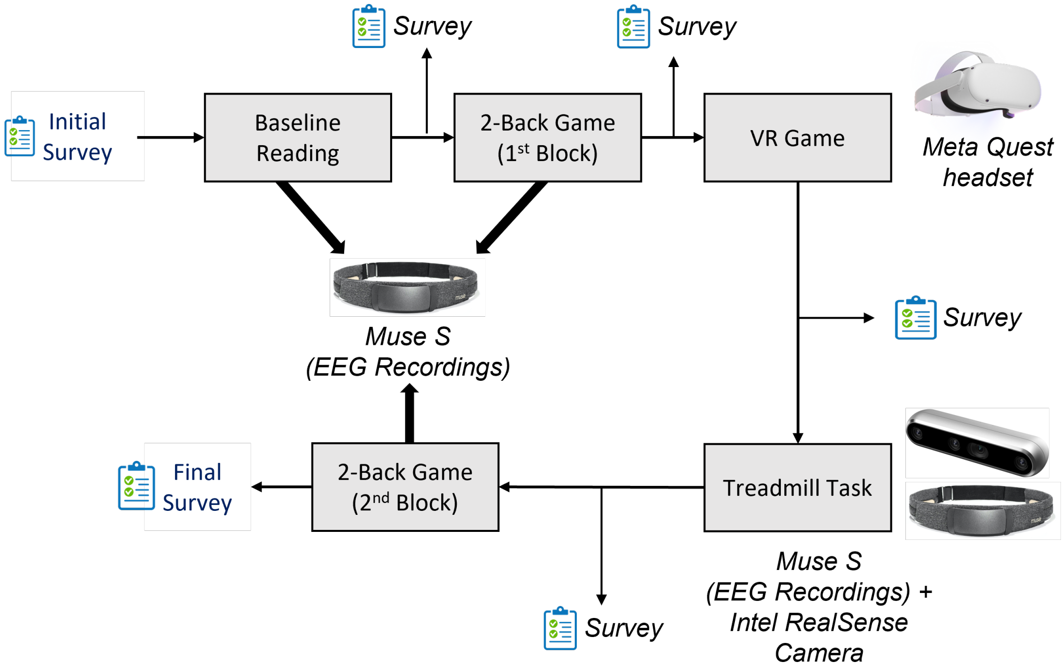


Figure 4.8: Experimental Setup that includes a VR game to induce cognitive fatigue in the participants along with the traditional N-back tasks.

As discussed earlier (see Section 4.5.1), raw timeseries EEG data was initially transformed into individual spectral components using Fast-Fourier Transform (FFT). The primary goal of this research was to ascertain the presence of cognitive fatigue solely through EEG readings.

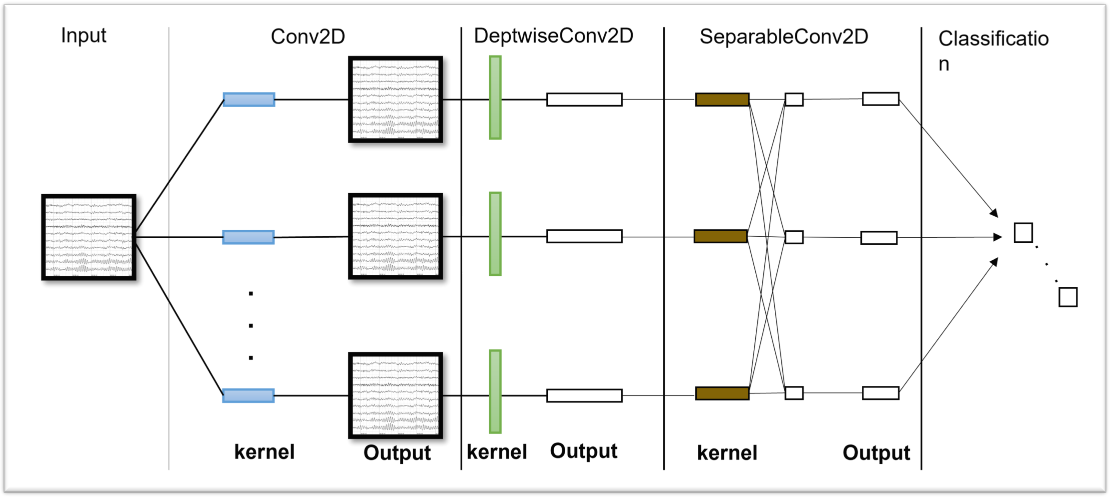


Figure 4.9: EEGNet architecture [81]

Two classification methods: **SVM** and **EEGNet** [81] were utilized to predict the classification labels. SVM models were trained using features extracted from the spectral components of the EEG data, achieving a maximum accuracy of 68%. Conversely, as depicted in Figure 4.9, causal EEG features were employed to train the convolutional network (EEGNet). The rationale behind selecting EEGNet was its ability to deliver high performance with limited training data and a modest number of model parameters. The Scikit-learn package in Python [106] was employed for model training on hardware consisting of an Intel Core i7-8750 quad-core CPU with 16 GB of RAM and an NVIDIA GTX 1060 GPU.

The dataset was partitioned into 60% for training, 20% for validation, and 20% for testing. For EEGNet, a filter length of 128 and a dropout rate of 0.2 were chosen. The kernel length was set to half of the sampling rate (256 Hz). Training utilized the Adam optimizer with a learning rate of 0.00001, a batch size of 16 for 500 epochs, resulting in the optimal performance with an accuracy of 91.40%. Precision, recall, and specificity values were reported at 0.91, 0.92, and 0.91, respectively.

