INTELLIGENT SYSTEM FOR THE CLASSIFICATION OF MENTAL STATE PARAMETERS

A Thesis

Submitted in partial fulfilment for the Dual Degree of

Doctor of Philosophy under the

Faculty of Engineering, Amrita Vishwa Vidyapeetham, India

and

Doctor of Philosophy in

Materials, Mechatronics and Systems Engineering

University of Trento, Italy

by

JYOTSNA C.

COMPUTER SCIENCE AND ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

BENGALURU CAMPUS (INDIA)

DEPARTMENT OF INDUSTRIAL ENGINEERING

DOCTORAL SCHOOL IN MATERIALS, MECHATRONICS AND SYSTEMS ENGINEERING

UNIVERSITY OF TRENTO (ITALY)

July 2024

AMRITA SCHOOL OF COMPUTING
AMRITA VISHWA VIDYAPEETHAM
BENGALURU CAMPUS

DOCTORAL SCHOOL IN MATERIALS, MECHATRONICS AND SYSTEMS ENGINEERING
UNIVERSITY OF TRENTO
ITALY

BONAFIDE CERTIFICATE

This is to certify that the thesis entitled "**Intelligent System for the Classification of Mental State Parameters**" submitted by **Jyotsna C., BL.EN.D*CSE16001** for the award of the **Dual Degree of Doctor of Philosophy** under the **Faculty of Engineering** and **Doctor of Philosophy** in **Materials, Mechatronics and Systems Engineering** is a bonafide record of the work carried out by her under our guidance and supervision at "Department of Computer Science and Engineering", Amrita Vishwa Vidyapeetham, Bengaluru Campus and at "Department of Industrial Engineering", University of Trento, Italy as part of a cotutelle agreement between Amrita Vishwa Vidyapeetham, India and University of Trento, Italy.

Dr. Amudha J.
Thesis Advisor,
Professor,
Department of Computer Science and Engineering,
Amrita Vishwa Vidyapeetham,
Bengaluru Campus,
India.

Dr. Giandomenico Nollo
Thesis Co-Advisor,
Associate Professor,
Department of Industrial Engineering,
University of Trento,
Trento,
Italy.

AMRITA SCHOOL OF COMPUTING
AMRITA VISHWA VIDYAPEETHAM
BENGALURU CAMPUS

DOCTORAL SCHOOL IN MATERIALS, MECHATRONICS AND SYSTEMS ENGINEERING
UNIVERSITY OF TRENTO
ITALY

I, **Jyotsna C.**, **BL.EN.D*CSE16001**, hereby declare that this thesis entitled "**Intelligent System for the Classification of Mental State Parameters**" is the record of the original work done by me under the guidance of **Dr. Amudha J.**, Professor, Department of CSE, Amrita Vishwa Vidyapeetham, Bengaluru Campus and **Dr. Giandomenico Nollo**, Associate Professor, Department of Industrial Engineering, University of Trento, Italy. To the best of my knowledge, this work has not formed the basis for the award of any degree/diploma/associateship/fellowship or a similar award to any candidate in any University.

Place: Bengaluru
Date: 25/07/2024

Signature of the Student

COUNTERSIGNED

Dr. Amudha J.
Thesis Advisor,
Professor,
Department of Computer Science and Engineering,
Amrita Vishwa Vidyapeetham,
Bengaluru Campus,
India.

Dr. Giandomenico Nollo
Thesis Co-Advisor,
Associate Professor,
Department of Industrial Engineering,
University of Trento,
Trento,
Italy.

Contents

Acknowledgements

The satisfaction that accompanies successful completion of any task would be incomplete without mention of people who made it possible, and whose constant encouragement and guidance have been source of inspiration throughout the course of this thesis work.

I offer my sincere pranams at the lotus feet of "AMMA", MATA AMRITANANDAMAYI DEVI who showered her blessing upon me throughout the course of this research work.

I owe my gratitude to Prof. Manoj P., (Director, Amrita School of Engineering, Bangalore), Dr. Bharat Jayaraman (Dean, Amrita School of Computing, Bengaluru), Dr. Sriram Devanathan (Principal, Amrita School of Engineering, Bengaluru) and Dr. Gopalakrishnan E. A., (Principal, Amrita School of Computing, Bengaluru) for their support and inspiration. I would like to use this opportunity to express my sincere gratitude to Amrita Vishwa Vidyapeetham for providing the facilities and necessary infrastructure to conduct the research.

It is with great pleasure that I express my gratitude and indebtedness to my thesis advisor, Dr. Amudha J., Professor in the Department of Computer Science and Engineering at Amrita School of Computing, Bangalore. Her invaluable guidance, encouragement, moral support, and affection have been instrumental throughout the thesis work. Her profound knowledge in the research area has served as a guiding light, enabling me to navigate through challenges and achieve milestones in this endeavor. Her continuous motivation and unwavering support provided me with the confidence and determination to undertake research work at the Behavior Analysis Lab at the University of Trento, Italy.

I extend my heartfelt appreciation to Prof. Giandomenico Nollo, Associate Professor in the Department of Industrial Engineering at the University of Trento, Italy, my thesis co-advisor, for his exceptional support and guidance throughout the research process. Prof. Nollo's willingness to mentor and advise me has been instrumental in enabling me to conduct research at their Behavioral Analysis Lab at the University of Trento. His expertise and encouragement have been invaluable in facilitating the progress and success of this endeavor.

I am grateful for the valuable suggestions and insightful discussions provided by my doctoral committee members, Dr. Amritanshu Ram (Subject Expert), Dr. Mariolino De Cecco (External Member), and Dr. B. Venkatesh (Faculty from Minor Area), during the review presentations. Their expertise and guidance have greatly contributed to the enhancement of my work. Additionally, I extend my thanks to past committee members, Dr. Rakesh S.G and Dr. Manoj Kumar Panda, along with Dr. Raghavendra Rao, for their unwavering support in their respective roles.

I express my sincere gratitude to Dr. Amritanshu Ram, Head of the Department of Yoga at HCG, Bengaluru; Dr. Raghavendra, Director at the Central Council for Research in Yoga and Naturopathy, Ministry of AYUSH, Government of India; and Dr. Ravi Nayar, former Dean of Academics at HCG, Bengaluru, for their invaluable assistance in obtaining ethical committee approval for data collection from HCG, Bangalore, and for their unwavering support throughout the data collection process. I am also deeply thankful to Prof. Giandomenico Nollo, Prof. Claudio Mulatti, Dr. Damiano Fruet, and Ms. Deborah Ferrante for their invaluable assistance in data collection from the University of Trento, Italy. I extend my heartfelt thanks to Ms. Namratha, Professional Counselor at Amrita School of Engineering, Bangalore, for her invaluable assistance in the research study.

I sincerely thank Dr. Rajkumar Elagiri Ramalingam, Former Chief Operating Officer and Principal Data Scientist at IIT Madras, Chennai, for his invaluable criticism and insightful feedback during the comprehensive examination. I am grateful to all my course mentors, Dr. Amudha J., Dr. Geetha K. N., and Dr. Srinivas Rao, for their invaluable contributions in broadening my knowledge base, which proved immensely beneficial during the research process.

I extend my gratitude to Amrita Vishwa Vidyapeetham for the infrastructure at the Eye Tracking and Computer Vision Research Lab, which was crucial for this study. Special thanks to the dedicated lab members for their invaluable support, and to the participants of the eye-tracking experiment for their contribution. I also thank my friends and well-wishers at Amrita School of Computing, Bengaluru, for their encouragement and support.

I am filled with immense gratitude for the unwavering support of my beloved husband, Mr. Ranjith Sundaram, whose encouragement has been my guiding light throughout this journey. My heartfelt thanks to my daughters, Esha R. and Diya R., for their cooperation and understanding. My dear parents, Mr. P. Chandrasekharan and Ms. K. Meenakumari, and my beloved in-laws, Mr. Sundaram K.G. and Ms. Reetha S., your unwavering love and encouragement have been a constant source of strength and inspiration.

(JYOTSNA C.)

List of Figures

List of Tables

Abbreviations

ACF	–	Autocorrelation Function
AD	–	Alzheimer's Disease
ADHD	–	Attention Deficit Hyperactivity Disorder
AI	–	Artificial Intelligence
ALS	–	Amyotrophic Lateral Sclerosis
ANFIS	–	Adaptive Neuro-Fuzzy Inference System
ANN	–	Artificial Neural Network
ANOVA	–	Analysis of Variance
AOI	–	Area of Interest
ARIMA	–	Auto-Regressive Moving Average
ASD	–	Autism Spectrum Disorder
AUC	–	Area Under the Curve
CI	–	Cognitive Impairment
CL	–	Cognitive Load
CNN	–	Convolutional Neural Network
DL	–	Deep Learning
DT	–	Decision Tree
ECG	–	Electrocardiogram
EEG	–	Electroencephalogram
ETMT	–	Eye Tracking Based Trail Making Test
FES	–	Fuzzy Expert System
FIS	–	Fuzzy Inference System
GSR	–	Galvanic Skin Response
IAPS	–	International Affective Picture System
IDT	–	Identification by Dispersion Threshold
KNN	–	K-Nearest Neighbour
LDA	–	Linear Discriminant Analysis
LR	–	Logistic Regression
LSL	–	Lab Streaming Layer
LSTM	–	Long Short-Term Memory
MAE	–	Mean Absolute Error
MAPE	–	Mean Absolute Percentage Error
MCI	–	Mild Cognitive Impairment
ME	–	Mean Error
ML	–	Machine Learning

MPE	–	Mean Percentage Error
MMSE	–	Mini-Mental State Examination
MoCA	–	Montreal Cognitive Assessment
MS	–	Multiple Sclerosis
MSE	–	Mean Squared Error
PACF	–	Partial Autocorrelation Function
PD	–	Parkinson's Disease
PPG	–	Photoplethysmography
RMSE	–	Root Mean Squared Error
RNN	–	Recurrent Neural Network
ROC	–	Receiver Operating Characteristic
RSS	–	Residual Sum of Squares
SAM	–	Self-Assessment Manikin
SMI	–	SensoMotoric Instruments
SVM	–	Support Vector Machine
SWAT	–	Subjective Workload Assessment Technique
TBI	–	Traumatic Brain Injury
TD	–	Typically Developing
TMT	–	Trail Making Test

Abstract

Mental health is essential for overall well-being, focusing emotional, psychological, and social aspects. Assessing and managing mental health requires understanding mental state parameters, including cognitive load, cognitive impairment, and emotional state. Advanced technologies like eye tracking provide valuable insights into these parameters, transformed mental health evaluation and enabled more targeted interventions and better outcomes.

Thesis focused towards developing intelligent system to monitor mental health, focusing on cognitive load, cognitive impairment, and emotional state. The research has three main objectives, including creating four eye-tracking-based unimodal datasets and a multimodal dataset to address the lack of publicly available mental health assessment datasets. Each dataset is designed to study cognitive load, cognitive impairment, and emotional state classification using varied stimuli.

In addition to dataset creation, the thesis excels in feature extraction, introducing novel features to detect mental state parameters and enhancing assessment precision. High-level features such as error rate, scanpath comparison score, and inattentional blindness are incorporated, contributing to find cognitive impairment scores.

Five models are developed to detect mental states by separately monitoring the mental state parameters, cognitive load, cognitive impairment, and emotional state. The models employ statistical analysis, machine learning algorithms, fuzzy inference systems, and deep learning techniques to provide detailed insights into an individual's mental state.

The first two models, Eye-Tracking Cognitive Load models (ECL-1 and ECL-2) focus on cognitive load assessment during mathematical assessments and Trail Making Test tasks. ECL-1 model utilizes statistical analysis to understand the correlation between eye tracking features like pupil diameter and blink frequency with the cognitive load while performing mathematical assessments. With the identification of relevant features while performing Trail Making Test (TMT), the ECL-2 model effectively classifies low and high cognitive load states with a notable 94% accuracy, utilizing eye-tracking data and machine learning algorithms.

The third model, the ETMT (Eye tracking based Trail Making Test) model,

uses a fuzzy inference system and adaptive neuro-fuzzy inference system to detect mental states associated with cognitive impairment. It provides detailed scores in visual search speed and focused attention, important for understanding the exact cognitive deficits of a patient. This greatly aids in understanding the cognitive states of an individual and addresses deficits in executive functioning, memory, motor function, attentional disengagement, neuropsychological function, processing speed, and visual attention.

The fourth model, PredictEYE, utilizes a deep learning time-series univariate regression model based on Long Short-Term Memory (LSTM) to predict future sequences of each feature. Machine learning-based Random Forest algorithm is applied on the predicted features for mental state prediction and identifying the mental state as calm or stressful based on a person's emotional state. The personalized time series methodology makes use of the power of time series analysis, identifying patterns and changes in data over time to enable more precise and individualized mental health assessments and monitoring. Notably, PredictEYE outperforms ARIMA with an accuracy of 86.4%.

The fifth model introduced in this study is based on a multimodal dataset, incorporating physiological measures such as ECG, GSR, PPG, and respiratory signals, along with eye tracking data. Two separate models, one based on eye tracking data and the other based on all other physiological measures developed for understanding the emotional state of a person. These models demonstrate comparable performance, with notable proficiency in binary classification based on arousal and valence. Particularly, the Binary-Valence model achieves slightly higher accuracy when utilizing eye tracking data, while other physiological measures exhibit stronger classification performance for the Binary-Arousal model.

The thesis makes substantial progress in mental health monitoring by providing accurate, non-intrusive evaluations of an individual's mental state. It emphasizes mental state parameters such as cognitive load, impairment, and emotional state, with AI-based methods incorporated to improve the precision in detection of mental state.

Chapter 1

Introduction

1.1 Mental Health - State of Well-being

Mental health is a person's emotional, psychological, and social well-being [1–3]. It encompasses a person's ability to handle stress, make choices, maintain fulfilling relationships, and cope with life's challenges. Good mental health doesn't mean the absence of difficulties but rather the ability to manage them positively and effectively [4]. It's about achieving balance in one's emotions, thoughts, and behaviors, contributing to an overall sense of contentment and fulfillment.

However, when the delicate balance of mental health is disrupted, individuals may experience medical conditions known as mental disorders. It can affect a person's normal cognitive, emotional, and behavioral functioning. These conditions can range from mild to severe and may encompass disorders such as depression, anxiety, schizophrenia, bipolar disorder, and more. Mental disorders can significantly impact a person's thoughts, feelings, and actions, often leading to distress and impairment in daily life [5]. Proper diagnosis, treatment, and support are crucial for managing mental disorders effectively.

Mental health, when reaching its optimal stage, is defined as mental wellness, where an individual experiences a positive sense of self, effective coping mechanisms, and the ability to function well in various aspects of life [6]. It involves a balance between cognitive, behavioral, and emotional dimensions. Mental wellness is characterized by self-awareness, resilience, positive relationships, a sense of purpose, and the ability to adapt to change [7]. Conversely, when mental health falls below a specific threshold, it may indicate the presence of a mental disorder. This distinction underscores the spectrum within mental health, ranging from the flourishing state of mental wellness to potential challenges that warrant clinical attention.