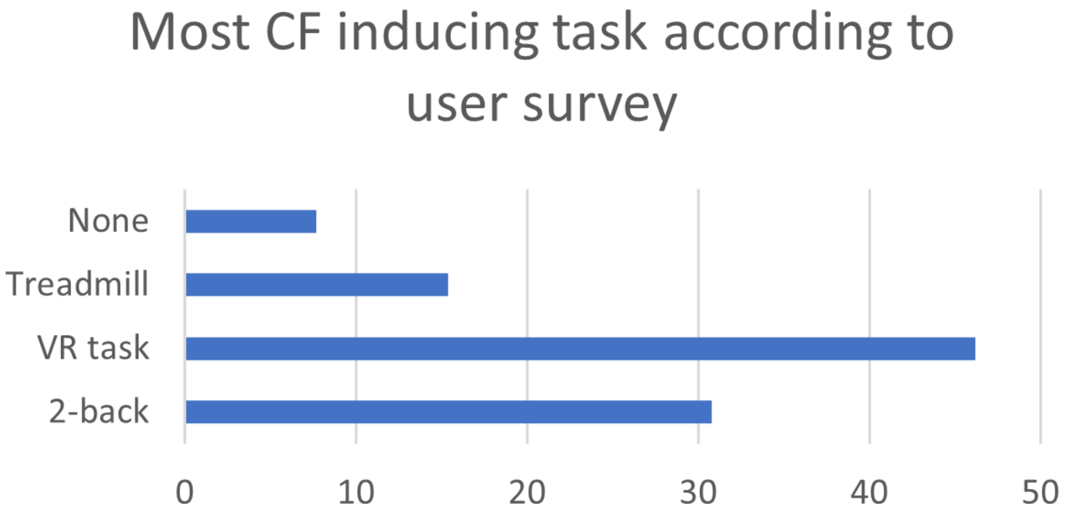


Figure 4.10: Percentage of CF inducing tasks according to user survey

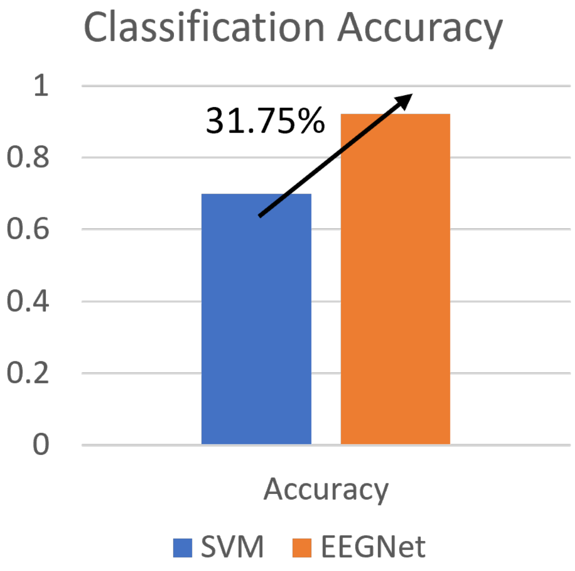
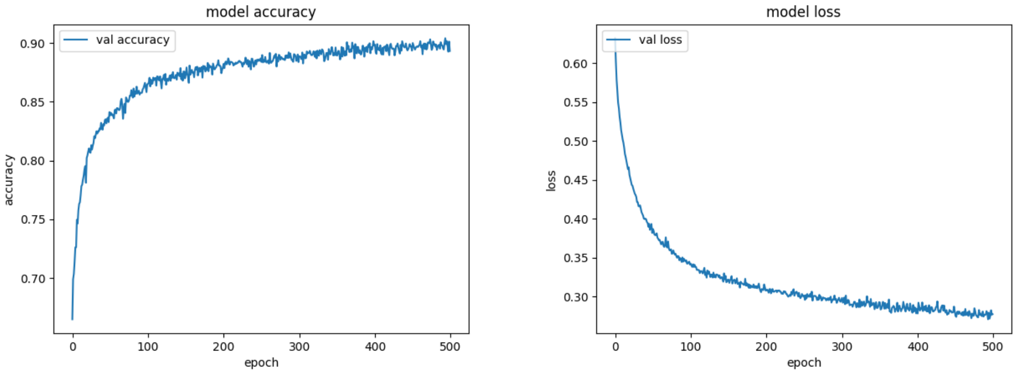


Figure 4.11: Classification Accuracy comparison between SVM and EEGNet models

The data presented in Figure 4.10 unmistakably indicates that participants experienced greater cognitive fatigue during two specific tasks: the VR game and the N-back task. The incorporation of the VR game aimed not only to enhance the experimental enjoyment but also to impose a higher cognitive load



(a) Validation Accuracy

(b) Validation Loss

Figure 4.12: Validation accuracy and validation loss of EEGNet

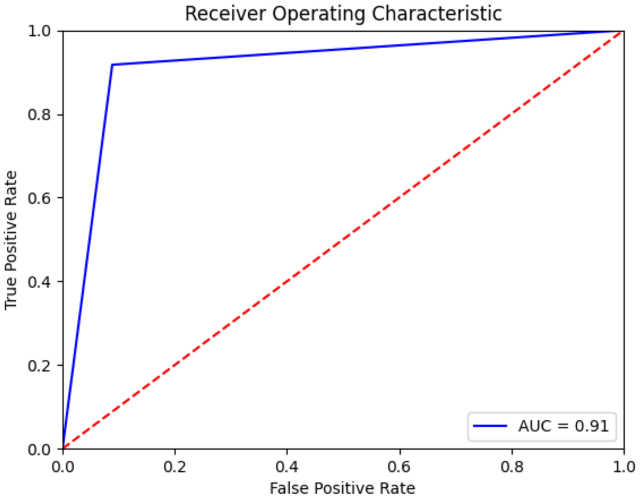


Figure 4.13: ROC Curve

on the participants, presenting a promising technique for analyzing an individual's cognitive state [4]. Within the VR task, participants engaged in two rounds of the Beat Saber game while wearing the Meta Quest 2 headset. This VR rhythm game involves participants slashing beats to exhilarating music as they approach, all within a futuristic virtual environment. According to user survey results, the VR task emerged as the most significant contributor to cognitive weariness among participants.

4.8 Discussion and Applications

This research presents an innovative method that utilizes a unique combination of physiological sensors (ECG, EDA, EMG) and brain sensors (EEG) to simultaneously identify cognitive fatigue (CF) and physical fatigue (PF). The designed task flow for data collection effectively induced CF (reported by over 80% of participants) and PF (reported by over 90% of participants), as evidenced by subjective Visual Analog Scale (VAS) scores. While the Random Forest classifier excelled in detecting PF, the LSTM model demonstrated significant success in predicting CF, eliminating the need for extensive data preprocessing and feature extraction. Overall, even the top-performing models in the system exhibited minimal failure rates, detecting actual PF in less than 12% of cases (recall=0.88) and CF in less than 10% (recall=0.90), showcasing promising outcomes. Furthermore, our models outperformed state-of-the-art approaches in detecting cognitive and physical fatigue. Future research directions will explore integrating visual sensor data to analyze facial expressions and gait movements, enhancing fatigue prediction. The inclusion of subjects with severe conditions impacted by fatigue can further enrich the study, aiding in the detection of symptoms related to those diseases.

This research holds significant promise for various practical applications. An immediate application is in the realm of human performance optimization, particularly in work environments where cognitive and physical fatigue can impact productivity and safety. Industries such as aviation, healthcare, and manufacturing could benefit from real-time fatigue monitoring to prevent errors and accidents. Additionally, the integration of physiological and brain sensors opens avenues for personalized health monitoring and wellness programs. Athletes and fitness enthusiasts could utilize this technology to optimize training regimens by understanding the interplay between cognitive and physical fatigue.

Finally, **our our main focus is to deploy the system to provide assistive care using robots for individuals with disabilities and mobility issues**. For example, a robot could assist someone with motor impairments in performing daily tasks, such as preparing lunch. While current research mainly emphasizes safe Human-Robot Cooperation (HRC) in industrial settings to ensure the well-being of human collaborators, there is a noticeable gap in investigating the cognitive or mental state of individuals engaging with robots on a daily basis. Our specific focus lies in empowering assistive robots to evaluate the Cognitive Fatigue (CF) level of their human counterparts. Moreover, the potential integration of visual sensor data to analyze facial expressions and gait movements could extend the system's utility to mental health applications, aiding in the early detection of fatigue-related symptoms in individuals with conditions such as chronic fatigue syndrome or neurological disorders. Overall, the system's innovative approach has broad implications for enhancing human performance, safety, and well-being across various domains.

CHAPTER 5

APPLICATION: ASSISTIVE ROBOTIC SYSTEM FOR COGNITIVE STATE ASSESSMENT IN INDIVIDUALS WITH SPINAL CORD INJURY

5.1 Introduction

This chapter delineates the practical application of our research in evaluating cognitive fatigue to design and create a comprehensive personalized assistive robotic system named **iRCSA** (*Intelligent Robotic Cooperation for Safe Assistance*). The primary objective of this system is to recognize, assess, and respond to Cognitive Fatigue (CF) levels in individuals with Spinal Cord Injury (SCI) during human-robot cooperation (HRC) tasks. With the increasing prevalence of robotics and Artificial Intelligence (AI), assistive robots hold immense potential for enhancing the independence and quality of life for individuals with disabilities. While existing research primarily concentrates on ensuring safe HRC in industrial settings, there exists a significant gap in comprehending the cognitive states of individuals interacting with robots in their everyday lives.

To bridge this gap, we integrate our multi-sensory system for detecting participants' CF levels and an assistive robot capable of providing corresponding support. Alongside physiological data (ECG, EDA, and EEG), audio and video data are gathered from individuals on wheelchairs during our pre-designed HRC tasks. Employing advanced machine learning algorithms, relevant features are extracted from the collected data, automatically evaluating the individual's CF level. Based on this evaluation, the iRCSA system dynamically adjusts the robot's behavior to provide personalized support. However, **our primary focus in this preliminary work is to understand the cognitive state of**

the participants during HRC tasks.

The development and evaluation of iRCSA adheres to the Participatory Action Research (PAR) approach, involving SCI subjects at every stage of the project. Their invaluable insights and feedback is taken to ensure the acceptability and usability of the proposed system. HRC scenarios, encompassing daily tasks such as cooking and preparing for work, are orchestrated to facilitate cooperative interactions between individuals with SCI and the assistive robot. The potential outcomes of this research are significant, promising to elevate the quality of life for individuals with SCI by enabling assistive robots to comprehend and respond to their cognitive state. By addressing the cognitive aspect of HRC, the iRCSA system stands to enhance the safety, efficiency, and effectiveness of assistive robotic systems in delivering support and care to individuals with SCI.

In this chapter, we conduct a thorough examination of the design of our assistive robotic system, explaining the functionality of each module: the mobile robot assistant, physiological sensor setup, facial expressions recording, and speech recognition. **As this study is in its initial stages, our emphasis lies on detailing the design, methodology, data collection, and labeling processes**. Additionally, we validate our cognitive fatigue assessment system as introduced in Chapter 4 through a preliminary analysis of the gathered sample data. Recognizing that this work is ongoing, we wrap up by delineating potential applications and proposing directions for future research.

5.2 System Design

In this section, we delve into the specifics and design of the experimental setup, as depicted in Figure 5.1. We have formulated a system in which participants engage in two Human-Robot Collaboration (HRC) tasks featuring robotic assistance, designed to simulate real-world situations. These tasks, namely ***Cooking Pasta Sauce*** and ***Getting Ready for Work***, are aimed at assisting wheelchair-bound participants in the performance of everyday activities. To ensure the seamless execution of these tasks, a robotic manipulator, illustrated in Figure. 5.2, is utilized to fetch objects for participants as required.

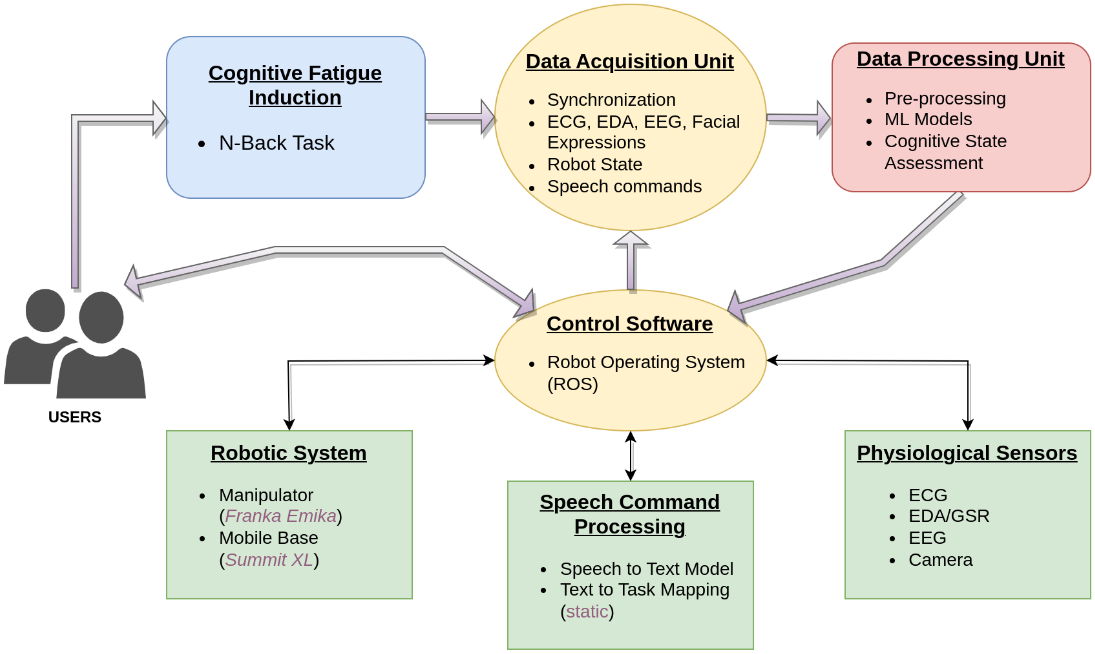
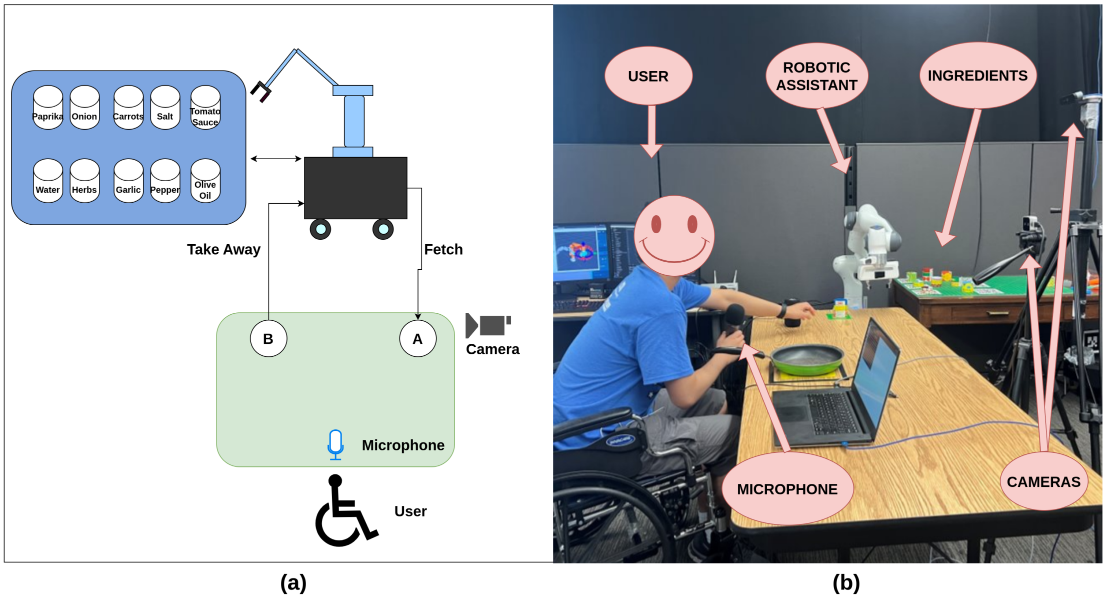


Figure 5.1: System Design: A robotic system that can monitor the user's cognitive state using physiological, vision, and audio sensors and adapt its behavior accordingly.

The robotic system is intricately linked with a multi-sensory system to gather physiological data, facilitating the assessment of participants' cognitive fatigue states. Furthermore, a speech assistant module is incorporated to aid participants in interacting with the robot using natural language. The entirety of the participant's activity, encompassing facial expressions, is recorded through RGB cameras.



(a)

(b)

Figure 5.2: Experimental Setup: (a) Overview and (b) A subject on wheelchair performing the simulated task–***Cooking Pasta Sauce***

The experimental arrangement is depicted in Figure. 5.2, featuring two tables: one designated for task execution by the subject and the other for the arrangement of required items. The figure highlights the ingredients required for the ***Cooking Pasta Sauce*** task. In this task, participants receive a list of cooking ingredients (such as tomato sauce, mushrooms, salt, etc.) to memorize before initiating the task. Subsequently, they direct the robot to retrieve each ingredient individually, commencing the pasta sauce cooking process. The sequence and quantity of ingredients fetched by the robot entirely hinge on the participant's instructions. Similarly, in the ***Getting Ready for Work*** task, the robot assists the participant in procuring six common items essential for preparing to leave for work. These items include a cellphone, laptop, headphones, keys, etc.