The growth and development of society are intrinsically linked to the health of its members, making health a fundamental cornerstone for happiness and well-being worldwide. In this context, mental health stands out as an integral and crucial component of overall health, impacting individuals of all ages and genders. With the rising prevalence of brain and mind-related health conditions, regular monitoring of mental health has become imperative for everyone in society [8, 9].

There is a complex relationship between mental and physical health. Mental illnesses like anxiety and depression can make it difficult for people to practice healthy habits, which emphasizes their critical role in preserving general well-being [10, 11]. Mental health care faces more obstacles than physical health care, including lack of mental health professionals that prevent people from accessing the treatment they need [12].

Many countries worldwide rely on primary healthcare systems like community health centers, primary health clinics, regional hospitals, and specialized institutions. Integrating mental health care into these primary care programs is a beneficial approach in addressing mental health conditions. This could allow patients to receive personalized care [13].

Mental health professionals shortage and unavailability of necessary care are significant challenges to comprehensive mental health services [3]. Implementing an assistive model can streamline operations, provide more time for patients with serious conditions, and ensure mental health services are more accessible and effective in primary care settings [14, 15]. This holistic approach addresses systemic challenges hindering timely and adequate mental health services.

Mental health assessment is crucial for effective treatment and support. Traditional methods, such as self-reporting through questionnaires and interviews, have limitations due to biases and inaccuracies [16–19]. To address these, objective measures like behavioral and physiological data have been introduced. Behavioral data, including observable actions like facial expressions and eye movements [20], provides real-time insights into an individual's mental state [21–26].

Physiological data collected through various sensing technologies, including accelerometers, gyroscopes, pupil corneal reflection, Electroencephalogram (EEG) [16, 27], Electrocardiogram (ECG), photoplethysmography (PPG), Galvanic Skin Response (GSR) [28], pressure-sensitive touch screen [29], heart rate, sound, temperature, oxygen saturation, and salivary cortisol levels [30], provide continuous monitoring. Among these physiological measures, eye tracking presents a unique opportunity to objectively assess mental health by observing subtle changes in eye movements and gaze patterns, which can serve as biomarkers for various mental health conditions.

Eye tracking is a sensor technology that detects gaze, providing insights into visual attention and subconscious processing. It has become a promising tool for monitoring human behavior and cognition [31]. It allows precise measurement and analysis of eye gaze movements, providing valuable insights into visual attention and cognitive processes. Its potential is particularly significant in mental health monitoring, as it can help understand cognitive processes [32].

Research indicates that individuals with mental health disorders often display unique eye movement patterns and gaze behaviors, such as reduced eye contact and altered visual attention [33]. This can be attributed to depression and anxiety disorders, where individuals may have a greater attentional bias towards threat-related stimuli. Eye tracking technology can help assess, diagnose, and monitor these conditions [34, 35].

Furthermore, eye tracking has the potential to overcome some of the limitations associated with traditional assessment methods. It collects data objectively and non-intrusively, eliminating the cognitive biases in self-reporting. Eye tracking can provide real-time and continuous monitoring, capturing minute changes in eye movements that may indicate the presence or progression of mental health conditions. It also enables early identification that facilitates early intervention, enabling healthcare professionals to provide timely support and prevent the worsening of symptoms [36].

The unique insights from eye movements and gaze patterns provide a new dimension to understanding mental health [37]. Through objective assessment and continuous monitoring, eye tracking contributes to accurate diagnoses, personalized treatment plans, and improved overall mental well-being [38]. Early detection through gaze behavior analysis enables timely interventions, revolutionizing mental health assessment with cutting-edge technologies and advancing us toward a future where mental health can be effectively monitored and managed.

Eye tracking provides various benefits; however, there are some drawbacks that need to be addressed. One significant drawback is the expensive cost of hardware, which can be replaced with cheaper webcam-based technology. The data obtained from an eye tracker is always raw and requires further processing to yield relevant information [39]. Developing supporting tools will be beneficial for data exploration. The difficulty in interpreting eye-tracking data, can be overcome by developing efficient software that provides clear insights. Additionally, eye-tracking data is idiosyncratic, varying from one participant to another, necessitating a clear and detailed study for accurate interpretation. The accuracy of eye-tracking systems can be affected by lighting conditions, user movement, and the need for calibration. Advancements in algorithms and user-friendly calibration processes can address these issues. Privacy and ethical concerns regarding the sensitive information revealed by eye-tracking data can be managed through robust data protection protocols and obtaining informed consent.

Mental state parameters encompass observable aspects that provide insights into an individual's cognitive and emotional well-being. These parameters, including attention, engagement, fatigue, stress, cognitive processes, social interaction, behavior, anticipation, and performance, collectively contribute to the intricate fabric of mental states [10, 11]. However, this thesis predominantly focuses on three pivotal mental state indicators: cognitive load, cognitive impairment, and emotional state. These collectively provide valuable insights into an individual's mental state and overall well-being. These specific dimensions serve as a supportive tool for healthcare professionals, offering in-depth insights into an individual's mental health. By observing and classifying these parameters, the goal is to assess and understand mental states, thereby facilitating the monitoring of individuals' mental health [10, 11].

Building upon the exploration of mental state parameters, this thesis delves into the technical intricacies of eye tracking, with a specific emphasis on classifying mental states. The study highlights the profound impact of these parameters on mental health. The overarching goal is to contribute to a comprehensive understanding of individuals' mental states by emphasizing the pivotal role of eye tracking technology in effectively detecting and understanding mental state indicators.

The following sections will delve into mental state parameters, specifically focusing on cognitive load, cognitive impairment, and emotional state.

1.1.1 Cognitive Load

Cognitive load refers to the amount of information that working memory can hold and process effectively. When the cognitive load increases, it can lead to mental fatigue and decreased cognitive performance, affecting tasks requiring concentration, decision-making, and memory [40, 41]. Eye tracking can effectively monitor cognitive load by tracking gaze patterns and fixation duration. Longer fixations on specific areas of a screen or frequent eye movements between multiple points may indicate high cognitive load [42]. By identifying the patterns indicating cognitive load, eye tracking can help individuals and healthcare professionals recognize when mental fatigue occurs, allowing prompt interventions like taking breaks or work adjustments. [43]. This monitoring can contribute to improved cognitive performance and overall well-being [44].

1.1.2 Cognitive Impairment

Cognitive impairment encompasses difficulties in memory, learning, concentration, and decision-making. It can affect mental flexibility, visual attention, and motor control. Eye tracking can assist in the early diagnosis of cognitive impairment by detecting changes in gaze behavior and visual attention patterns [45, 46]. Individuals with cognitive impairment may exhibit irregular or unfocused gaze patterns when performing tasks that require attention and memory. Early detection through eye tracking allows for timely intervention and personalized treatment plans, potentially slowing down the progression of cognitive decline and improving the quality of life for affected individuals [46].

1.1.3 Emotional State

Emotions are mental states brought about by neurophysiological changes, influencing thoughts, feelings, and behavioral responses. Negative emotions, such as stress and anxiety, can harm hormonal balance and weaken the immune system, impacting overall mental and physical health [47]. Eye tracking can play a crucial role in understanding and managing emotions. Tracking gaze patterns and pupil dilation can detect emotional states in response to visual stimuli or tasks [48]. For example, increased pupil size and shifts in gaze towards emotionally charged images or scenarios can indicate emotional reactions. Individuals and therapists can discover emotional triggers and build stress management and emotional regulation strategies with this data.

Cognitive load and cognitive impairment can reveal a person's mental condition and capacity to handle daily tasks and stress. Higher cognitive load or impairment may indicate mental health disorders like depression or anxiety [49]. Additionally, emotional state can indicate mental well-being, with heightened emotional arousal potentially suggesting distress or difficulties in emotional regulation.

The cognitive load, cognitive impairment, and emotional state are influenced by a variety of factors, including medical conditions, medication usage, lifestyle choices, and more. Therefore, seeking guidance from a healthcare professional is recommended for a precise evaluation of mental health. They can conduct a thorough assessment to comprehend and address these complex issues.

Acknowledging the shortage of mental health professionals and the need for timely mental illness detection, this thesis centers on predicting an individual's mental state through the analysis of cognitive load, cognitive impairment, and emotional state using eye gaze data. This focus aims to provide a support in healthcare professionals for detecting mental states and facilitating diagnoses within the individual's comfort zone. It supports the healthcare experts by providing them with valuable indicators of an individual's mental state through the comprehensive analysis of eye gaze data.

1.2 Motivation

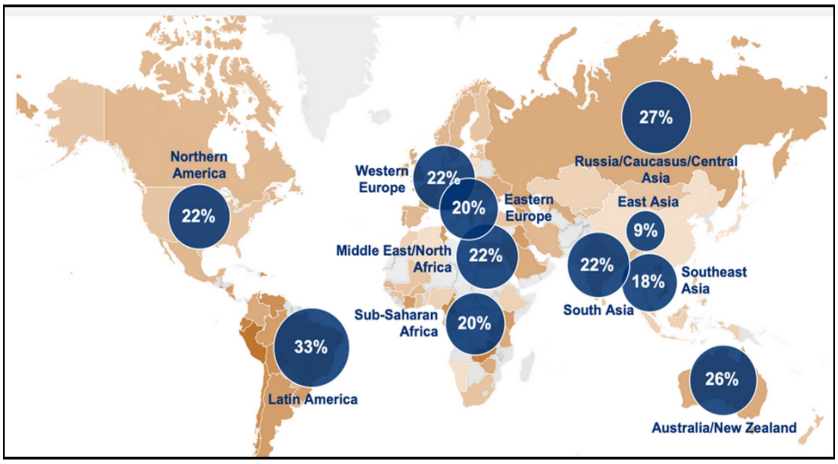


Figure 1.1: The survey on experience of anxiety or depression by Wellcome Global Monitor in 2020 [1]

The motivation behind mental health monitoring is deeply rooted in addressing the pressing mental health challenges faced by India and many other countries. Several key factors have highlighted the urgency for such a model.

Mental health disorders have emerged as the leading cause of non-fatal disease burden worldwide, emphasizing the critical need for effective mental health solutions [1], [50]. Figure 1.1 illustrates this trend. Unfortunately, societal stigma often leads to the concealment of mental health disorders, as individuals fear judgment or societal pressures, hindering their willingness to seek help. Based on the 'Mental Health in India' statistics provided by IPSOS Global Health Monitor 2019, Figure 1.2 illustrates the frequency of visiting or consulting with mental health professionals in India [2].

Compounding this issue is the significant concern of unavailability and inaccessibility of timely mental health care, particularly in regions like India where many individuals struggle to access the necessary services [51]. The National Mental Health Survey in 2016 revealed alarmingly low treatment rates, with a substantial percentage of individuals with mental illness in India receiving no treatment [51]. Further exacerbating the situation is the shortage of mental health professionals, as highlighted by a 2021 study by the Observer Research Foundation [52]. Figure 1.3 based on 'Major Healthcare Challenges' statistics provided by IPSOS Global Health Monitor 2021, 41% of people in India reported not having access to treatment or facing long waiting queues, while 39% mentioned insufficient staff as a significant issue [3]. Recognizing the importance of early intervention, it becomes imperative to address these challenges, emphasizing the need for timely diagnosis and intervention to reduce the burden of mental illness.

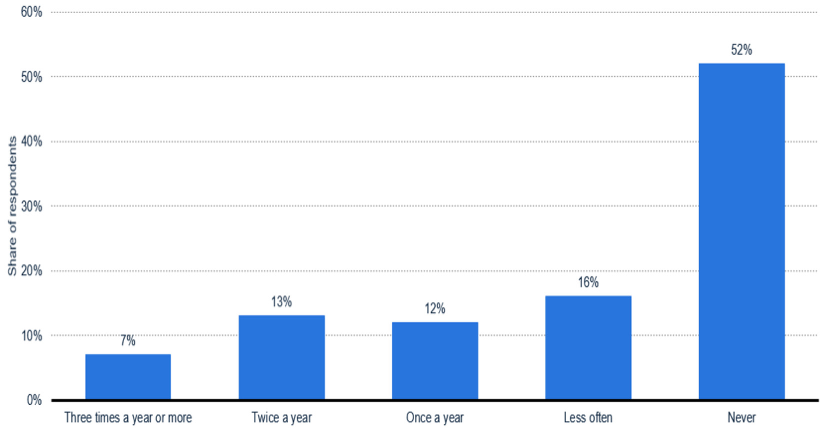


Figure 1.2: "Mental Health in India" statistic provided by IPSOS Global Health Monitor 2019. [2]

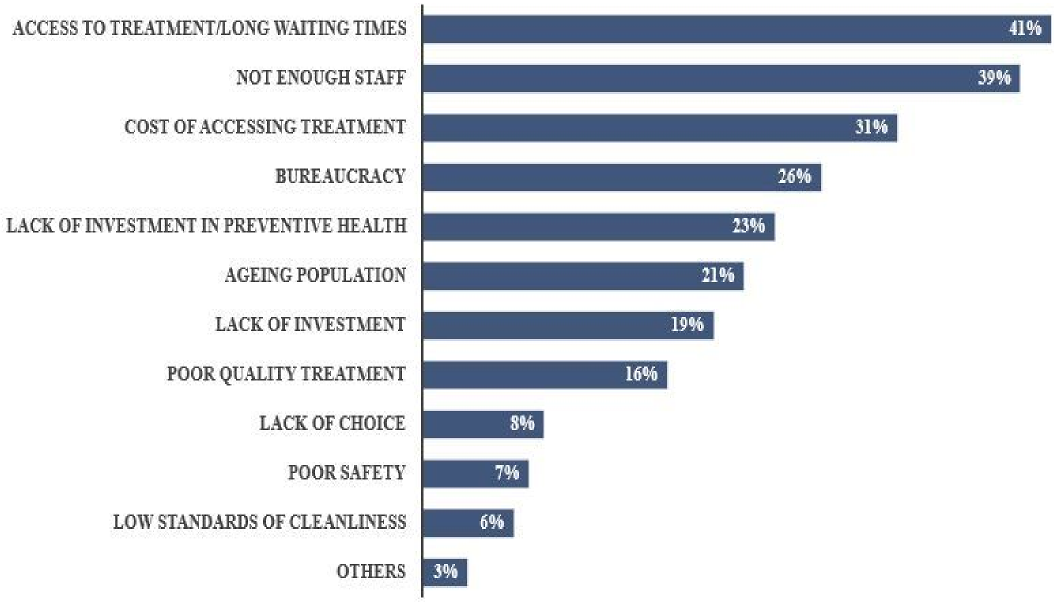


Figure 1.3: Major healthcare challenges, statistic provided by IPSOS Global Health Monitor 2021. [3]

1.3 Challenges

The development of a model for mental health assessment using eye gaze data faces several challenges. The lack of publicly available eye tracking datasets is a major obstacle, making it difficult to train and test machine learning (ML) and deep learning (DL) models. Obtaining labeled data for mental health indicators like cognitive load, cognitive impairment, and emotional state is also a challenge. This requires the expertise of mental health professionals and may be resource-intensive. Collaboration between researchers, mental health experts, and data annotators is needed to create datasets that can drive advancements in this critical domain.