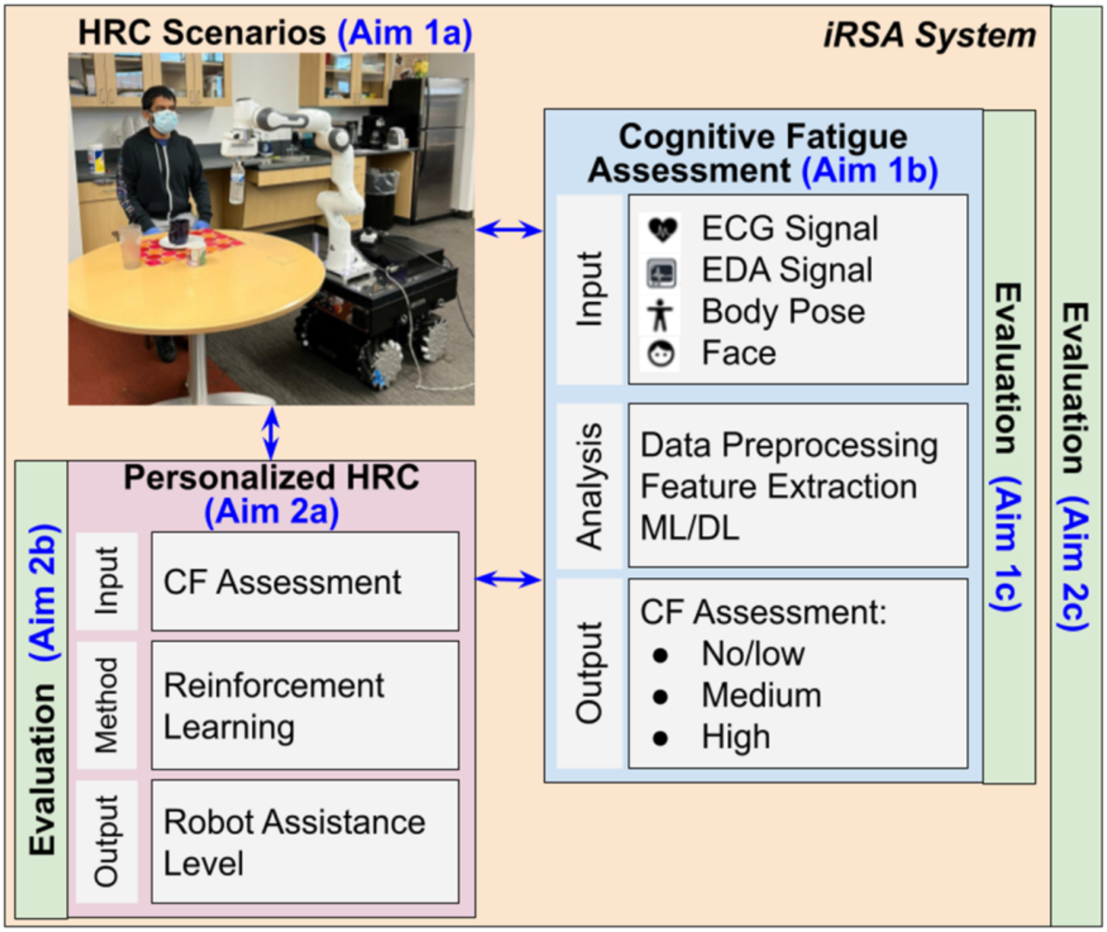


Figure 5.3: Cognitive Fatigue (CF) assessment and personalized Human-Robot Cooperation (HRC).

5.2.1 Mobile Robot Assistant

The mobile robotic assistant module within our system consists of two key components: the Summit XL omnidirectional mobile robot and the 7-DOF Franka Emika Panda robotic manipulator, commonly known as the Panda arm. The Summit XL robot enables agile and versatile movement in confined spaces. Atop its base, the Panda arm is equipped with a 2-finger gripper capable of handling objects with dimensions of up to 80mm in width and a weight of up to 3 kg.

For precise interactions with objects and enhanced safety in close proximity to humans during collaborative tasks, the Panda arm is fitted with torque sensors on each joint. Additionally, an RGBD camera is integrated into the robotic arm, improving its ability to grasp and pick up objects by providing visual information.

To facilitate navigation and interaction, the robot base is outfitted with LiDAR sensors and cameras. This sensor suite assists the robot in perceiving its surroundings and navigating through the environment effectively. Participants can guide the robots through predefined sequences of actions or make real-time adjustments using spoken commands, enabling intuitive and dynamic interaction with the mobile robotic assistant module.

5.2.2 Physiological Sensor Setup

Our human-centric framework incorporates a diverse array of physiological sensors seamlessly integrated into a multi-sensory system. This system is adept at collecting electroencephalogram (EEG), electrocardiogram (ECG), and electrodermal activity (EDA). These physiological sensors play a crucial role in evaluating cognitive fatigue (CF) levels during Human-Robot Collaboration (HRC) scenarios. The PLUX Biosignals sensor module [14] is utilized for gathering ECG and EDA data, while the Muse S headset [96] is employed for collecting EEG signals. These sensors actively monitor electrical activity in the heart, skin, and the brain and are significant in detecting the cognitive state of a person [86, 64, 70].

During various tasks performed by subjects, EEG signals are captured using the Muse S headset, which features four electrodes (AF7, AF8, TP9, TP10) strategically positioned on different areas of the head. The EEG signals provide quantifiable data on brain electrical activity, categorized into five frequency bands: alpha, beta, delta, gamma, and theta, each corresponding to a distinct brain state. As illustrated in Fig. 1.2, electrodes for ECG and EDA from the PLUX Bluetooth module are attached to different points on the body, with red and white dots denoting the front and the black dot denoting the back. Both the Muse headset and the PLUX module are connected to the ROS system via Bluetooth, ensuring a continuous stream of data to their respective topics.

5.2.3 Facial Expressions

In addition to physiological sensors, our setup incorporates two types of camera sensors to meticulously capture the physical activities and facial responses of individuals with spinal cord injury (SCI) as they engage in their daily tasks. These RGB-D cameras, equipped with both color and depth capabilities, are designed to provide a comprehensive view of the environment. One camera focuses on capturing the overall activity area (the table) where the participant performs tasks, while the other offers a close-up, dynamic view of the participant during the ***Cooking Pasta Sauce*** and ***Getting Ready for Work*** tasks.

An additional aim of our system is to extract vital features such as human body position and facial landmarks by leveraging advanced libraries like OpenPose [19] and OpenFace [8]. Recognizing human activity is pivotal, and our system endeavors to predict and interpret captured behaviors using robust computer vision algorithms. These inferred actions can provide crucial insights into cognitive fatigue levels, especially during interactions with the robot. For example, by monitoring eye motion, blink rate, and utilizing facial landmarks, we can discern different eye movement patterns, which serve as essential indicators of cognitive fatigue. However, it's important to note that the exploration of vision data goes beyond the scope of our preliminary work.

5.2.4 Speech Recognition Module

To facilitate communication between humans and the robot, we integrate a speech recognition module for receiving commands from participants and controlling the mobile robot. The speech recognition pipeline, implemented through the Google Cloud Speech library for Python [1], involves the following steps: i) adapting to ambient noise, ii) recognizing a trigger word and identifying the task keyword, and iii) dispatching the pick-and-place command to the robotic system. In the initial step, the speech recognizer adapts to ambient environmental noise for 0.5 seconds to improve speech command recognition in subsequent steps. During the command recognition stage, participants are instructed to:

- Begin by saying, "Hi/Hey Robot" to prompt the robot to listen to the command,
- Subsequently, state, "I would like to get an *item_name*" as the command to instruct the robot to fetch the specified item.

Items are fetched one at a time, with the robot initiating the task by respond-

[1] `https://cloud.google.com/speech-to-text/docs/speech-to-text-client-libraries`

ing with ”Sure, fetching *item_name* for you.” After delivering the item to the designated location, the robot notifies the user of task completion. The predetermined items for the ”Cooking Pasta Sauce” and ”Getting Ready for Work” tasks are fixed and include:

- **Cooking Pasta Sauce**: *pasta, cheese, carrots, tomato sauce, green beans, mushrooms, garlic, chili, butter, salt, bell pepper, and corn*
- **Getting Ready for Work**: *cellphone, coffee, keys, calculator, laptop, and headphones*

Upon issuing the command, we capture both the speech audio and transcribed text commands that can be used to facilitate the downstream cognitive fatigue detection task in the future. The speech audio is stored in WAV format, while the transcribed text is obtained from the Google Cloud Speech library and saved in CSV format. The rationale behind collecting speech data from participants is rooted in previous studies indicating that fatigue can lead to potential dangers, accidents, or a decline in life quality [49, 33].

5.3 Experimental Phases

The experimental study is structured into three distinct phases, each aimed at systematically assessing the impact of cognitive fatigue on performance. In the initial phase, participants commence the experiment in a rested state, devoid of cognitive load. During this baseline phase, physiological measurements (ECG, EMG, and EEG) and facial expressions are recorded through cameras.

The baseline data is crucial for normalizing signals for each participant. Following this, participants engage in N-Back tasks (2-Back) known for inducing cognitive fatigue and mental workload [99]. After each N-Back round, participants complete a VAS-F questionnaire, gauging their level of cognitive fatigue. Once moderate cognitive fatigue is induced, participants perform daily tasks (cooking pasta and getting ready for work). The severe cognitive fatigue phase is initiated when the VAS-F score surpasses 70, prompting participants to repeat the daily tasks.

Throughout the Cooking Pasta and Getting Ready for Work tasks, physiological signals and facial expressions are continuously recorded as participants interact with the robotic system. The robot aids participants in task completion, responding to speech commands issued by the participant. All data modalities are captured during these tasks, resulting in a comprehensive dataset for analysis.

5.4 Data Collection and Labeling

The recruitment of participants was a collaborative effort between the Student Access & Resource (SAR) and the Office of Accessible Education (OAE) at the university. Specifically, we reached out to members of the UTA basketball team who use wheelchairs, also known as the MovinMavs. For our initial study, we invited eight participants, ensuring a balanced representation by sex. The primary objective was to gather feedback on the design of two Human-Robot Collaboration (HRC) tasks: **Cooking Pasta Sauce** and **Getting Ready for Work**. Participants, assisted by the mobile robot assistant, engaged in these daily activities and interacted with the robot using natural language (English) through speech. The study was structured to assess task performance at three distinct phases of cognitive fatigue: ***baseline***, ***moderate***, and ***severe***.

Cognitive fatigue was induced through multiple rounds of N-back tasks, with participants completing a minimum of six rounds. Additional rounds were administered if the desired level of ***severe*** fatigue was not reached. The VAS-F questionnaire gauged the level of cognitive fatigue experienced by participants after each N-back task. Physiological sensor data, collected using Muse EEG headbands, ECG sensors, and Bioplux EMG sensors, provided insights into cognitive workload, heart rate variability, and skin sensitivity, respectively. Facial expressions captured by a camera will be used to analyze signs of stress, fatigue, and changes in emotional state. Speech commands were recorded and can be used to assess variations in participants' vocal characteristics, potentially correlating with cognitive fatigue levels.

Sensor data were synchronized using timestamps from the robot operating system (ROS). Preprocessing steps were applied to EEG, ECG, and EDA data to eliminate noise, artifacts, and baseline shifts. Spectral analysis was conducted on EEG data to extract cognitive load-related frequency bands. ECG data were processed to compute heart rate variability parameters, and EDA data were filtered and normalized for skin conductivity quantification.

The VAS-F questionnaire, providing subjective cognitive fatigue scores between 0-100, offered a general rating of overall fatigue intensity perceived by participants. Scores below **40** were categorized as *No Fatigue (< 40)*, those between **40 and 70** as *Moderate Fatigue (> 40 and < 70)*, and anything surpassing **70** as *Severe/Extreme Fatigue (> 70)*. The VAS-F scores served as a benchmark for evaluating the effectiveness of induced cognitive fatigue and validating the multi-modal data analysis.

5.5 Preliminary Analysis

The primary focus of this study is the detection of participants' cognitive fatigue states during Human-Robot Collaboration (HRC) tasks. Although vision data, speech transcriptions, and robot state are recorded during the experiments, their primary role is to contribute to the development of an intervention system, a topic that will be explored in future research. In this section, we outline the pre-processing pipeline for physiological signal data and delve into their significance in identifying the three pre-defined levels of cognitive fatigue.

For the detection of different levels of cognitive fatigue, 100 statistical features are extracted from EEG signals, and 129 combined features are derived from ECG and EDA signals. These features encompass a variety of aspects at different frequency levels, such as peaks, rates, onsets, offsets, and more. Rather than processing the entire signal for a task as a single input, we partition the temporal signals into multiple slices based on various window sizes (5 seconds, 10 seconds, and 20 seconds) for training the machine learning models. Each signal slice inherits the same label as its parent signal, and features are then extracted. This approach increases the volume of input data points for ML model training. However, we also evaluate the models using complete signal blocks as inputs. Similarly, during inference, the input signal is subdivided into smaller slices based on the window size established during training. Each slice is classified individually by the model, and ultimately, the entire signal block is classified based on the predominant class among the classified slices. This technique enhances the model's resilience to noise or outliers in the signals, as noise within certain slices may have minimal impact on the final classification result.

Despite the limited sample size of eight participants in our current preliminary study, we employ transfer learning to leverage data acquired from our previous study for cognitive fatigue detection [64] in Chapter 4. The dataset originates from a similar experimental setup involving N-back tasks designed to induce cognitive fatigue. We possess physiological sensor data (ECG, EEG, EDA, and EMG) from 32 healthy participants, who provided self-reported subjective VAS-F scores after each round of N-back. The features extracted from these samples are incorporated into the dataset from our preliminary study to facilitate the classification of the three intended levels of cognitive fatigue.

The entire aggregated dataset, comprising data from a total of 40 subjects, is randomly partitioned into training (70%, 28 subjects), validation (15%, 6 subjects), and test (15%, 6 subjects) sets. Stratified sampling is employed during the partitioning process to address potential imbalances in the dataset. Additionally, 5-fold cross-validation is performed for each of the models. Four distinct machine learning models—Logistic Regression (Log Reg.), Support Vector Machines (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) recurrent neural network—are employed in the analysis. Various combinations of features extracted from the signals are utilized to predict cognitive fatigue.

Classifiers such as Logistic Regression, SVM, and RandomForest are trained on features extracted from physiological signals. However, LSTM models (with 256 hidden layers) are trained on raw signals, given their proficiency in processing time-series data. The LSTM models follow a similar window-based method

Table 5.1: Detection of Cognitive Fatigue (CF) with EDA/GSR + EMG Features

Model	Accuracy (Window Size)				Avg. Recall
	5s	10s	20s	Full Block	
Log Reg.	68.1%	68.7%	68.9%	71.8%	0.59
SVM	**77.3%**	79.7%	80.1%	82.1%	0.68
RF	71.1%	**79.9%**	76.4%	80.9%	0.73
LSTM	68.6%	79.2%	**84.2%**	**84.5%**	**0.77**

Table 5.2: Detection of Cognitive Fatigue (CF) with EEG + EDA/GSR + EMG Features

Model	Accuracy (Window Size)				Avg. Recall
	5s	10s	20s	Full Block	
Log Reg.	64.2%	64.9%	66.7%	66.7%	0.69
SVM	**77.1%**	**80.3%**	80.3%	80.9%	0.77
RF	73.7%	77.8%	78.9%	78.8%	0.70
LSTM	68.8%	77.1%	**84.4%**	**85.7%**	**0.87**

for training, with the input size of EEG signals being *t x 20 x 1* (five frequency bands from each electrode). Conversely, ECG and EDA signals are combined to form inputs of size *t x 2 x 1*. Ultimately, the LSTM is trained on *t x 23 x 1* inputs for all signals combined, where ”t” represents the number of timesteps in the signal, varying based on the window size.

Table 5.3: Comparison of different models with the state-of-the-art algorithms

Model	**Accuracy**	**Avg. Recall**	**Ref.**
RF	64.69%	0.65	[86]
RF	66.20%	0.66	[86]
LSTM	84.1%	0.90	[64]
LSTM (Ours)	**85.7%**	0.87	Table 5.2

The Avg. Recall presented in the tables represents the average recall for **Moderate Fatigue** and **Severe Fatigue** conditions obtained across 5-fold cross-validation for each ML model. The best-performing value for each model among different window sizes is considered. Remarkably, the LSTM model outperforms others with an accuracy of **85.7%** in predicting cognitive fatigue states. The recall value of **0.87** indicates that actual fatigue cases are correctly identified 87% of the time, with only a 13% false positive rate.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

6.1 Conclusion

This dissertation delves deep into the intricate domain of frameworks for intelligent cognition assessment, investigating the union of human behavior, cognition, and technology. The research initiated with a thorough examination of the constituents that form such frameworks, explaining the roles of cognitive assessments, intelligent interfaces, and data acquisition processes, including the crucial integration of machine learning in chapters 1 & 2. Task-based cognitive assessment frameworks were then analyzed in chapter 3, with a focus on innovative tests such as the Activate Test of Embodied Cognition (ATEC) and the study of cognitive fatigue in subjects with traumatic brain injuries. The subsequent phase of the study in chapter 4 concentrated on the meticulous evaluation of cognitive fatigue using a multi-modal sensor approach, encompassing physiological sensory systems and advanced data processing techniques. Moreover, in chapter 5, we explain the application of these findings in the creation of an assistive robotic system for individuals with spinal cord injuries was explored, highlighting the practical implications of the research. The results of this comprehensive exploration not only enhance our comprehension of cognitive assessments but also present tangible applications in real-world scenarios.