1.4 Scope

The scope of the thesis encompasses several critical areas within the mental health assessment and intervention field.

1. **Dataset Creation and Expansion**: Since there is lack of multimodal datasets, the creation of multimodal datasets enables the exploration multiple physiological signals and that leads to the exploration of multiple insightful features and technological advancements.

2. **Mental Health Screening and Monitoring**: The scope of this research involves establishing a supportive tool that provides more insightful features to the mental health professionals. It offers clear analysis and providing in depth features and that can help the health professionals to save their time in monitoring each person.

3. **Advanced Cognitive Assessment**: This research presents the new approach to the enhanced cognitive test, focusing on eye tracking of individuals' abilities in the TMT. In addition to typical completion time measurements, the model analyzes a wide variety of gaze features in order to assess the subject's cognitive abilities as informative as possible.

4. **Adaptability and Integration**: The model has the adaptability of including various physiological data and can be incorporated in various domains like healthcare and education.

5. **Reallocation of Healthcare Resources**: It ability to provide more insightful features enables optimized reallocation of healthcare professionals to patients who need immediate care.

1.5 Thesis Objectives

The objective of the thesis is to develop an Intelligent system that can monitor the mental health of a person by assessing key mental state parameters, such as cognitive load, cognitive impairment, and emotional state. The system analyzes various physiological measures, offering valuable insights to support healthcare professionals.

The main objective of the study is divided into three sub-objectives, each focusing on specific aspects to achieve the overall research goal.

1. Generate Eye Gaze Dataset and develop Meaningful Eye Gaze Features:

(a) Identify various tests or stimuli used in the study of mental health monitoring.

   The study reviews existing literature to identify established mental health assessment tests and stimuli. Consulting with mental health professionals helps select appropriate assessments, encompassing subjective measures like self-report questionnaires and objective measures, including physiological and behavioral indicators.

(b) Establish an environment for data collection and identify appropriate stimuli.

   Establish a controlled environment for optimal data collection, encompassing quiet rooms and proper lighting. Ensuring participant privacy and confidentiality is a priority. Data collection accuracy and reliability are maintained by calibrating and validating measurement tools like physiological sensors and questionnaires.

(c) Extraction of features and creation of datasets.

   Develop algorithms that extract relevant features from eye gaze data, such as fixations, saccades, and blinks. The dataset creation includes administering selected mental health tests and stimuli to participants, followed by collecting responses and physiological data to construct comprehensive datasets.

2. Design an AI system to make Intelligent decisions to identify various mental health indicators.

   (a) Identify the eye gaze biomarkers, indicating a person's mental state parameters like cognitive load, cognitive impairment, and emotional state.

      Perform an extensive literature survey to pinpoint distinctive eye gaze patterns and biomarkers associated with cognitive load, cognitive impairment, and emotional state. This involves an in-depth exploration of existing research to discern specific ocular behaviors and physiological indicators that serve as reliable markers for these mental states.

   (b) Apply Computational Intelligence, Machine Learning and Deep Learning technologies that help the system to make Intelligent decisions on the classification of those identified mental state parameters and monitor the user's mental health. Implement decision-making algorithms that consider the mental state parameters, incorporating eye gaze biomarkers.

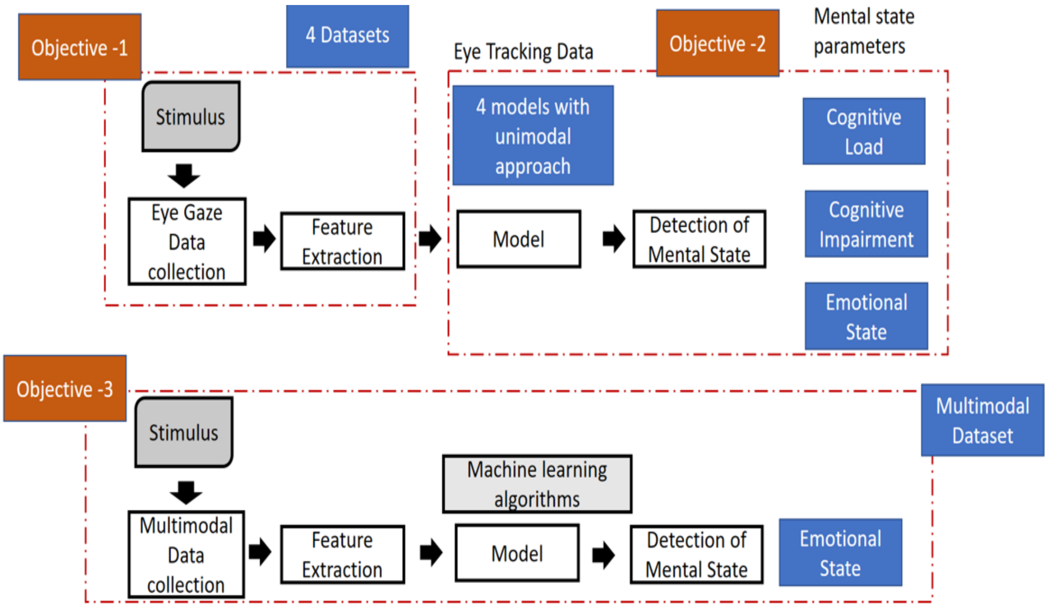


Figure 1.4: Major contributions

3. Create a multimodal dataset and implement emotional state detection models based on various physiological measures, such as ECG, GSR, PPG, Respiratory signals, and eye tracking data. The impact of integrating physiological signals as an additional component to the eye tracking-based emotional state detection model is evaluated, revealing potential enhancements or alterations in the system's performance.

1.6 Major Contributions

The significant contribution of the thesis, as shown in Figure 1.4, encompasses various aspects of mental health monitoring. Foremost, based on the first objective, it addresses the notable challenge of the scarcity of publicly available eye tracking datasets for mental health assessment. By creating four eye tracking-based datasets, this model bridges a critical gap in the field, offering researchers a valuable resource for further exploration and development in mental health assessment.

Four distinct unimodal datasets have been created for various research purposes, where unimodal refers to data collected solely through eye tracking, specifically eye gaze data. The dataset naming convention follows a systematic structure, starting with the data collection mode, followed by the stimulus used, and ending with the targeted mental health parameter. The first dataset, ET_MT_CL, is intended for investigating cognitive load based on eye tracking data and utilizes mathematical questions as stimuli [49]. The second, ET_TMT_CL, is utilized for studying cognitive load but employs the TMT as the stimulus. The third dataset, ET_TMT_CI, focuses on exploring cognitive impairment, utilizing the TMT as the experimental stimulus [14], [15]. Lastly, the fourth dataset, ET_Video_ES, has been curated for the examination of emotional states, utilizing videos as the stimuli [36].

In addition to dataset creation, the thesis focuses on the extraction of novel features, improving the dataset with distinctive markers for a comprehensive understanding of underlying patterns. These features are designed to detect mental state parameters, cognitive load, cognitive disability, and emotion. Their uniqueness and relevance make them powerful for enhancing the precision and efficacy of mental health assessment, providing a deeper understanding of an individual's mental state. The thesis highlights the significance of incorporating high-level features such as error rate, scanpath comparison score, total time, and inattentional blindness for generating comprehensive cognitive impairment scores

In addition to the creation of the new dataset, the thesis focuses on the discovery of new features, enhancement of the dataset with unique identifiers for the essential pattern recognition. These features are aimed to identify parameters of a subject's mental state, his or her cognitive load, cognitive impairment, and emotional state. Their uniqueness and relevance make them effective in advancing evaluation of mental disorders' precision and effectiveness, as well as offering deeper insight into the subject's psychological state. The thesis highlights the significance of high-level features including error rate, scanpath comparison score, and inattentional blindness for deriving cognitive impairment scores [15].

Based on the second objective, the thesis has contributed to developing four models for detecting mental state by monitoring cognitive load, cognitive impairment, and emotional state of a person. The first two models, ECL-1 and ECL-2 (Eye-Tracking Cognitive Load), employ statistical analysis and machine learning algorithms to discern a person's mental state by focusing on cognitive load. It incorporates key features such as pupil diameter and blink frequency to understand variations in cognitive load during mathematical assessments [49] and TMT tasks.

The third model, the ETMT (Eye tracking based Trail Making Test) model, make use of a fuzzy inference system and adaptive neuro-fuzzy inference system, plays a major role in the detection of mental states associated with cognitive impairment by providing detailed scores in visual search speed and focused attention, for understanding the exact cognitive deficits of a patient [15]. This application introduces an eye tracking version of the TMT, enabling the observation of specific features through the integration of fuzzy logic technologies [14]. The eye tracking version of this test significantly contributes to the understanding of a person's cognitive states, offering a non-invasive tool for mental health monitoring. The

ETMT model serves as a screening tool for cognitive impairments, allowing efficient allocation of healthcare resources to individuals with severe ailments.

The fourth model, PredictEYE, employs an approach by utilizing a deep learning time-series univariate regression model based on Long Short-Term Memory (LSTM) to predict future sequences of each feature, complemented by a machine learning-based Random Forest algorithm for mental state prediction [38]. This model predicts the mental state as calm or stressful based on a person's emotional state, introducing a personalized time series methodology that utilizes the benefit of time series analysis. PredictEYE's time series analysis plays a major role in identifying patterns and changes in data over time, enhancing predictive analytics and forecasting. By capturing the variations and interdependence in fluctuating data, this customized time series technique improves mental health assessments and monitoring.

Based on the third objective, a multimodal dataset labeled EmoRPhyE has been developed in addition to the unimodal datasets. This dataset is designed explicitly to research emotional states and incorporates pleasant and unpleasant images as stimuli. The multimodal data collected, including ECG, GSR, PPG, respiratory signal, and eye-tracking data, are utilized to assess emotional valence and arousal through physiological and eye-tracking-based approaches.

The thesis advances the field by underscoring the importance of individually analyzing cognitive load, cognitive impairment, and emotional state. While many previous studies have not explored the assessment of these critical factors, the developed models focus on monitoring mental states based on cognitive load, cognitive impairment, and emotional state separately.

Together, these contributions have shaped mental health monitoring by enabling more accurate, non-invasive examinations of mental health, cognitive performance, and emotional state. All the major contributions are summarized in the Table 1.1.

Table 1.1: Major contributions

Dataset	**Type of dataset**	**Model**	**Estimation**	**Publication**
ET_MT_CL	Unimodal	ECL-1	Cognitive Load	[49]
ET_TMT_CL	Unimodal	ECL-2	Cognitive Load	-
ET_TMT_CI	Unimodal	ETMT	Cognitive Impairment	[14, 15]
ET_Video_ES	Unimodal	PredictEYE	Emotional State	[36, 38]
EmoRPhyE	Multimodal	Model to detect emotional state	Emotional State	-

1.7 Thesis Organization

In Chapter 2, an extensive review of existing literature in the field of mental health monitoring is provided, focusing on the analysis of physiological data and exploring various computational techniques suitable for the proposed model. The chapter discusses both unimodal and multimodal approaches employed in mental health assessment, with a particular emphasis on understanding mental state parameters such as cognitive load, impairment, and emotional state. Chapter 3 delves into the creation of an eye gaze dataset, outlining protocols and rules for collecting eye tracking data and emphasizing its necessity for understanding an individual's mental health. Various datasets generated for mental health assessment purposes are elaborated upon. Chapter 4 concentrates on detecting cognitive load using eye gaze measures, detailing the selection of cognitive load-inducing tasks as stimuli and observing changes in eye gaze parameters for precise detection. Chapter 5 expands on understanding cognitive state, exploring how specific physiological data can provide insights into cognitive functioning. Chapter 6 introduces a personalized unimodal time series approach for predicting emotional states, detailing methodology, data processing techniques, and modeling approaches using eye gaze time series data. In Chapter 7, a multimodal dataset is created, incorporating physiological measures alongside eye tracking data to understand emotional states. The chapter focuses on assessing emotional valence and arousal through various physiological measures and eye-tracking-based models. Lastly, Chapter 8 serves as the conclusion, summarizing key findings, contributions, and implications, discussing study limitations, and suggesting areas for future research in mental health monitoring.

Chapter 2

Related Works

2.1 Introduction

Mental health monitoring involves assessing cognitive and emotional functioning. Eye tracking is significant in this area, providing reliable information for researchers, clinicians, and patients. Eye tracking technology is effective in capturing eye movement impairments which is effective in detecting age-related conditions and neurological disorders. This chapter explores eye tracking research in mental health monitoring and the role of mental state parameters. This chapter explains the role of eye tracking in mental health monitoring in section 2.2, traditional methods for evaluating mental states in section 2.3, and AI-based approaches for quantifying mental health indicators in section 2.4. Section 2.5 explains the role of mental state parameters in monitoring mental health. Section 2.6 to 2.8 explains each mental state parameter, cognitive load, cognitive impairment, and emotional arousal. Section 2.9 explains the need for other physiological measures.

2.2 Role of Eye Tracking in Mental Health Monitoring

Eye tracking technology is crucial for mental health monitoring, providing accurate and quantitative data to researchers, physicians, and patients to better understand and manage various illnesses [53]. The adaptability and effectiveness of this technology make it ideal for aging and age-related neurological and mental health disorders. It extends its capabilities in enabling early detection, providing objective biomarkers, prognosis, and real-time monitoring while delving into the intricate relationship between eye movements and cognitive function. The promise of its integration into healthcare and telehealth systems holds the potential to enhance the quality of life for the elderly population and alleviate the burden of age-related conditions on healthcare systems.

Eye tracking technology has offered a versatile and non-invasive approach to understand various neurological and psychological conditions [54]. Given its close connection to the central nervous system, it emerged as a sensitive tool in detecting disorders and diseases affecting critical regions such as the cerebral cortex, brainstem, and cerebellum. In the quest to better understand compromised areas of the brain, eye movement dysfunction, often detectable through eye tracking, provides vital insights.

Notably, research endeavors have extensively explored the impact of mental disorders on eye movements, resulting in a diverse spectrum of applications for analyzing eye movements in patients. These applications encompass a wide range of conditions, positioning eye tracking as a dependable marker for various brain-related diseases, including dementia [54].