6.2 Future Directions

As we conclude this dissertation, it is imperative to outline the avenues for future research stemming from this study. Firstly, ongoing investigation into the continuous refinement and expansion of intelligent cognition assessment frameworks is vitally important, with a specific emphasis on improving the sensitivity and specificity of cognitive assessments. Furthermore, the incorporation of emerging technologies, such as augmented reality and virtual reality, holds the potential to introduce new dimensions to the assessment process.

Future research endeavors should also delve deeper into the personalized nature of cognitive assessments, recognizing the diversity in cognitive profiles and tailoring assessment frameworks accordingly. Additionally, sustained attention is necessary for exploring the long-term implications of cognitive fatigue and its influence on cognitive decline. The field stands to gain from collaborative efforts between researchers and practitioners, facilitating the bridging of the gap between laboratory findings and real-world applications. This ensures the seamless integration of cognitive assessment technologies into healthcare and everyday life.

In summary, the outlined future directions underscore the dynamic and evolving nature of this field, urging researchers to persevere in advancing our understanding of human cognition and its assessment in diverse contexts.

BIBLIOGRAPHY

[1] Thomas M Achenbach, Thomas M Ruffle, et al. The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in review*, 21(8):265–271, 2000.

[2] Neusa R Adão Martins, Simon Annaheim, Christina M Spengler, and René M Rossi. Fatigue monitoring through wearables: a state-of-the-art review. *Frontiers in physiology*, page 2285, 2021.

[3] Reem S AlOmar, Nouf A AlShamlan, Saad Alawashiz, Yaser Badawood, Badr A Ghwoidi, and Hassan Abugad. Musculoskeletal symptoms and their associated risk factors among saudi office workers: a cross-sectional study. *BMC Musculoskeletal Disorders*, 22(1):1–9, 2021.

[4] A Armougum, E Orriols, A Gaston-Bellegarde, C Joie-La Marle, and P Piolino. Virtual reality: A new method to investigate cognitive load during navigation. *Journal of Environmental Psychology*, 65:101338, 2019.

[5] Under Armour. Men's ua heatgear armour sleeveless compression shirt.

[6] Marc S Atkins, William E Pelham, and Mark H Licht. A comparison of objective classroom measures and teacher ratings of attention deficit disorder. *Journal of abnormal child psychology*, 13:155–167, 1985.

[7] Ashwin Ramesh Babu, Mohammad Zakizadeh, James Robert Brady, Diane Calderon, and Fillia Makedon. An intelligent action recognition system to assess cognitive behavior for executive function disorder. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pages 164–169. IEEE, 2019.

[8] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. P. Morency. Openface 2.0: Facial behavior analysis toolkit. In *IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2018.

[9] Pouya Bashivan, Irina Rish, and Steve Heisig. Mental state recognition via wearable eeg. *arXiv preprint arXiv:1602.00985*, 2016.

[10] Tim Bayne, David Brainard, Richard W Byrne, Lars Chittka, Nicky Clayton, Cecilia Heyes, Jennifer Mather, Bence Ölveczky, Michael Shadlen, Thomas Suddendorf, et al. What is cognition? *Current Biology*, 29(13):R608–R615, 2019.

[11] Morris D Bell, Andrea J Weinstein, Brian Pittman, Richard M Gorman, and Maher Abujelala. The activate test of embodied cognition (atec): Reliability, concurrent validity and discriminant validity in a community sample of children using cognitively demanding physical tasks related to executive functioning. *Child Neuropsychology*, 27(7):973–983, 2021.

[12] A Belmont, N Agar, C Hugeron, B Gallais, and Philippe Azouvi. Fatigue and traumatic brain injury. In *Annales de réadaptation et de médecine physique*, volume 49, pages 370–374. Elsevier, 2006.

[13] Salaheddine Bendak and Hamad SJ Rashid. Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, 76:102928, 2020.

[14] Plux Biosignals. Plux biosignals: Customize your biosignalsplux kit according to your needs.

[15] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9):1063–1074, 2003.

[16] Ingrid Broch-Due, Hanne Lie Kjærstad, Lars Vedel Kessing, and Kamilla Miskowiak. Subtle behavioural responses during negative emotion reactivity and down-regulation in bipolar disorder: A facial expression and eye-tracking study. *Psychiatry research*, 266:152–159, 2018.

[17] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8*, pages 25–36. Springer, 2004.

[18] Fabio Buttussi, Luca Chittaro, Roberto Ranon, and Alessandro Verona. Adaptation of graphics and gameplay in fitness games by exploiting motion and physiological sensors. In *International Symposium on Smart Graphics*, pages 85–96. Springer, 2007.

[19] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017.

[20] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017.

[21] Nadine Chang, John A Pyles, Austin Marcus, Abhinav Gupta, Michael J Tarr, and Elissa M Aminoff. Bold5000, a public fmri dataset while viewing 5000 visual images. *Scientific data*, 6(1):1–18, 2019.

[22] Abhijit Chaudhuri and Peter Behan. Fatigue and basal ganglia. *Journal of the neurological sciences*, 179:34–42, 11 2000.

[23] Abhijit Chaudhuri and Peter O Behan. Fatigue in neurological disorders. *The lancet*, 363(9413):978–988, 2004.

[24] Pranee Chavalitsakulchai and Houshang Shahnavaz. Musculoskeletal discomfort and feeling of fatigue among female professional workers: The need for ergonomics consideration. *Journal of human ergology*, 20(2):257–264, 1991.

[25] Jerry Chen, Maysam Abbod, and Jiann-Shing Shieh. Pain and stress detection using wearable sensors and devices—a review. *Sensors*, 21(4):1030, 2021.

[26] Amy Chinner, Jasmine Blane, Claire Lancaster, Chris Hinds, and Ivan Koychev. Digital technologies for the assessment of cognition: a clinical review. *Evidence-based mental health*, 21(2):67–71, 2018.

[27] Trevor T-J Chong, Matthew Apps, Kathrin Giehl, Annie Sillence, Laura L Grima, and Masud Husain. Neurocomputational mechanisms underlying subjective valuation of effort costs. *PLoS biology*, 15(2):e1002598, 2017.

[28] C Christodoulou, J DeLuca, JH Ricker, NK Madigan, BM Bly, G Lange, AJ Kalnin, WC Liu, J Steffener, BJ Diamond, et al. Functional magnetic resonance imaging of working memory impairment after traumatic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*, 71(2):161–168, 2001.

[29] Mario Cifrek, Vladimir Medved, Stanko Tonković, and Saša Ostojić. Surface emg based muscle fatigue evaluation in biomechanics. *Clinical biomechanics*, 24(4):327–340, 2009.

[30] Robert W Cox. Afni: software for analysis and visualization of func-

tional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.

[31] Catherine L Davis and Stephanie Cooper. Fitness, fatness, cognition, behavior, and academic achievement among overweight children: do cross-sectional associations correspond to exercise trial outcomes? *Preventive medicine*, 52:S65–S69, 2011.

[32] Michael E Dawson, Anne M Schell, and Christopher G Courtney. The skin conductance response, anticipation, and decision-making. *Journal of Neuroscience, Psychology, and Economics*, 4(2):111, 2011.

[33] Carla Aparecida de Vasconcelos, Maurílio Nunes Vieira, Göran Kecklund, and Hani Camille Yehia. Speech analysis for fatigue and sleepiness detection of a pilot. *Aerospace medicine and human performance*, 90(4):415–418, 2019.

[34] John DeLuca, Helen M Genova, Frank G Hillary, and Glenn Wylie. Neural correlates of cognitive fatigue in multiple sclerosis using functional mri. *Journal of the neurological sciences*, 270(1-2):28–39, 2008.

[35] Alex Dillhoff, Konstantinos Tsiakas, Ashwin Ramesh Babu, Mohammad Zakizadehghariehali, Benjamin Buchanan, Morris Bell, Vassilis Athitsos, and Fillia Makedon. An automated assessment system for embodied cognition in children: from motion data to executive functioning. In *Proceedings of the 6th international Workshop on Sensor-based Activity Recognition and Interaction*, pages 1–6, 2019.

[36] Ekaterina Dobryakova, Helen Genova, John Deluca, and Glenn Wylie. The dopamine imbalance hypothesis of fatigue in multiple sclerosis and other neurological disorders. *Frontiers in Neurology*, 6, 03 2015.

[37] Ekaterina Dobryakova, Hanneke E Hulst, Angela Spirou, Nancy D Chiaravalloti, Helen M Genova, Glenn R Wylie, and John DeLuca. Fronto-striatal network activation leads to less fatigue in multiple sclerosis. *Multiple Sclerosis Journal*, 24(9):1174–1182, 2018.

[38] Willem Einthoven, G Fahr, and A De Waart. On the direction and manifest size of the variations of potential in the human heart and on the influence of the position of the heart on the form of the electrocardiogram. *American heart journal*, 40(2):163–211, 1950.

[39] Farnaz Farahanipad, Harish Ram Nambiappan, Ashish Jaiswal, Maria Kyrarini, and Fillia Makedon. Hand-reha: dynamic hand gesture recognition for game-based wrist rehabilitation. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–9, 2020.

[40] Mark Faul, Marlena M Wald, Likang Xu, and Victor G Coronado. Traumatic brain injury in the united states: emergency department visits, hospitalizations, and deaths, 2002-2006. 2010.

[41] Annalisa Franco, Antonio Magnani, and Dario Maio. A multimodal approach for human activity recognition based on skeleton and rgb data. *Pattern Recognition Letters*, 131:293–299, 2020.

[42] Roy Freeman and Anthony L Komaroff. Does the chronic fatigue syndrome involve the autonomic nervous system? *The American journal of medicine*, 102(4):357–364, 1997.

[43] Harshala Gammulle, Simon Denman, Sridha Sridharan, and Clinton Fookes. Two stream lstm: A deep fusion framework for human action recognition. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 177–186. IEEE, 2017.

[44] Richard C Gershon, David Cella, Nathan A Fox, Richard J Havlik, Hugh C Hendrie, and Molly V Wagster. Assessment of neurological and behavioural function: the nih toolbox. *The Lancet Neurology*, 9(2):138–139, 2010.

[45] Michael A Goodrich, Alan C Schultz, et al. Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3):203–275, 2008.

[46] Google. Movenet: Ultra fast and accurate pose detection model.

[47] Klaus Gramann, Joseph T Gwin, Daniel P Ferris, Kelvin Oie, Tzyy-Ping Jung, Chin-Teng Lin, Lun-De Liao, and Scott Makeig. Cognition in action: imaging brain/body dynamics in mobile humans. 2011.

[48] Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45, 2012.

[49] Harold P Greeley, Eric Friets, John P Wilson, Sridhar Raghavan, Joseph

Picone, and Joel Berg. Detecting fatigue from voice using speech recognition. In *2006 IEEE International Symposium on signal processing and information technology*, pages 567–571. IEEE, 2006.

[50] Yao Guo, Xiangyu Liu, Shun Peng, Xinyu Jiang, Ke Xu, Chen Chen, Zeyu Wang, Chenyun Dai, and Wei Chen. A review of wearable and unobtrusive sensing technologies for chronic disease management. *Computers in Biology and Medicine*, 129:104163, 2021.

[51] Peter Hagell and Lena Brundin. Towards an understanding of fatigue in parkinson disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 80(5):489–492, 2009.

[52] James Hale, Vincent Alfonso, V Berninger, Bruce Bracken, Catherine Christo, Elaine Clark, M Cohen, Andrew Davis, Scott Decker, M Denckla, et al. Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly*, 33(3):223–236, 2010.

[53] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Learning spatio-temporal features with 3d residual networks for action recognition. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 3154–3160, 2017.

[54] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[55] Ernst A Heinz, Kai S Kunze, Matthias Gruber, David Bannach, and Paul Lukowicz. Using wearable sensors for real-time recognition tasks in games of martial arts-an initial experiment. In *2006 IEEE Symposium on Computational Intelligence and Games*, pages 98–102. IEEE, 2006.

[56] Christian Herff, Dominic Heger, Ole Fortmann, Johannes Hennrich, Felix Putze, and Tanja Schultz. Mental workload during n-back task—quantified in the prefrontal cortex using fnirs. *Frontiers in human neuroscience*, 7:935, 2014.

[57] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[58] Shitong Huang, Jia Li, Pengzhu Zhang, and Weiqiang Zhang. Detection of mental fatigue state with wearable ecg devices. *International journal of medical informatics*, 119:39–46, 2018.

[59] Josie Hughes and Fumiya Iida. Multi-functional soft strain sensors for wearable physiological monitoring. *Sensors*, 18(11):3822, 2018.

[60] Curtis S Ikehara and Martha E Crosby. Assessing cognitive load with physiological sensors. In *Proceedings of the 38th annual hawaii international conference on system sciences*, pages 295a–295a. IEEE, 2005.

[61] Shamsi T Iqbal, Piotr D Adamczyk, Xianjun Sam Zheng, and Brian P Bailey. Towards an index of opportunity: understanding changes in mental workload during task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 311–320, 2005.

[62] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A survey on contrastive self-supervised learning. *Technologies*, 9(1):2, 2020.

[63] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Fillia Makedon, and Glenn Wylie. Understanding cognitive fatigue from fmri scans with self-supervised learning. *arXiv preprint arXiv:2106.15009*, 2021.

[64] Ashish Jaiswal, Mohammad Zaki Zadeh, Aref Hebri, Ashwin Ramesh Babu, and Fillia Makedon. A smart sensor suit (sss) to assess cognitive and physical fatigue with machine learning. In *International Conference on Human-Computer Interaction*, pages 120–134. Springer, 2023.

[65] Budi Thomas Jap, Sara Lal, Peter Fischer, and Evangelos Bekiaris. Using eeg spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications*, 36(2):2352–2359, 2009.

[66] Qiang Ji, Peilin Lan, and Carl Looney. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, 36(5):862–875, 2006.

[67] Michael J Kane, Andrew RA Conway, Timothy K Miura, and Gregory JH Colflesh. Working memory, attention control, and the n-back task: a question of construct validity. *Journal of Experimental psychology: learning, memory, and cognition*, 33(3):615, 2007.

[68] Georgios Kapidis, Ronald Poppe, Elsbeth Van Dam, Lucas Noldus, and Remco Veltkamp. Egocentric hand track and object-based human action recognition. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 922–929. IEEE, 2019.

[69] Rithik Kapoor, Ashish Jaiswal, and Fillia Makedon. Light-weight seated posture guidance system with machine learning and computer vision. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 595–600, 2022.

[70] Enamul Karim, Hamza Reza Pavel, Ashish Jaiswal, Mohammad Zaki Zadeh, Michail Theofanidis, Glenn Wylie, and Fillia Makedon. An eeg-based cognitive fatigue detection system. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 131–136, 2023.

[71] Gary Kielhofner. *A model of human occupation: Theory and application*. Lippincott Williams & Wilkins, 2002.

[72] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[73] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5253–5263, 2020.

[74] Alexander D Kohl, Glenn R Wylie, HM Genova, Frank Gerard Hillary, and J Deluca. The neural correlates of cognitive fatigue in traumatic brain injury using functional mri. *Brain injury*, 23(5):420–432, 2009.

[75] Bon Mi Koo and Lisa M Vizer. Mobile technology for cognitive assessment of older adults: a scoping review. *Innovation in aging*, 3(1):igy038, 2019.

[76] Annica Kristoffersson and Maria Lindén. A systematic review of wearable sensors for monitoring physical activity. *Sensors*, 22(2):573, 2022.

[77] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[78] Gerald P Krueger. Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work & Stress*, 3(2):129–141, 1989.

[79] Lauren B Krupp, Luis A Alvarez, Nicholas G LaRocca, and Labe C Scheinberg. Fatigue in multiple sclerosis. *Archives of neurology*, 45(4):435–437, 1988.

[80] Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, 10:90–115, 2013.