In different domains of mental health monitoring, eye tracking has demonstrated its versatility and efficacy [55]. Detecting eye movement abnormalities including smooth pursuit dysfunction and inhibitory saccade deficiencies can help diagnose and assess Alzheimer's disease severity [56, 57]. In schizophrenia, smooth pursuit deficiencies and higher saccade frequencies provide crucial insights into the illness and its genetic markers. Eye tracking can detect multiple sclerosis early by assessing saccade peak velocity changes between both eyes, providing reliable disease progression monitoring. [57].

Eye tracking has become effective in monitoring mental health conditions and neurological disorders. Its non-invasiveness, early detection, and continuous monitoring make it useful for studying, diagnosing, and managing many illnesses. This technology contributes to advancing mental health research and developing more effective monitoring and intervention strategies.

Eye tracking research was conducted in various disciplines, including psychology, neuroscience, marketing, and education [49]. Researchers tried to understand how eye movements are related to visual attention, behavior, and cognition [15]. It has also been used to evaluate information-processing and decision-making skills and measure emotional and cognitive responses. Eye tracking has also been used to study cognitive development, plasticity, and visual processing.

Eye tracking is widely used in mental health research since it provides a window into cognitive processes that may be difficult to assess with traditional methods [15]. Eye tracking can determine emotional responses to stimuli, attentional focus, and signs of cognitive decline. It is also used to evaluate the efficacy of mental health treatments, such as psychotherapy and pharmacological interventions, and to develop more targeted interventions [14].

Cognitive load, cognitive impairment, and emotional state are very key features to mental health metrics. The monitoring of such measures significantly aids in the assessment of an individual's mental well-being. The eye tracking data analysis is the main aspect that the proposed model focuses on, as it gives a thorough outlook on such key metrics. Eye tracking technology has become a sensitive yet valid way to determine cognitive load, cognitive ability, and emotional state.

Cognitive load is a very important element in monitoring mental health. It can be easily measured by analyzing information obtained from eye tracking data. When experiencing high cognitive loads, eye movements include long fixation duration, increasing saccadic activity, and focusing on the relevant area for a task [43]. This becomes important in designing interfaces for users, educational systems, and work environments to optimize cognitive performance in a way that eliminates overload and maximizes productivity.

The changes in oculomotor movement patterns can assist in the early detection of cognitive impairment, particularly for conditions such as dementia and Alzheimer's disease. Inhibitory saccade dysfunctions and smooth pursuit impairments are the early markers of such diseases [58]. Regular eye-tracking to monitor disease progression and effectiveness of interventions will add up to deliver better patient care [45, 46].

With eye tracking technology, even the emotional state of the respondent can be analyzed. Gaze patterns during the reaction to emotionally provocative stimuli present important insights [59, 60]. People tend to focus their attention on emotionally salient places while having an elevated emotional state [47]. This information may be used to evaluate the emotional impact of different stimuli and material; it proves helpful for user experience design, marketing, and psychological research.

Eye tracking technology, combined with advanced computational techniques, can predict mental states by analyzing eye movement features related to cognitive load and emotional state. This capability has applications in personalized mental health interventions, human-computer interaction, and assistive technologies [38, 49]. Eye tracking provides a non-invasive, objective, and real-time means to assess cognitive load, impairment, and emotional state, offering significant insights for improving mental health research, monitoring, and intervention strategies. The following sections discuss the workings of an eye tracker, applications of eye tracking technology, and different types of eye trackers [48, 61, 62].

2.2.1 Working of an Eye Tracker

Eye tracking technology offers deep insights into how individuals control their gaze and perceive the world. The pupil center corneal reflection method is effective

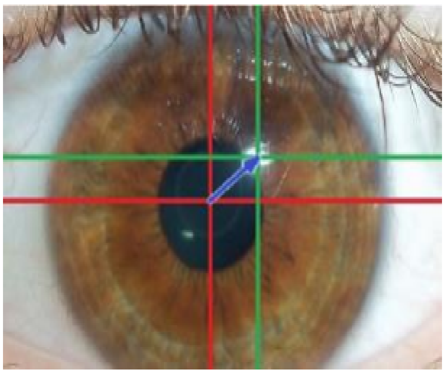


Figure 2.1: Insights into eye tracking functionality

among the various available techniques [63,64]. The pupil center corneal reflection method provides real-time accuracy and stability, addressing common issues of low accuracy and restricted head movement in existing systems [65]. Enhancements in the method, such as those focusing on multiple corneal reflections and the use of visible light, further improve its accuracy and robustness [63, 64]. These advancements make this method a promising approach for gaze tracking in various applications.

An eye tracker is a comprehensive hardware device designed to capture and analyze eye movements, determining a user's precise gaze point. It includes cameras, illuminators, and software that work together seamlessly [66]. Illuminators project near-infrared light onto the eyes, while high-resolution cameras capture detailed images of the eyes and reflected patterns. These images are then analyzed using complex algorithms to identify the pupil's boundaries and accurately determine its center, essential for pinpointing the gaze location [67].

The eye tracker pinpoints the reflections of infrared illuminators on the corneal surface, appearing as bright spots near the pupil's center, as shown in Figure 2.1 . With this precise knowledge, the system uses geometric computations and advanced algorithms to determine the user's gaze direction in real-time [55]. This generates a continuous stream of gaze data for studying visual behaviors, cognitive processes, and human-computer interaction. The Pupil Center Corneal Reflection method involves illuminating the eye, capturing and analyzing images, and identifying the pupil center and corneal reflections. This technology is crucial for cognitive research and user interface design.

2.2.2 Applications of Eye Tracking Technology

Eye tracking technology plays a crucial role across various fields, offering diverse applications. Human-Computer Interaction (HCI) facilitates hands-free control for individuals with physical disabilities [68]. Usability testing benefits from eye tracking in optimizing digital interfaces and enhancing user experience [69]. Market research and advertising leverage eye tracking to analyze consumer behavior and improve marketing strategies [70]. In medical and psychological research, eye trackers aid in diagnosing conditions by studying eye movement patterns [38]. Educational settings use eye tracking to understand reading and comprehension patterns among students. The automotive industry employs eye tracking for driver monitoring, enhancing road safety. Virtual and Augmented Reality (VR/AR) applications integrate eye tracking for more immersive experiences [71]. In the gaming industry, eye tracking enhances gameplay by responding to users' gaze and intentions [72]. Overall, the versatility of eye tracking technology makes it invaluable in technology, healthcare, marketing, and research, providing insights into human behavior and cognition.

2.2.3 Different Types of Eye Trackers

1. **Remote Eye Trackers:** Remote eye trackers, selected for their non-intrusive design utilizing infrared light, monitor gaze without direct eye contact. This ensures a natural user experience in mental state classification, enabling the collection of genuine data on cognitive processes and emotional states without interference [73].

2. **Head-Mounted Eye Trackers:** Head-mounted eye trackers can be worn by users and it move with head motions, offering a first-person view of visual attention. Particularly valuable in scenarios with dynamic head movements, they provide a comprehensive perspective, contributing valuable data for applications such as virtual reality. This technology enhances understanding by capturing nuanced interactions in dynamic environments [74].

3. **Wearable Eye Trackers:** Portable eye trackers like glasses or wearable offer adaptability for data collection in diverse environments. Since it is able to capture real behaviours in daily life conditions, it is appropriate for real-world studies, especially in mental state classification [75].

4. **Mobile Eye Trackers:**

   Mobile eye trackers, intended for use in conditions of movement, can be installed on a car or a bicycle. They are particularly useful in investigating visual attention in real-world scenarios and offer important information about people's behavior in their environment. Especially beneficial for the mental state classification in the mobile environment, they provide a wide view of behavior in different situations [76].

Eye tracking is essential for classification of mental state parameters and offers various solutions for specific tasks. In this thesis, remote eye trackers has been used since it is non invasive in nature. Eye tracking systems that can be operated remotely using infrared light allow for gaze observation without physical contact. It is important to avoid interfering with the user's action since subjects should display real behaviors.

2.3 Traditional Methods for Evaluating Mental States

Traditional mental state assessments include clinical interviews, psychometric tests, and observations. Clinical interviews are direct or remote interactions between mental health professionals and individuals to assess cognitive capacity, emotional condition, and behavior. This method reveals a person's mental state [77].

Psychometric examinations, a well-established approach, utilize standardized tests such as the Beck Depression Inventory and the Mini-Mental State Examination [35] to measure various elements of mental well-being. These exams provide quantifiable data, which helps in the examination and comparison of individuals' mental states, hence improving diagnostic abilities.

Observation, the third conventional method, records an individual's behavior, expressions, and social interactions. It gives behavior facts to help understand a person's psychological condition. It provides external validation of self-reported data and mental health indicators.

These conventional methodologies allow mental health professionals to assess and track many elements of a person's mental health using personal views, measurable data, and observational perspectives.

Relying completely on traditional methods may not provide a detailed understanding of all mental state parameters. The exploration integrates innovative technologies, leveraging advanced algorithms to enhance the depth of insights. These computational techniques aim to extract more refined inferences, surpassing the limitations of conventional approaches.

2.4 AI-Based Approaches for Quantifying Mental Health Metrics

Eye-tracking technology provides a rich resource of data about human behavior and mental states. This data includes eye movements, gaze points, fixations, and saccades, collected when a person interacts with stimuli like pictures, videos, or interfaces, resulting in large datasets.

Before implementing machine learning (ML) and deep learning (DL) techniques, meaningful features must be derived from the raw eye-tracking data. These features typically include fundamental eye movement parameters such as fixation duration, saccade amplitude, and pupillary response, which help in understanding the user's visual behavior. All the extracted features may not be equally important for characterizing the user's mental state or behavior. Feature selection techniques identify the relevant ones while reducing dimensionality and computational complexity. Pre-processing includes steps for noise reduction, data cleaning, and normalization to ensure quality and consistency. It may also involve co-registering eye-tracking data with other relevant sources, like physiological sensors or self-report measures of emotion.

ML and DL algorithms process the eye-tracking data. ML algorithms classify or predict data points, such as gaze patterns corresponding to different emotional or cognitive states. Unsupervised ML techniques, like clustering, group similar eye movement patterns among respondents. Deep learning models, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), analyze temporal sequences within eye-tracking data to track changes in mental states. Trained models infer the user's mental state based on eye movement patterns.

For example, increased stress or heightened attention can be detected based on gaze behavior. These analyses yield mental health monitoring parameters like stress levels, cognitive workload, and emotional engagement, providing insights into the user's psychological well-being. Intelligent systems can use these parameters for real-time feedback or interventions.

Combining eye-tracking technology with computational methods, mainly ML and DL algorithms, leads to inferring patterns from eye movement data. These patterns relate to human cognition, emotion, and mental health, resulting in intelligent systems that support psychological well-being by monitoring and responding to the user's mental state. The next section covers statistical models, fuzzy models, ML, and DL models.

2.4.1 Statistical Models

Correlation analyses, linear regression models, t-tests, and analysis of variance (ANOVA) are commonly used statistical methods in mental health monitoring research. Traditional statistical techniques imply linearity between variables and may be biased by outliers and non-normal data. To address these issues, researchers have introduced robust statistical methods like robust regression, trimmed means, and bootstrapping [78]. These methods provide more precise and reliable results, enhancing the validity and reliability of the analyses. Multiple regression and correlation analysis effectively help researchers analyze complex relationships among multiple variables [79]. Neural networks and decision trees, among machine learning models, better handle non-linear relationships and outliers.

Time series data analysis methods also incorporate statistical methodologies. Time series models can be linear or nonlinear; classic linear types include autoregressive (AR), moving average (MA), auto-regressive moving average (ARMA), and auto-regressive integrated moving average (ARIMA) [80]. Autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis help identify appropriate models for time series data by showing how data sequences are related. Non-linear models like autoregressive conditional heteroskedasticity, generalized ARCH, exponential GARCH, threshold autoregressive, and nonlinear autoregressive are also available for better analysis and prediction.

2.4.2 Fuzzy Models

Fuzzy models are crucial for assessing mental states, including cognitive load, impairment, and emotional state, by addressing complexities in understanding and monitoring these states. They handle ambiguity and uncertainty by using linguistic terms and allowing gradual transitions between states, providing realistic and nuanced assessments. These models can account for various factors affecting mental states, enhancing accuracy. They also accommodate individual differences in responses to cognitive tasks and emotional stimuli, adjusting membership functions based on personal experiences and preferences.

The adaptability of fuzzy systems allows them to respond dynamically to changes in mental states, making them suitable for real-time monitoring and interventions. Their transparency and interpretability are vital in clinical and human-computer interaction contexts, aiding in trust-building and effective interventions. Fuzzy models are effective in healthcare, driver monitoring, education, and human-computer interaction, offering insights for improving safety, well-being, and performance. They can be integrated with machine learning or neural networks to enhance predictive accuracy and robustness, combining fuzzy logic's interpretive strengths with machine learning's pattern recognition abilities.

In CAD interfaces, fuzzy logic models map psychophysiological signals to key emotions, improving user interfaces and engineering decision-making [81]. This non-invasive approach provides valuable insights into CAD tasks' affective and cognitive dimensions, leading to more intuitive design processes. A driver fatigue detection system uses non-intrusive video analysis [82], evaluating alertness through visual cues like eye closure duration and yawning. This system employs a Fuzzy Expert System (FES) for real-time driver state classification, optimizing feature extraction and response time. The study also developed spatial fuzzy c-means for mouth detection, enhancing system robustness, contributing to advanced driver vigilance and fatigue monitoring.

Fuzzy models effectively monitor mental states due to their ability to address the complexities, uncertainties, and subjectivity inherent in human cognition and emotion. Their adaptability, interpretability, and established applicability make them invaluable in various domains where understanding and responding to human mental states is crucial.

2.4.3 Machine Learning and Deep Learning Models

Artificial Intelligence(AI) based techniques and statistical analysis applied to the physiological data make better predictions and make an efficient health monitoring system [83]. The enormous data from the physiological sensors can be labeled and used for training. The machine learning or deep learning model trained using physiological data can detect the mental state of unseen observations. Classification methods such as decision trees, Naive Bayes, etc., can find patterns automatically and are widely used to detect mental conditions [84].

Machine learning and deep learning techniques applied to physiological data have shown potential in monitoring mental health and assessing human mental states [83]. Machine learning models can be trained on labeled eye-tracking data to classify the user's mental state [85,86]. However, more investigation is required to develop the ideal eye-tracking paradigms and machine-learning algorithms for correctly diagnosing people with Autism Spectrum Disorder (ASD) and other neurological and neuropsychiatric illnesses.

Machine learning based on eye-tracking data could classify individuals with ASD and typically developing (TD) individuals with an 81% pooled accuracy [87]. A study on young children indicated that with an accuracy of 85.1 %, fixation times at the lips and body could significantly distinguish ASD from TD [88]. Video-based eye-tracking method for measuring brain function using readily available webcams has the potential for early detection, diagnosis, and remote/serial monitoring of neurological and neuropsychiatric disorders [89].