[81] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

[82] Wei-Han Lee and Ruby B Lee. Implicit smartphone user authentication with sensors and contextual machine learning. In *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 297–308. IEEE, 2017.

[83] Tao Lin, Masaki Omata, Wanhua Hu, and Atsumi 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*, pages 1–10. Citeseer, 2005.

[84] Simon P Liversedge and John M Findlay. Saccadic eye movements and cognition. *Trends in cognitive sciences*, 4(1):6–14, 2000.

[85] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015.

[86] Hongyu Luo, Pierre-Alexandre Lee, Ieuan Clay, Martin Jaggi, and Valeria De Luca. Assessment of fatigue using wearable sensors: a pilot study. *Digital biomarkers*, 4(1):59–72, 2020.

[87] Haojie Ma, Wenzhong Li, Xiao Zhang, Songcheng Gao, and Sanglu Lu. Attnsense: Multi-level attention mechanism for multimodal human activity recognition. In *IJCAI*, pages 3109–3115, 2019.

[88] Mehran Maghoumi and Joseph J LaViola. Deepgru: Deep gesture recognition utility. In *Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7–9, 2019, Proceedings, Part I 14*, pages 16–31. Springer, 2019.

[89] Scott Makeig, Klaus Gramann, Tzyy-Ping Jung, Terrence J Sejnowski, and Howard Poizner. Linking brain, mind and behavior. *International Journal of Psychophysiology*, 73(2):95–100, 2009.

[90] Dominique Makowski, Tam Pham, Zen Juen Lau, Jan Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Chen. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53, 02 2021.

[91] Andrea Mannini, Diana Trojaniello, Andrea Cereatti, and Angelo M Sabatini. A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients. *Sensors*, 16(1):134, 2016.

[92] Daniel McDuff, Sarah Gontarek, and Rosalind Picard. Remote measurement of cognitive stress via heart rate variability. In *2014 36th annual international conference of the IEEE engineering in medicine and biology society*, pages 2957–2960. IEEE, 2014.

[93] Ryan S McGinnis, Nikhil Mahadevan, Yaejin Moon, Kirsten Seagers, Nirav Sheth, John A Wright Jr, Steven DiCristofaro, Ikaro Silva, Elise Jortberg, Melissa Ceruolo, et al. A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis. *PloS one*, 12(6):e0178366, 2017.

[94] Beat Meier, Nicolas Rothen, and Stefan Walter. Developmental aspects of synaesthesia across the adult lifespan. *Frontiers in human neuroscience*, 8:129, 2014.

[95] Maria Laura Mele and Stefano Federici. Gaze and eye-tracking solutions for psychological research. *Cognitive processing*, 13:261–265, 2012.

[96] MUSE. Muse s - the next generation of muse s.

[97] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress.

[98] Richard Nelesen, Yasmin Dar, KaMala Thomas, and Joel E Dimsdale. The relationship between fatigue and cardiac functioning. *Archives of internal medicine*, 168(9):943–949, 2008.

[99] Adrian M Owen, Kathryn M McMillan, Angela R Laird, and Ed Bullmore. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping*, 25(1):46–59, 2005.

[100] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, pages 230–236, 1985.

[101] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas Huang. Human computing and machine understanding of human behavior: A survey. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 239–248, 2006.

[102] Michalis Papakostas, Akilesh Rajavenkatanarayanan, and Fillia Makedon. Cogbeacon: A multi-modal dataset and data-collection platform for modeling cognitive fatigue. *Technologies*, 7(2):46, 2019.

[103] Hamza Reza Pavel, Enamul Karim, Ashish Jaiswal, Sneh Acharya, Gaurav Nale, Michail Theofanidis, and Fillia Makedon. Assessment of cognitive fatigue from gait cycle analysis. *Technologies*, 11(1):18, 2023.

[104] Hamza Reza Pavel, Enamul Karim, Mohammad Zaki Zadeh, Ashish Jaiswal, Rithik Kapoor, and Fillia Makedon. Automated system to measure static balancing in children to assess executive function. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 569–575, 2022.

[105] Vladimir I Pavlovic, Rajeev Sharma, and Thomas S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):677–695, 1997.

[106] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[107] Hai Pham, Paul Pu Liang, Thomas Manzini, Louis-Philippe Morency, and Barnabás Póczos. Found in translation: Learning robust joint representa-

tions by cyclic translations between modalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6892–6899, 2019.

[108] Majid Ali Khan Quaid and Ahmad Jalal. Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm. *Multimedia Tools and Applications*, 79(9):6061–6083, 2020.

[109] Md Mustafizur Rahman, Ajay Krishno Sarkar, Md Amzad Hossain, Md Selim Hossain, Md Rabiul Islam, Md Biplob Hossain, Julian MW Quinn, and Mohammad Ali Moni. Recognition of human emotions using eeg signals: A review. *Computers in Biology and Medicine*, 136:104696, 2021.

[110] Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Ashish Jaiswal, Alexis Lueckenhoff, Maria Kyrarini, and Fillia Makedon. A multi-modal system to assess cognition in children from their physical movements. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 6–14, 2020.

[111] Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*, (3):257–266, 1983.

[112] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[113] Harry T Reis, W Andrew Collins, and Ellen Berscheid. The relationship context of human behavior and development. *Psychological bulletin*, 126(6):844, 2000.

[114] David Richley. New training and qualifications in electrocardiography. *British Journal of Cardiac Nursing*, 8(1):38–42, 2013.

[115] Markus Ringnér. What is principal component analysis? *Nature Biotechnology*, 26:303–304, 2008.

[116] Samuel Rota, Baptiste Morel, Damien Saboul, Isabelle Rogowski, and Christophe Hautier. Influence of fatigue on upper limb muscle activity and performance in tennis. *Journal of Electromyography and Kinesiology*, 24(1):90–97, 2014.

[117] Laura Sevilla-Lara, Yiyi Liao, Fatma Güney, Varun Jampani, Andreas Geiger, and Michael J Black. On the integration of optical flow and action recognition. In *Pattern Recognition: 40th German Conference, GCPR 2018, Stuttgart, Germany, October 9-12, 2018, Proceedings 40*, pages 281–297. Springer, 2019.

[118] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1010–1019, 2016.

[119] Anita Valanju Shelgikar, Patricia F Anderson, and Marc R Stephens. Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest*, 150(3):732–743, 2016.

[120] Ben Shneiderman and Pattie Maes. Direct manipulation vs. interface agents. interactions, 4 (6): 42–61. *Google Scholar Google Scholar Digital Library Digital Library*, 1997.

[121] Steven Shorrock. Four kinds of 'human factors': 2. factors of humans, Mar 2019.

[122] Dee Unglaub Silverthorn. *Human physiology*. Jones & Bartlett Publishers, 2015.

[123] Martin J Sliwinski, Jacqueline A Mogle, Jinshil Hyun, Elizabeth Munoz, Joshua M Smyth, and Richard B Lipton. Reliability and validity of ambulatory cognitive assessments. *Assessment*, 25(1):14–30, 2018.

[124] Oliver Wasenmüller and Didier Stricker. Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In *Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II 13*, pages 34–45. Springer, 2017.

[125] John B Watson and Gregory A Kimble. *Behaviorism*. Routledge, 2017.

[126] GR Wylie, E Dobryakova, J DeLuca, N Chiaravalloti, K Essad, and H Genova. Cognitive fatigue in individuals with traumatic brain injury is associated with caudate activation. *Scientific reports*, 7(1):1–12, 2017.

[127] Wenyuan Xu, Chen Yan, Weibin Jia, Xiaoyu Ji, and Jianhao Liu. Analyzing

and enhancing the security of ultrasonic sensors for autonomous vehicles. *IEEE Internet of Things Journal*, 5(6):5015–5029, 2018.

[128] Mohammad Zaki Zadeh, Ashwin Ramesh Babu, Ashish Jaiswal, Maria Kyrarini, Morris Bell, and Fillia Makedon. Automated system to measure tandem gait to assess executive functions in children. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*, pages 167–170, 2021.

[129] Mohammad Zaki Zadeh, Ashwin Ramesh Babu, Ashish Jaiswal, and Fillia Makedon. Self-supervised human activity representation for embodied cognition assessment. *Technologies*, 10(1):33, 2022.