Machine learning models, including Decision Trees, Naive Bayes, Support Vector Machines, and Random Forests, can analyze vast amounts of data and detect patterns [84] that may not be visible to human observers, providing an efficient health monitoring system. The studies showed that machine learning algorithms could analyze the data from these sources to accurately classify and predict psychological conditions, with classification accuracy ranging from 66% to 90% depending on the dataset and features used.

Recent technological advancements have enabled deep learning algorithms to automatically detect mental states, using various physiological signals such as visual metrics, EEG, and eye-tracking [19, 83]. One study used CNN-LSTM algorithms to analyze visual metrics time-series data and accurately classified individuals' mental health metrics levels with high accuracy, highlighting the potential benefits of home-based mental health monitoring for patients after oncologic surgery [19]. The other study proposed the EYE-CNN-DLSTM algorithm for psychological testing based on eye movement tracking data, which uses a fusion strategy combining CNN and DLSTM to evaluate patients with mental disorders [90].

A recent method combines EEG and eye-tracking signals in deep learning to boost emotion recognition accuracy [91]. This study highlighted the advantages of using physiological signals and the crucial role of eye-tracking in improving emotion recognition. The method employs a fusion model with a Gaussian mixed model, signal filters, feature extraction techniques, and normalization methods for precise emotion classification.

Algorithms like LSTM can be optimized through hyperparameter tuning to better predict trends and fluctuations [92]. Deep learning models also support multivariate forecasting by mapping inputs and outputs of linear and nonlinear models [93]. However, these models often lack personalization, meaning they are trained on general populations and may not account for individual differences in behavior, preferences, and emotional responses.

Personalized time series models can address this by incorporating individual-level data, tailoring the model to each person's unique characteristics, and improving mental health monitoring accuracy and effectiveness, especially for psychiatric disorders. Additionally, deep learning models require large datasets, which can be challenging in mental health monitoring due to limited data collection. Personalized models can work with smaller datasets by incorporating prior knowledge and individual characteristics, leading to more efficient and cost-effective mental health monitoring [94].

Time series is a machine learning technique that predicts target values only based on a known history of target values. It is a type of regression known as auto-regressive modeling in the literature. The main goal of time series modeling is to meticulously gather and analyze past information from the time series in order to develop a model that accurately captures the series' structure. It forecast future values by understanding the past. There are linear and nonlinear traditional models for time series data analysis. Two widely used linear models are Auto-regressive (AR) and Moving Average(MA). Auto-regressive Moving Average(ARMA) and Auto-regressive Integrated Moving Average(ARIMA) models combine AR and MA models. Autocorrelation function(ACF) and partial autocorrelation function(PACF) analysis can determine a proper model for the time series data. The ACF and PACF plots help forecast the time series data. These statistical measures show how the data sequence in the time series is related. Furthermore, nonlinear time series models are used for better analysis and prediction. Some of the non-linear models are Autoregressive Conditional Heteroskedasticity(ARCH), Generalized ARCH(GARCH), Exponential GARCH(EGARCH), Threshold Autoregressive model(TAR), Non-linear Autoregressive model(NAR).

The prediction of the impact of COVID-19 using stacked LSTM, Bi-directional LSTM, and convolution LSTM was proposed in [95]. Deep learning models can understand trends, seasons, and fluctuations. LSTM is one of the most widely used deep learning algorithms, which makes better predictions by hyperparameter tuning [92]. In addition, deep learning models can be trained to learn and understand the mapping between inputs and outputs of linear and nonlinear models. While most traditional methods are linear, deep learning models can promptly learn the linear and nonlinear relationships and support multivariate forecasting [93].

Time series analysis helps us comprehend the root causes of trends or systemic patterns over time. When data is examined at regular intervals, time series forecasting calculates the likelihood of future events. Time series data can be subjected to predictive analytics to identify anticipated data changes, such as seasonality or cyclical behavior, which helps to forecast and provides a deeper knowledge of data elements. According to the research on related works, the time series model will be a more effective method for a customized model that can track a person continuously. Considering the proposed personalized model, the time series model helps to understand the changes in a person's mental state based on the observed trend of the eye tracking measures and helps forecast future events.

2.5 Role of Mental State Parameters in Monitoring Mental Health

In mental health monitoring, multiple indicators provide an in-depth understanding of an individual's mental health. These dimensions, ranging from attention and engagement to fatigue, stress, cognitive processes, social interaction, behavior, anticipation, and performance, collectively contribute to the complex tapestry of mental states. Each parameter plays a significant role in unraveling the intricacies of an individual's cognitive and emotional functioning.

However, cognitive load, cognitive impairment, and emotional state have specific importance among these parameters. Cognitive load reflects the mental effort expended in processing information [40], while cognitive impairment signifies deficits in cognitive functions. Emotional state encapsulates the broader spectrum of an individual's overall emotional well-being. Understanding these parameters is pivotal in comprehensively assessing mental health, guiding interventions, and identifying potential challenges. The following section explains the mental state parameters, cognitive load, cognitive impairment, and emotional arousal. The conventional methods, eye tracking techniques, and computational techniques for detecting each mental state parameter and the associated disease are also explained in each of the following sections.

2.6 Cognitive Load

Cognitive load, defined as the cumulative mental effort held within a person's working memory at any given moment, is a crucial factor that impacts task performance [41]. It is noteworthy, however, that cognitive load is not a uniform experience; it varies among individuals based on age and gender. Our working memory has a finite capacity, and it becomes overloaded when handling complex information, often resulting in problematic or confusing learning experiences [44].

High cognitive load significantly contributes to reduced performance, increased stress, and errors, particularly in industrial settings where individuals work continuously for extended periods [42]. This elevated mental workload can lead to mental fatigue and stress, which, in turn, can adversely affect mental health and overall performance [96]. Timely identification of cognitive load levels can help individuals manage and mitigate mental fatigue without compromising their well-being.

2.6.1 Conventional Methods for the Detection of Cognitive Load

Various standard questionnaires, such as the NASA Task Load Index (NASA-TLX) [97] and Subjective Workload Assessment Technique (SWAT) [98], are commonly employed to subjectively assess cognitive load. These tools capture individuals' perceptions of mental and physical demand, effort, and workload-related dimensions. Researchers and practitioners choose specific questionnaires based on contextual demands, gaining valuable insights into subjective experiences. In subjective measurement, users rate their mental effort using tools like the NASA-TLX questionnaire [97]. These subjective assessments and objective measurements offer a comprehensive understanding of cognitive experiences and workload demands in diverse settings, informing strategies for optimizing task performance and mental well-being. Performance-based approaches involve monitoring users' achievements, such as task completion times and accuracy rates. Although both methods are convenient, they are unsuitable for real-time cognitive load assessment, as evaluation can only occur after task completion.

2.6.2 Eye Tracking Techniques for the Detection of Cognitive Load

Eye tracking techniques have proven valuable in assessing cognitive load. A model was presented describing the relationship between eye movements and cognitive load [43]. This model highlights correlations between major eye movements and cognitive load, making eye tracking an effective means of measurement. Longer fixation duration and reduced fixation rate indicate increased task complexity, while higher saccade velocity and longer saccades suggest greater cognitive load. Pupil dilation and blink rate reduction during cognitive load increase further indicate this relationship.

Increased cognitive activity can lead to mental fatigue, affecting behavioral performance. A novel method was introduced for detecting mental fatigue based on cognitive load increase [99].The study found that engaging in cognitive tasks can induce mental fatigue by examining features like saccade amplitude, duration, rate, inter-saccade interval, mean saccadic velocity, fixation duration, blink duration, blink rate, inter-blink interval, pupil diameter, constriction velocity, and amplitude of each eye before and after the tasks. Eye-tracking technology, due to its non-intrusive nature, is pivotal in monitoring mental health. Pupillary responses and other eye measures reliably indicate cognitive load, allowing parameter tracking without additional devices.

Eye-tracking for cognitive load detection involves diverse stimuli: military aviation simulators [40], gaming activities [96], arithmetic problem-solving [96], coding tasks on whiteboards and paper [42], and varying levels of mental calculations [41]. These stimuli provide comprehensive scenarios to assess cognitive load through eye-tracking measurements, offering insights into the cognitive demands of individuals across different tasks and settings.

2.6.3 Computational Techniques for the Detection of Cognitive Load

Computational techniques and machine learning methods enable the extraction of meaningful patterns from behavioral and physiological data [100]. Various studies employ statistical and machine learning techniques, including one-way ANOVA [40, 96], t-test [40], ANCOVA [41], and machine learning models like Naive Bayes, Random Forest, Multi-Layer Perceptron, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Logistic Regression (LR), and Decision Tree (DT) [42]. Researchers used a vision-based approach to extract parameters from a driver's eye movements, combining manual feature extraction with deep learning for cognitive load detection. Results showed high classification accuracies, with SVM at 92% and CNN at 91%. This non-contact technology promises advanced driver-assistive systems, effectively detecting and classifying a driver's cognitive load [101].

2.6.4 Cognitive Load and the Associated Diseases

Cognitive load is intricately linked with various conditions, and eye tracking is crucial for its detection. For instance, Attention Deficit Hyperactivity Disorder (ADHD) involves challenges in sustaining attention and managing cognitive load. Eye tracking studies reveal distinct gaze patterns and attentional focus, offering an objective method to assess cognitive load [102].

Autism Spectrum Disorder (ASD) presents unique cognitive load challenges, especially in social settings. Eye tracking studies of gaze patterns help deepen our understanding of cognitive processing and cognitive load in individuals with ASD [88].

Stroke and Traumatic Brain Injury (TBI) recovery involves eye tracking measures, including gaze patterns and visual scanning, facilitating cognitive load assessments and enhancing the understanding of post-injury cognitive changes [103].

Schizophrenia disrupts cognitive load dynamics, impacting various cognitive functions. Eye tracking studies uncover alterations in gaze behavior and visual processing, offering valuable insights into cognitive load during tasks [104].

As individuals age, cognitive load management becomes intricate. Eye tracking becomes a valuable tool to study age-related changes in visual attention, working memory, and cognitive load, enriching our comprehension of cognitive decline in the elderly. Eye tracking is an indispensable instrument for detecting and understanding cognitive load in various conditions.

2.7 Cognitive Impairment

Cognitive impairment, marked by challenges in memory, comprehension, attention, and judgment, significantly affects daily life [14]. While costly and incurable, timely diagnosis and care can slow its progression [105, 106]. The growing prevalence, especially among the aging population, raises significant public health concerns [107]. Risk factors include family history, injuries, substance exposure, and education level [108]. Conditions like mild cognitive impairment (MCI) result from various factors and range from subtle abnormalities to severe deficits, impacting mental functions and independence. Early detection is crucial for optimal outcomes.

2.7.1 Conventional Methods for the Detection of Cognitive Impairment

Various approaches are employed to assess cognitive impairment, often involving neuropsychological tests administered by trained healthcare professionals. Widely used tests such as the Mini-Mental State Examination (MMSE) [35], the TMT [109], Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog), and the Frontal Assessment Battery (FAB) [106] are recognized for their validity but may not be brief enough for routine dementia screening. Self-reporting questionnaires, while well-validated, can be susceptible to biases. These traditional assessments, though accurate, have drawbacks. Older patients may require more time, and the evaluator's proficiency can impact outcomes, necessitating administration by professional neuropsychologists. Some tests involving writing and drawing tasks may pose challenges for those with motor dysfunction, affecting results [110, 111].

The TMT is a widely-used neuropsychological tool assessing visual attention and task-switching capabilities, offering insights into executive functioning, processing speed, focused attention, and mental flexibility [112, 113]. Particularly effective in detecting brain dysfunction, especially frontal lobe impairments, the TMT was historically employed to evaluate soldiers' brain damage during World

War II. Administered conventionally with paper and pencil, the test considers error rates and completion time, but limitations include a lack of detailed analysis [114]. The TMT aims for prompt and precise examinations to identify potential cognitive impairment indicators, necessitating psychologist assistance. Motor-impaired individuals may face challenges and prolonged completion times. Ongoing studies explore various TMT variants to overcome conventional method limitations [115].

The digital TMT (dTMT) precisely monitors a variety of distinct elements along with the overall completion time and the number of errors, such as the number of viewing pauses, the duration of each pause, lifts, lift duration, duration within the circle, and the amount of time across the circles [115].

The benefits of combining infrared eye tracking with the TMT task is another significant factor [116]. Research on infrared eye tracking is becoming well-known for its ability to diagnose cognitive impairment. Using eye tracking in conjunction with the TMT task was relatively unexplored. This gap in exploration presents an opportunity for further investigation into the potential benefits of employing eye tracking technology to enhance the TMT task as a tool for cognitive assessment.

2.7.2 Eye Tracking Techniques for the Detection of Cognitive Impairment

Eye tracking data is crucial for capturing involuntary physiological responses and revealing insights into cognitive inhibitions [49]. In contrast to clinical assessments, eye tracking measures prove effective in identifying early-stage cognitive inhibitions [45, 46]. These measures can distinguish subtypes of mild cognitive impairment, particularly through error rates in the antisaccade task [46]. Studies involving participants aged 55 to 90 showed significant differences in error rates among various mild cognitive impairment types, determined through Analysis of Variance (ANOVA) tests.

Eye tracking scores correlate with the Mini-Mental State Examination (MMSE) score, effectively differentiating control groups from those with Alzheimer's disease (AD) or Mild Cognitive Impairment (MCI) in memory tasks [35]. Gaze tracking has gained widespread acceptance for monitoring cognitive functions and neurological disorders [117].

Eye tracking technology is beneficial for assessing conditions like amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), Parkinson's disease (PD), multiple sclerosis (MS), and epilepsy [56, 57]. Tasks such as the antisaccade, smooth pursuit, and visual scanning tasks, observed through eye tracking, can detect early cognitive impairments. Difficulties or errors in these tasks provide insights into changes in inhibitory control, motor control, attention, and scanning abilities, enabling early detection of cognitive decline before conventional evaluations [46].

Combining infrared eye tracking with the TMT task proves significant [116]. Infrared eye tracking, known for diagnosing cognitive impairment, gathers objective, quantitative data on visual, attentional, and memory functions in ongoing studies [118]. Alongside basic completion-time metrics, eye tracking offers insightful and sensitive measurements.

In the detection of cognitive impairment using eye tracking, various tasks are employed to assess cognitive functions. These tasks include memory tests [106], attention assessments [106], calculation tasks, and the Trail Making Test (TMT) [115]. Memory tasks involve presenting words or images and tracking eye movements to analyze memory-related processes [119]. Attention tasks measure gaze patterns to evaluate focus and attention. Calculation tasks assess numerical processing, with eye tracking providing insights into visual attention allocation. The Word Memory Test assesses recall or recognition through eye movements. The TMT helps evaluate cognitive flexibility and executive function. Together, these tasks measure different aspects of cognitive performance. Integrated with eye tracking, they offer comprehensive insights into visual attention and processing, contributing to a detailed understanding of cognitive impairment.

2.7.3 Computational Techniques for the Detection of Cognitive Impairment

Statistical analyses, including ANOVA and Chi-square tests, were conducted to assess the correlation between eye tracking-based cognitive scores and those derived from neuropsychological tests [106]. The results showed a strong correlation, validating the effectiveness of eye tracking in measuring cognitive performance. Additionally, logistic regression was used to classify participants into groups with and without cognitive impairment based on correlations between paper and digital variants of the TMT [115,119]. Machine learning algorithms, such as Naive Bayes, SVM, and logistic regression, were also used to successfully distinguish between participants with and without cognitive impairment [120]. These findings highlight the utility of eye tracking and advanced statistical methods in accurately assessing cognitive function and detecting cognitive impairment.

2.7.4 Cognitive Impairment and the Associated Diseases

Amyotrophic lateral sclerosis (ALS) patients, facing both motor and cognitive deficits, exhibit higher error rates and longer saccadic latency. Traditional paper-pencil assessments become unsuitable as the disease progresses [57]. The Edinburgh Cognitive and Behavioral ALS Screen (ECAS) offers a standardized assessment, but it lacks sensitivity to early-stage cognitive changes [121, 122]. An eye tracking version of the ECAS test could streamline evaluations for ALS patients [123, 124].

In Alzheimer's disease (AD), eye tracking provides insights into progressive memory loss. AD patients display slower fixation, reduced fixation spans, and less precise saccadic movements [57]. Gradual declines in attention and visual attention impact performance in memory tasks, deductive reasoning, and working memory assessments [35, 106, 125, 126].

Parkinson's disease (PD) is marked by cognitive decline and ocular abnormalities. PD patients show prolonged response times during saccadic tasks and impaired focused attention [57, 127]. Standard assessments may lack sensitivity in early PD stages, while eye tracking measures correlate with disease severity [128].

Table 2.1: Cognitive impairments and associated factors analysis.

Paper	Disease	Cognitive impairments	Stimulus	Standard cognitive assessment tools	Shortcomings of standard assessment tools	Eye tracking measures	Observations
[57]	ALS	Motor impairment	Visual paired comparison task	ECAS	Cannot handle lower motor neuron atropy	Anti-saccade error rate, saccadic latency	Higher anti-saccade, error rate, saccadic latency
[57] [125] [126] [106] [35] [58]	AD	Memory impairment impaired visual attention, Attentional disengagement	Working memory tasks, deductive reasoning, memory recall task, visual memory task	ADAS-Cog, MMSE, MOCA	Longer time duration, not simple, subjects may feel highly stressed	Saccade, fixation, smooth pursuit	Longer time to fixate the target, shorter fixation duration, imprecise saccadic movements
[57] [129]	PD	Focused attention impairment, movement impairment memory impairment	Saccadic task, TMT	SCOPA-COG PD-CRS, MOCA	Low sensitivity to recognize cognitive deficits in the early stages of PD	Pupil diameter	Ocular abnormalities, longer response time
[130] [131] [132]	MCI	Memory impairment	Visual paired comparison task, Animal Fluency, WLM, Constructional Praxis, TMT, Digit Span subtest, Clock Drawing Test	MMSE, MOCA	Expensive, invasive, can't detect early stages of the disease	Fixation, saccade, re-fixation, pupil diameter, Total looking time, fixation count, percentage looking time on novel image	Percentage time in viewing the novel pictures could differentiate control group from MCI group.
[57]	MS	Impairment of attention, executive function impairment, memory impairment	Saccadic task, ocular working memory task	MRI MACFIMS	A trained evaluator needs at least 90 min for a full evaluation	Fixation, saccade latency	Fixation instability, higher saccade error rates, impaired pursuit.
[57]	Epilepsy	Neuropsychological impairment	Vision-guided saccade, antisaccade response inhibition, prosaccade, antisaccade tasks	ET PNS	Limited sensitivity, unsuitability for repeated assessment, sole focus on one aspect of cognition	Saccade, fixation	Increased error rate, longer reflexive time at the initial of saccade

MACFIMS − Minimal Assessment of Cognitive Function in Multiple Sclerosis, MoCA − Montreal Cognitive Assessment, ET − Epitrack, PNS − Portland Neurotoxicity Scale, WLM − Word List Memory

Cognitive impairment in older individuals involves disengagement of visual attention, particularly during saccadic tasks [58]. Visual paired comparison tasks utilizing eye movement analysis effectively distinguish normal controls from Mild Cognitive Impairment (MCI) individuals [130, 131].

Eye tracking measures show promise in the early diagnosis of cognitive impairments, including Mild Cognitive Impairment (MCI), during working memory tasks [133]. In diseases such as multiple sclerosis (MS), stroke, and Huntington's disease (HD), eye tracking proves valuable for detecting abnormalities and serves as a diagnostic tool [57, 134, 135]. Table 2.1 provides a comprehensive overview of various diseases linked to cognitive impairments, the standard assessment tools used, their limitations, and the effectiveness of eye tracking measures in overcoming these challenges.

For screening mild cognitive impairment, eye tracking measures offer a sensitive, non-invasive, and cost-effective approach, especially when traditional assessments fall short [106, 109, 119, 136–138]. This underscores the potential of eye tracking not only as a diagnostic tool but also as a valuable alternative in cases where conventional methods prove inadequate.

Studies on the related works could reveal a variety of cognitive deficits, including memory loss, lower visual interest, atypical visuospatial behavior, attentional disengagement, motor impairment, and impaired mobility that are associated with diseases like AD, PD, MCI, MS, ALS, HD, stroke and Epilepsy. The eye tracking features like inattentional blindness, error rate, total completion time, scanpath comparison score, fixation duration, saccadic latency, and smooth pursuit could bring the inferences to understand those impairments as shown in Table 2.2. The eye tracking version of the TMT provides those eye tracking features and can bring out the inferences on detecting cognitive impairments.

Table 2.2: Eye tracking features for the detection of cognitive impairment.

Feature	Impairments	Disease
Inattentional blindness	Memory	PD, AD, MCI
Error rate	Memory, imprecise saccadic movement	AD, ALS, MCI, MS
Total completion time	Memory	AD, PD, MCI
Scanpath comparison score	Visual attention, diminished visual curiosity, abnormal visuo-spatial behavior	AD
Less Fixation duration	Attentional disengagement	AD, MS
Higher Saccadic latency	Motor impairment	ALS
Fixation instability	Impaired mobility and cognition	MS
Impaired pursuit	Impaired mobility and cognition	MS

PD − Parkinson's Disease, AD − Alzheimer's Disease, MCI − mild cognitive impairment,
ALS − amyotrophic lateral sclerosis, MS − multiple sclerosis

2.8 Emotional State

Understanding and detecting emotional states is crucial for effective mental health monitoring, providing valuable insights into an individual's psychological well-being. Emotional variations, such as depression, anxiety, and stress, are key indicators of underlying mental health issues [49]. Recognizing and assessing these changes is fundamental for providing timely support and intervention [139]. Accurate detection of emotional states allows healthcare professionals to tailor interventions and therapies to individual needs, improving overall mental well-being [47]. A comprehensive approach that considers various aspects of emotional expression and experience enhances the precision and reliability of emotional state assessments, leading to more effective mental health support [60].

2.8.1 Conventional Methods for the Detection of Emotional State

Conventional methods for detecting emotional states often rely on self-report measures, behavioral observations, and physiological indicators. Self-report measures involve individuals verbally expressing their emotions through questionnaires or interviews [140]. Behavioral observations assess visible behaviors associated with emotions, such as facial expressions, body language, and vocal cues [141]. Physiological indicators monitor changes linked to emotions, such as heart rate, skin conductance, and hormonal levels [142].

Integrating eye tracking technology with these conventional methods provides a more comprehensive and objective understanding of emotional responses by capturing subtle, non-verbal cues related to gaze patterns, pupil dilation, and other eye-related parameters.

2.8.2 Eye Tracking Techniques for the Detection of Emotional State

Understanding and monitoring emotional states are crucial for mental health assessment, providing insights into individuals' psychological well-being amid factors like depression and stress. Eye tracking technology serves as a pivotal tool, offering objective data on emotional responses through the analysis of blink frequency, gaze patterns, and pupil diameter variations. Pupillary changes in response to emotionally arousing stimuli signify the activation of the sympathetic nervous system, providing valuable glimpses into emotional states [59, 60].

Studies investigating the impact of emotional stimuli on eye behavior reveal distinct patterns. Negative images prompt extensive and faster saccades, indicating agitation, discomfort, and avoidance behavior [47]. In contrast, positive images lead to a strong center bias in latitude. Various eye tracking measures, including pupil diameter, blink frequency, saccadic angle, gaze patterns, and fixation duration, directly correlate with emotional reactions, offering insights into emotional states.

The role of eye tracking in emotional state detection is exemplified in research utilizing dynamic cinematic content to evoke emotional reactions. Utilizing eye-tracking variables related to fixations, saccades, and pupil diameter, a study achieved an 80% accuracy in emotion classification with a support vector machine (SVM) classifier. Another study involving emotional movie snippets and automatic categorization based on eye activity achieved a 66% accuracy rate, highlighting statistically significant changes in ocular activity patterns for positive and negative emotions [48, 61, 62].

Various stimuli have been employed in studies investigating emotional states using eye tracking to elicit specific emotional responses. These stimuli encompass medical and non-medical scenarios, including images with negative and neutral content and positive, neutral, and unpleasant images [47]. Additionally, participants are exposed to emotionally charged text comprising negative, positive, and neutral content for reading [143]. These diverse stimuli aim to evoke genuine emotional reactions, and eye tracking measures, such as gaze patterns and pupil dilation, are analyzed to gain objective insights into participants' emotional states during the observation of different visual and textual materials.

2.8.3 Computational Techniques for the Detection of Emotional State

A novel method for assessing visual information utility in emotion recognition is introduced, employing Gaussian fixation distribution and machine learning. In participants with autism, distinctive fixation patterns, especially in emotion recognition tasks, suggest differences in early face processing stages and a potential association with emotion recognition deficits in autism [144].

In a study on emotion recognition utilizing eye-tracking technology, dynamic movie stimuli were presented to evoke emotions in 30 participants. Eye-tracking signals were recorded and analyzed for 18 features related to eye movements and pupil diameter. The study achieved a notable 80% classification accuracy using a support vector machine (SVM) classifier with leave-one-subject-out validation [61].

The study on emotion recognition through eye tracking employed a comprehensive analysis utilizing machine learning, deep learning, and statistical techniques. Various eye tracking features, including fixations, saccades, and pupil diameter, were subjected to analysis. The combination of these analytical approaches contributed to a robust understanding of the relationship between eye movements and emotional responses.

2.8.4 Emotional State and the Associated Diseases

Emotional states significantly influence various diseases and health conditions. The intricate connection between mental and physical well-being is well-established, with emotions impacting both susceptibility to and progression of diseases. Conditions such as cardiovascular diseases, immune system disorders, and chronic inflammatory conditions can be influenced by chronic stress and negative emotional states [145]. Mental health disorders, including depression and anxiety, are recognized contributors to a range of physical ailments. Additionally, unhealthy emotional states may exacerbate existing health conditions and hinder the recovery process. Understanding and addressing the interplay between emotional states and diseases are crucial for comprehensive healthcare and effective disease management.

2.9 Need of Other Physiological Measures

Mental health assessment is intricate, involving a blend of physiological and psychological measures. While eye tracking offers valuable insights, its comprehensive understanding is augmented by other physiological measures. Mental health, with its diverse conditions, requires a multifaceted approach. Physiological measures, such as neuroimaging and genetic markers, provide an objective dimension to subjective experiences. Though various measures contribute, eye tracking stands out for its unique benefits, offering efficient insights into emotional and cognitive states.

Continuous monitoring of human behaviors and physiological responses, utilizing non-intrusive or intrusive sensing techniques, provides real-time insights into cognitive states. Signals related to heart rate, eye movements, brain activity, muscle activity, skin conductance, and speech signals contribute to this understanding [44, 146–150].

Research efforts have successfully utilized a combination of EEG and ECG-based metrics to detect changes in cognitive states, establishing a strong correlation with cognitive conditions [148]. Techniques incorporating speech and handwriting analysis have been proposed for detecting cognitive load, revealing linguistic, grammatical, and handwriting features that offer insights into cognitive processes [150].

In mental health assessment, physiological measures provide additional insights, addressing potential under reporting or self-awareness limitations. While eye tracking measures are valuable, a multifaceted approach encompassing various measures is often useful for a thorough understanding, diagnosis, and treatment of mental health conditions. This inclusive approach allows for a more detailed and individualized assessment and treatment plan, with other physiological measures complementing these tools for a comprehensive understanding of mental health.

Data can be categorized as either subjective or objective. Subjective data relies on personal experiences and perceptions, typically gathered through questionnaires, interviews, and checklists. These assessments have limitations, such as the inability to collect real-time data and the influence of psychological biases. Examples of standardized questionnaires used for subjective assessment include the NASA Task Load Index (TLX) [16], Trier Social Stress Test [17], State-Trait Anxiety Inventory (STAI) [18], Karolinska Sleepiness Scale (KSS) [18], Shortened State Stress Questionnaire (SSSQ) [18], and Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) [19]. Objective data, based on observable and measurable factors, is less susceptible to biases. This type of data can be divided into two categories: behavioral and physiological.

Behavioral data includes information such as facial expressions [21], audio signals [22], gestures [23], head movements [24], hand movements [151], leg movements [25], eye movements [26], and eye contacts [152, 153]. These data can be collected through audio-video recordings or using wearable devices. While behavioral data can be collected in real time, they may be influenced by voluntary control and cultural differences, potentially leading to inaccuracies.

2.9.1 Physiological Data

It provides continuous and consistent monitoring and is less affected by cultural or linguistic variations. This category includes signals such as eye tracking [14,20,36], accelerometers, gyroscopes, pupil corneal reflection, EEG [16, 27], ECG, photoplethysmogram, GSR, pressure-sensitive touch screen [29], heart rate, sound, temperature, oxygen saturation, and salivary cortisol levels [30]. These physiological signals are collected using various sensing technologies and offer valuable insights into an individual's mental health status [84].

The integration of physiological data into mental health assessment has revolutionized the field, providing an understanding of emotional and psychological states. This shift acknowledges the intrinsic connection between mental health and physiological responses, from heart rate variability to brain activity patterns. Unimodal and multimodal physiological approaches offer valuable insights, empowering healthcare professionals and researchers for more precise mental health assessments, leading to enhanced interventions and improved quality of life.

Multimodal physiological approaches integrate data from various sources, recognizing the complex nature of mental health influenced by diverse physiological factors. In education, combining digital, physical, psychometric, physiological, and environmental data helps monitor students' cognitive load, attention levels, and emotional engagement [154]. This comprehensive approach, using eye movement patterns, EEG readings, and electrodermal activity, provides educators with detailed insights into the learning experience [155].

Research on participants' experiences during game sessions used a multimodal approach, incorporating eye tracking, EEG readings, video recordings, and wristband data [26]. This method offered a thorough understanding of cognition, attention, and emotion. Additionally, the CareCam system, which monitors employees' vital signs and facial expressions through a webcam, illustrates using multimodal data to address stress and well-being [156].

These examples demonstrate the power of multimodal data collection and analysis in various contexts, enabling tailored interventions and improving our understanding of complex phenomena, ultimately enhancing well-being and experiences [157].

2.10 Summary

Traditional mental health assessments, involving clinical interviews, psychometric tests, and observations, provide valuable insights but can be limited. Recently, technologies like eye tracking combined with machine learning and deep learning algorithms have revolutionized mental health monitoring. These techniques allow for a more detailed understanding of cognitive and emotional states by analyzing eye movement data, enhancing the precision of mental health assessments and enabling real-time insights for more effective interventions.

In mental health monitoring, diverse indicators such as attention, engagement, fatigue, stress, cognitive processes, social interaction, behavior, anticipation, and performance together contribute to understanding an individual's mental state. Cognitive load refers to the mental effort being put into the task, cognitive impairment reflects deficits in cognitive function, and emotional state includes emotional well-being [40]. These parameters are crucial for a detailed mental health assessment, guiding interventions, and identifying potential challenges.

Eye tracking technology holds much promise in the field of mental health monitoring because it provides objective and quantifiable data. It assesses mental state parameters, cognitive load, cognitive impairment, and emotional state. Non-intrusive effectiveness is shown by linking models of major eye movements and cognitive load [43]. It identifies task complexity and mental fatigue through characteristics like saccade features and fixation duration [99]. Additionally, it detects early-stage cognitive impairments and helps differentiate subtypes of mild cognitive impairment [45, 46]. It is useful in neurological disorders like Alzheimer's and Parkinson's, enabling early detection of cognitive decline [46, 56, 57].

Tasks such as gaming, arithmetic problem-solving, and coding reveal cognitive load, while memory and attention tests provide insights into cognitive impairment. Eye tracking combined with stimuli like pleasant, unpleasant, and neutral images, and emotionally charged text allows for the objective analysis of gaze patterns and pupil dilation. This comprehensive approach enhances understanding of cognitive load, cognitive impairment, and emotional states [47, 106, 115, 119, 143].

2.11 Research Gap

There exist research gaps that come with developing the classification system of the mental state parameter by eye gaze tracking. The most important is the absence of publicly available datasets for mental health eye tracking. This hampers benchmarking, replicability, and collaboration between researchers. These deprive the shared resources of the research community in the progress of understanding mental states using eye gaze tracking. User-specific mental health monitoring systems represent a challenge due to the absence of user-tailored labeled data, which blocks the rise of effective prediction models. This requires special effort in collecting and curating personalized datasets for model training while at the same time encouraging open collaboration in developing publicly available datasets for the field.

Second, the potential of the eye tracking parameters to influence indicating mental illness is still at the development stage. Identify highly specific eye movement patterns that are directly related to specific mental health issues and disorders, ideally offering a good biomarker of the condition. The exact specifications for how this eye tracking data will assist in helping to inform better markers or states of mental health is currently vastly unexplored. Continuing to build knowledge in these connections is an element that will increase the use of this technology in mental health assessment over time.

Second, there has been little work on developing assistive tools to help healthcare professionals use eye tracking effectively in mental health assessment. Integration of eye tracking with mental health indicators and modeling holistic assessments is an untouched domain but holds great promise in offering a detailed account of a person's mental state. Effective development of prediction systems for mental health with eye gaze tracking technology needs to address these research gaps.

This chapter presented a comprehensive literature survey on utilizing eye measurements to detect mental state parameters such as cognitive load, impairment, and emotional arousal. From this survey, one would gather that the most major challenge in this area is the absence of publicly available eye-tracking datasets. To address this gap, several datasets were created. The following chapter elaborates on the datasets with a focus on the procedures followed during their collection and details of datasets created.

Chapter 3

Data Collection and Analysis: Procedures, Datasets, and Visualization

3.1 Introduction

In the domain of mental health assessment, the utilization of eye gaze data holds immense promise as a powerful tool for understanding and monitoring various mental health indicators. However, the development of effective machine learning and deep learning models in this field faces substantial challenges, primarily attributed to the scarcity of publicly available eye tracking datasets specifically designed for mental health assessment. This scarcity hinders the training and validation of models that rely on diverse and labeled data to achieve robust performance.

This chapter seeks to address a critical gap in the existing research by introducing novel datasets specifically curated to facilitate advancements in mental health monitoring through eye gaze data analysis. Additionally, the challenges associated with the absence of specialized datasets and the need for labeled data pertaining to mental health indicators, such as cognitive load, cognitive impairment, and emotional state, underscore the importance of creating a resource that can drive innovation in this vital sphere.

The primary purpose of this dataset is multifaceted. First and foremost, it aims to overcome the challenge of limited publicly available eye tracking datasets for mental health assessment. The first objective is fulfilled through the creation of four distinct eye-tracking-based datasets, ET_MT_CL, ET_TMT_CL, ET_TMT_CI, and ET_Video_ES. Collectively, these datasets offer a valuable resource to the research community, providing a targeted solution to the scarcity of specialized datasets in the field.

ET_MT_CL, based on mathematical problems, captures eye gaze data to assess the cognitive load, providing insights into mental states based on their cognitive load during mathematical tasks. ET_TMT_CL, centered around the TMT, utilizes eye tracking data to detect cognitive load, enhancing understanding of mental states related to task performance. ET_TMT_CI, also based on the TMT, incorporates cognitive impairment into analysis, facilitating the detection of mental states associated with cognitive challenges. ET_Video_ES integrates eye tracking data with video stimuli to capture an emotional states, enriching the dataset for the detection of emotional states through visual responses. These datasets collectively contribute to the detection of emotional states based on cognitive load, impairment, and emotional state, advancing the understanding of mental health through diverse eye gaze analyses. It enables researchers to explore and refine machine learning and deep learning models for mental health assessment, providing a foundation for robust and reliable model development.

Furthermore, an additional multimodal dataset EmoRPhyE is introduced to provide a comprehensive understanding of emotional states. This dataset integrates data from diverse physiological measures and modalities, including electrocardiogram (ECG), Photoplethysmography (PPG), Galvanic Skin Response (GSR), respiratory signal, and eye tracking data. This innovative approach aims to capture a holistic view of individuals' emotional health by combining information from various sources, paving the way for more nuanced and accurate assessments. The first objective of satisfying the need for dataset creation is successfully achieved through these datasets, laying the groundwork for further exploration and development. The subsequent section outlines the data collection procedures, eye gaze dataset creation, details on eye gaze data analysis and feature extraction algorithms, various visualization techniques applied to eye tracking data, multimodal dataset creation and the software and hardware utilized in this study.

3.2 Data Collection Procedures

As part of the unimodal and multimodal data collection, specific procedures were meticulously followed to ensure the proper acquisition of data. Approval for the unimodal multimodal data collection was obtained from the ethical committee. The exclusive involvement of human subjects characterized the data collection process. The subsequent sections elaborate on the specific procedures employed for participant recruitment and the execution of experiments in the context of eye-tracking-based data collection. The dataset naming convention is systematically structured, commencing with the mode of data collection, succeeded by the stimulus employed, and concluding with the specific mental health parameter targeted by the dataset. This section explains the procedures for participant recruitment, experimental setup, privacy and confidentiality maintenance, and data exclusion criteria.

3.2.1 Participant Recruitment

The recruitment of participants for the study will adhere to strict inclusion criteria. Participants will be thoroughly informed about the purpose and objectives of the research study. Their informed consent, signifying their willingness to participate, will be obtained. Additionally, demographic information and signed consent forms will be collected from all participants before they participate in the study.

3.2.2 Experimental Setup

Participants will be instructed to sit comfortably in front of the eye tracker, ensuring they are within the recommended operating distance of 50cm to 60cm limit for optimal performance during the experiment. The eye camera view for each participant will be carefully adjusted to ensure high-quality data capture. The experiments were conducted in a controlled environment with constant room lighting to ensure data accuracy and reliability. Before the commencement of each experiment, participants will receive a detailed explanation of the data collection process. In addition, their demographic information will be collected, and if necessary, any presurvey required for the study will also be administered. To ensure the accuracy and reliability of the data collected, calibration procedures will be meticulously conducted for each participant before the start of every experiment. This calibration step is crucial in optimizing the performance of the eye-tracking equipment and ensuring that the subsequent data accurately reflects the participants' gaze patterns and responses to stimuli. In the multimodal approach, other necessary physiological sensors will also be connected to the participant to facilitate concurrent data collection, including gaze patterns and other relevant physiological signals.

Participants will be presented with stimuli during the experiment, and the corresponding data collection process will begin. Participants will be instructed to observe the stimuli while their gaze pattern and other associated physiological signals are recorded. It is important to note that the experimental setup will not harm or discomfort the participants. They will be seated in stationary chairs with limited head movements during the entire duration of the experiment, which includes the explanation of the process, calibration, and the recording of eye movements.

3.2.3 Confidentiality

The privacy and confidentiality of the participants will be rigorously maintained throughout the study. Participants will not be personally identified in any reports or publications resulting from this research. Any personal information collected during the study will be safeguarded and treated in accordance with legal regulations. This comprehensive data collection procedure ensures the ethical and meticulous acquisition of the necessary data while prioritizing participant safety, consent, and confidentiality.

3.2.4 Data Exclusion Criteria

To ensure data quality and reliability, participants who fail to successfully pass the calibration process of the eye-tracker or have a substantial amount of missing data in their eye-tracking records will be excluded from subsequent data analyses. In such cases, replacements will be recruited and subjected to the same exclusion criteria, if necessary.

3.2.5 Eligibility Criteria

1. Inclusion criteria
   - (a) Diagnosed case of mental illness
   - (b) No other ocular pathology that can affect the optic nerve or visual field
   - (c) No other ocular pathology that can affect eye movements
   - (d) No recent surgery within the last three months
   - (e) Normal healthy participants
   - (f) Typically developing participants
2. Exclusion criteria
   - (a) The participants with drooping eyelids, contact lenses, squint, or difficult glasses for the study.
   - (b) Use of alcohol and psychotropic medication is known to affect eye movements.
   - (c) Participants with glasses with more than one power: bifocals, trifocals, and progressives.
   - (d) Participants with eye surgery: corneal (e.g., LASIK), cataract, intraocular implants

(e) Participants with eye movement or alignment abnormalities: lazy eye, strabismus, nystagmus

3.3 Dataset Creation

As the primary objective of this thesis work, a total of five datasets were diligently collected and generated. These datasets are systematically categorized into four unimodal datasets derived from eye tracking data and a single multimodal dataset that integrates eye tracking with other physiological signals. Comprehensive details of each dataset are presented in the table 3.1.

Table 3.1: Dataset details

Task	Population	Size	Mental state	Dataset
Mathematical questions	Students age group 20-30	20	Cognitive load	ET_MT_CL
TMT	Students age group 19 to 22	100	Cognitive Load	ET_TMT_CL
TMT	Allied health professionals age group 20 to 54	31	Cognitive Impairment	ET_TMT_CI
Calm and Stress video	Allied health professionals age group 26 to 42	6	Calm/Stressful	ET_Video_ES
Pleasant & unpleasant images	Students age group 20- 40	30	Emotional state	EmoRPhyE

Ethical considerations were of utmost importance throughout the data collection process. The research team obtained proper clearance from the institutional ethical committee. Additionally, strict adherence to the established inclusion and exclusion criteria for data collection was maintained to ensure the quality and validity of the datasets.

Furthermore, prior to their participation in the experiments, all participants were comprehensively briefed about the study's objectives and procedures. Informed written consent was obtained from each participant, affirming their willingness to partake in the data collection. These ethical safeguards were meticulously observed to safeguard the participants' rights and privacy and maintain the research's ethical integrity. The following sections explain the details of the five datasets.

3.3.1 Eye Gaze Datasets

The subsequent sections provide comprehensive details regarding the eye tracking based datasets, encompassing the data collection process, the stimuli utilized, and the creation of sample datasets derived from each. Additionally, the extraction of features from eye gaze data is elaborated upon, including the algorithms employed for feature extraction. Furthermore, the section delves into various visualization techniques utilized to represent eye gaze data effectively.

ET_MT_CL Dataset

The ET_MT_CL dataset captures eye tracking data while presenting mathematical questions (MT) stimuli specifically designed to investigate the detection of the mental state parameter known as cognitive load (CL). The study was designed to investigate the impact of stressors on individuals by increasing cognitive load, utilizing a sample of 20 students aged between 20 and 30. The participants had an average age of 25, with a standard deviation of 4, and included ten female participants. Eye movements were recorded using the SMI RED-n Professional Eye Tracker, operating at a 60 Hz sample rate. Calibration and validation processes were conducted before each experiment to ensure precise and accurate eye tracking data. A consistent viewing distance of 50 to 60 cm was maintained between the participant's eyes and the computer screen.

The experimental study followed the procedures outlined in section 3.2. Participants were instructed to sit facing the machine connected to the eye tracker and undergo calibration. Upon successful calibration, stimuli were presented to the participants. These stimuli comprised mathematical questions, and participants were required to solve each expression and verbally communicate their answers within the allocated time frame.

The task involved the presentation of twenty mathematical questions categorized into four difficulty levels: easy, normal, moderate, and difficult, as shown in Figure 3.1. The difficulty levels were determined based on the number of operands and operators employed in the questions. For example, Easy-level questions featured two operands, while Normal-level questions incorporated three operands with left-to-right operator precedence. Moderate-level questions also used three operands but with right-to-left operator precedence, and difficult-level questions introduced a higher level of complexity with four operands and operators with randomly assigned precedence. Sample stimuli are shown in Figure 3.2. During the task, participants were given three seconds to answer the easy, normal, and moderate questions and five seconds for the difficult questions, which aimed to investigate how variations in cognitive load, induced by different levels of mathematical complexity, could affect stress levels. Notably, the study focused exclusively on physiological measures, particularly pupil diameter and the number of blinks, with the emotional states of participants not taken into account.

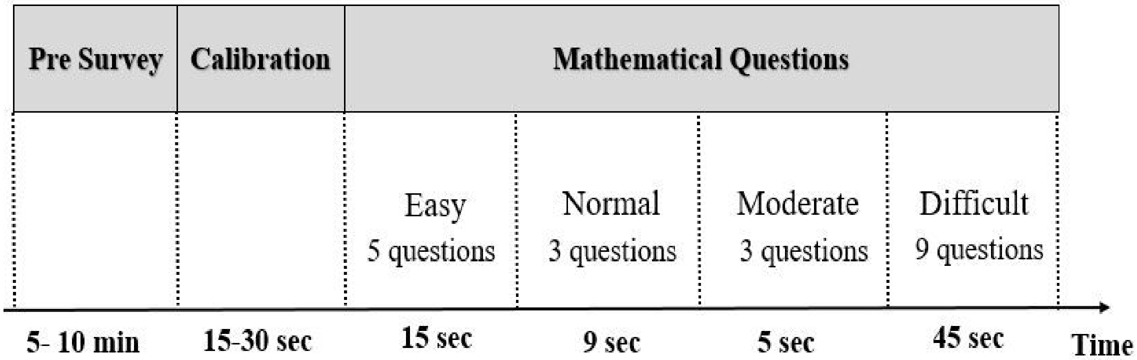


Figure 3.1: Experimental setup for ET_MT_CL data collection

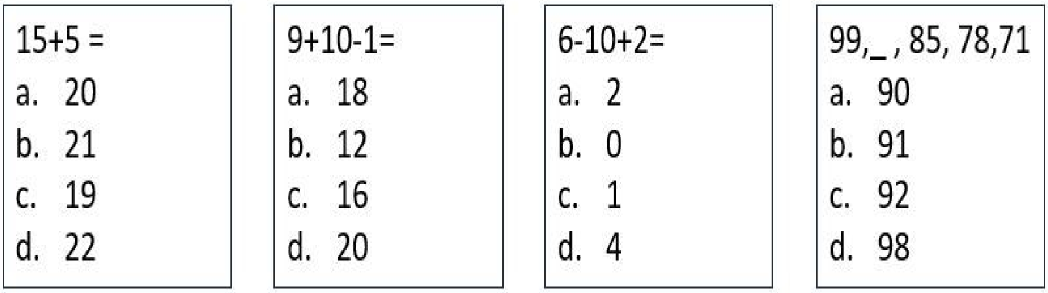


Figure 3.2: Stimuli used for ET_MT_CL datset creation

The collected eye tracking data comprises raw information containing the gaze X and Y coordinates along with pupil diameter as shown in Table 3.2. These details will be maintained for each participant. Chapter 4 provides a detailed explanation of the subsequent feature extraction process and its specifics. Based on the validity of the data, the data from any of the eyes will be considered for further feature extraction.

This structured data collection procedure was designed to comprehensively explore the relationship between cognitive load, stress, and physiological responses, ensuring data accuracy and participant comfort.

ET_TMT_CL Dataset

The ET_TMT_CL dataset encompasses eye tracking data collected with the TMT stimulus, designed to classify the mental state parameter cognitive load (CL). The study participants consisted of 100 students aged between 19 and 22. The gender distribution among the participants was 60 male students and 40 female students. The demographic details are shown in figure 3.3.

Table 3.2: The raw eye gaze data from SMI REDn eye tracker

Features	Description
Time	Timestamp
Type	Indication of different trials
L Raw X [px]	Left eye's Gaze X coordinate
L Raw Y[px]	Left eye's Gaze Y coordinate
R Raw X[px]	Right eye's Gaze X coordinate
R Raw Y[px]	Right eye's Gaze Y coordinate
L Diameter [mm]	Left eye's Pupil diameter
R Diameter [mm]	Right eye's Pupil diameter
L validity	Left eye data validity indicator. 1-valid, 0-not valid
R validity	Right eye data validity indicator.1-valid, 0-not valid

The experimental setup for data collection is depicted in Figure 3.4. Prior to commencing the main task, participants were briefed on the study objectives and the tasks involved in data collection. Calibration tests were then administered, serving as a crucial step in eye tracking to ensure precise identification and tracking of each participant's gaze point.

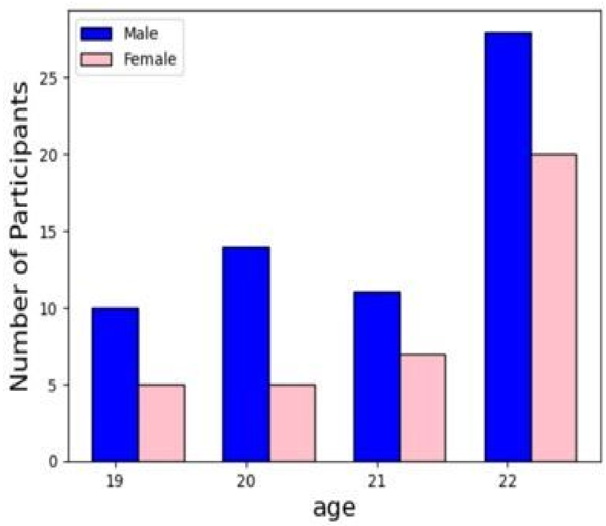


Figure 3.3: Demographic details of ET_TMT_CL dataset

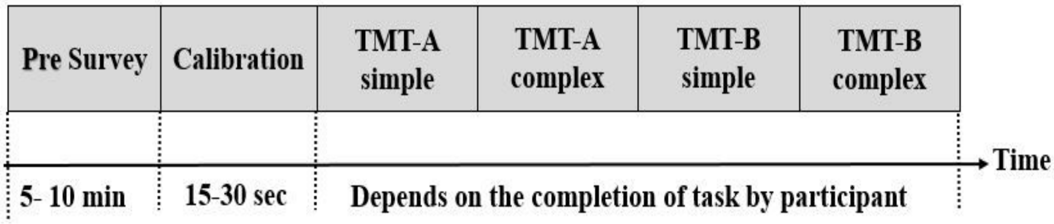


Figure 3.4: Experimental setup for ET_TMT_CL data collection

The chosen cognitive task for this investigation was the TMT, a widely used psychological assessment that evaluates various cognitive functions, including attention, visual scanning, and processing speed. Here, the eye tracking version of TMT is used as the stimulus. The TMT consists of two parts: TMT-A and TMT-B. TMT-A simple is a simple random number sequence from 1 to 8. TMT-A complex is a random assortment of numbers between 1 and 25. TMT-B simple is a random combination of alphabet letters from A to D and numbers from 1 to 4. TMT-B complex is a random combination of alphabet letters from A to L and numbers from 1 to 12. TMT stimuli are shown in figure 3.5. The data for this study was collected using the VT3 Mini eye tracker, a specialized device for recording eye movements and gaze data. Data was recorded at a sampling frequency of 60 Hz, allowing for capturing eye movements and gaze points at a rate of 60 times per second, providing fine-grained insights into participants' visual attention.

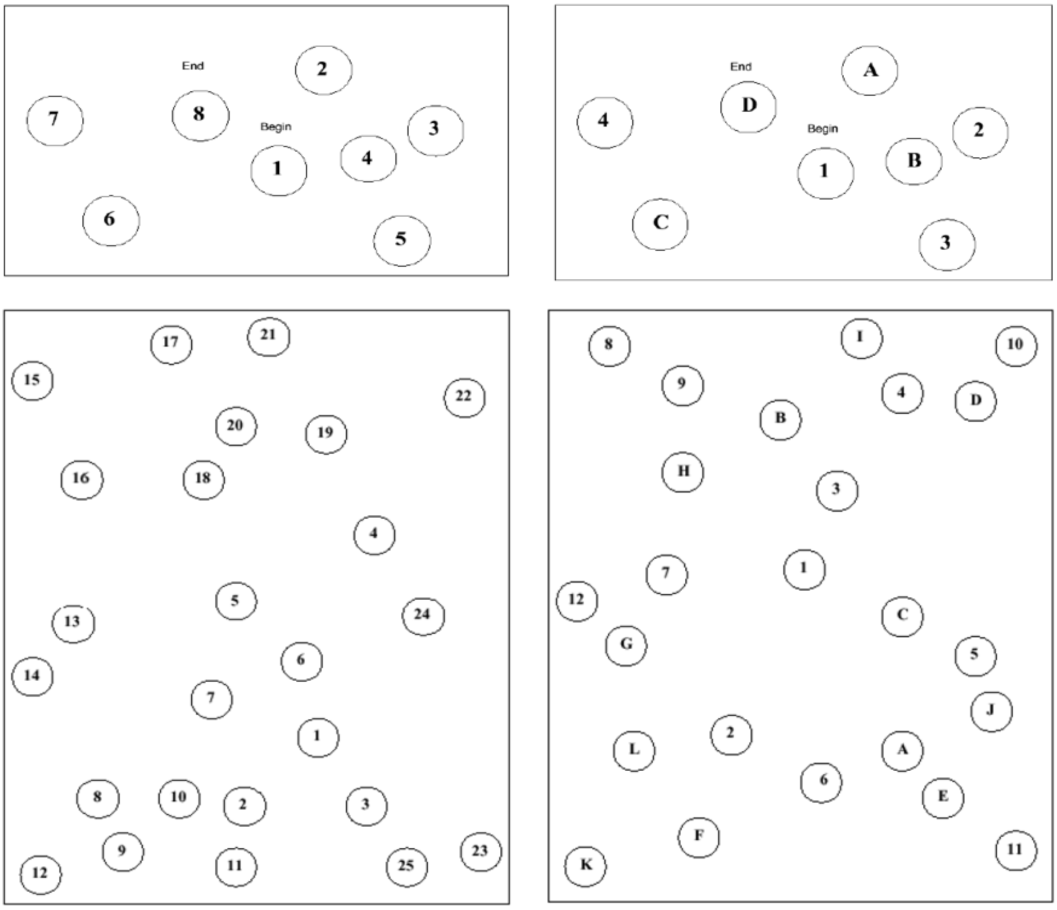


Figure 3.5: Trail Making Test stimuli

After achieving accurate calibration, participants were presented with the eye tracking version of the TMT. They were instructed to navigate through a series of numbered and lettered circles displayed on the screen as part of the stimulus. The stimulus was displayed in the order of TMT-A simple, TMT-A complex, TMT-B simple, and TMT-B complex. In TMT-A, they were asked to visually connect the numbers in ascending order, while in TMT-B, they needed to visually connect the numbers and letters in ascending order, alternating between numbers and letters. Participants were also directed to read the numbers and letters aloud during the test, introducing a verbal component that can offer insights into the task performance.

In the TMT-A and TMT-B assessments, two sets of stimuli were employed to gauge changes in cognitive load during participants' task performance. These variations were introduced to evaluate participants' cognitive abilities comprehensively and to assess their capacity to sequence and connect characters under differing levels of complexity. By comparing the performance in simple and complex tasks, the study aimed to shed light on how the cognitive demands of the TMT differ under varying levels of complexity, providing valuable insights into participants' cognitive processing capabilities.

The eye tracker captured participants' eye movements throughout the TMT, generating each participant's raw gaze points and pupil diameter as shown in Table 3.3. Further feature extraction and its analysis are explained in Chapter 4.

Table 3.3: Raw eye gaze data from VT3 Mini eye tracker

Features	**Description**
Time	Timestamp
Left Pupil Pos X	Left eye's Gaze X coordinate
Left Pupil Pos Y	Left eye's Gaze Y coordinate
Right Pupil Pos X	Right eye's Gaze X coordinate
Right Pupil Pos Y	Right eye's Gaze Y coordinate
Left Pupil Diameter (mm)	Left eye's Pupil diameter
Right Pupil Diameter (mm)	Right eye's Pupil diameter
Left Found	Left eye data validity indicator.
Left Found	Right eye data validity indicator.

ET_TMT_CI Dataset

The ET_TMT_CI dataset captures eye tracking data during the TMT stimulus presentation, specifically designed to assess the mental state parameter, cognitive impairment (CI). The data was collected from 31 healthy participants aged 20 to 54 working within a hospital environment. The mean age of the participants was 30.6 years, with a standard deviation of 8.6 years. The demographic details are shown in Figure 3.6. A total of 40 participants were recruited for the study, and we collected data from those participants. Out of the collected data, 9 participants' data were discarded due to technical issues and data loss, leaving a total of 31 participants' data for analysis.

The eye-tracking versions of the TMT-A and TMT-B, as shown in Figure 3.5 were utilized as stimuli. These tests encompassed simple and complex sequences of numbers and letters, evaluating participants' ability to follow ascending patterns. The data were collected using the SMI REDn Professional eye tracker, which operated at a sampling frequency of 60 Hz.

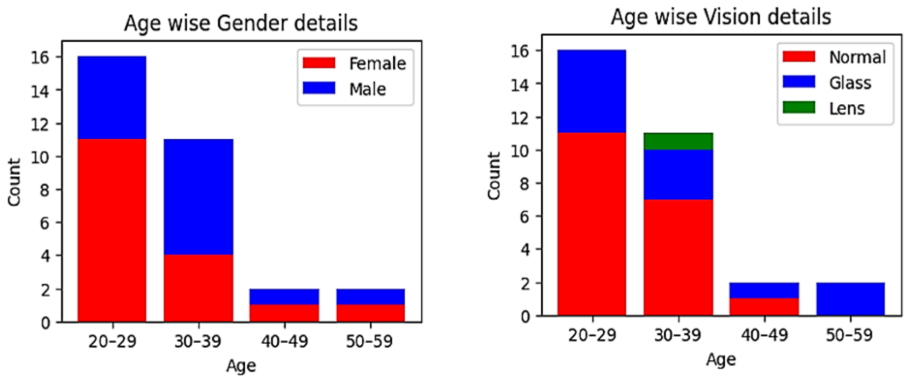


Figure 3.6: Demographic details of ET_TMT_CI dataset

The stimulus was displayed in the order of TMT-A simple, TMT-A complex, TMT-B simple, and TMT-B complex, as shown in Figure 3.7. The participants were instructed to look at the ascending numbers in TMT-A and the ascending combination of numbers and alphabets in TMT-B. While watching it, the participants were instructed to speak the number or alphabet aloud. The individual was then asked to complete the conventional TMT after the eye tracking version. It is based on the paper-pencil method, in which participants were instructed to link the alphabets or numerals in ascending sequence. The eye tracking and traditional TMT followed the same order of stimulus. The raw data obtained has features like timestamp, gaze, and pupil diameter details, as shown in Table 3.2. Since there were 31 participants, 31 raw data files were generated to store the raw data information. The BeGaze 3.7 has been used to extract low-level and middle-level features. Further high-level feature extraction is explained in Chapter 5.

ET_Video_ES Dataset

The ET_Video_ES dataset records eye tracking data while presenting calm and stressful video stimuli designed to assess a person's emotional state. The experimental study was performed on allied health professionals (n=6, 3 male, Mean age= 33.5, SD=5.6 , age range= 26 to 42) [36]. The demographic details are shown in Figure 3.8. Figure 3.9 shows the procedures followed in the experimental setup of ET_Video_ES dataset collection. A 10-minute video including 5 minutes of calm and 5 minutes of stressful video was utilized as a stimulus [99]. Before the actual stimulus, an introductory segment with a short animated movie was played for

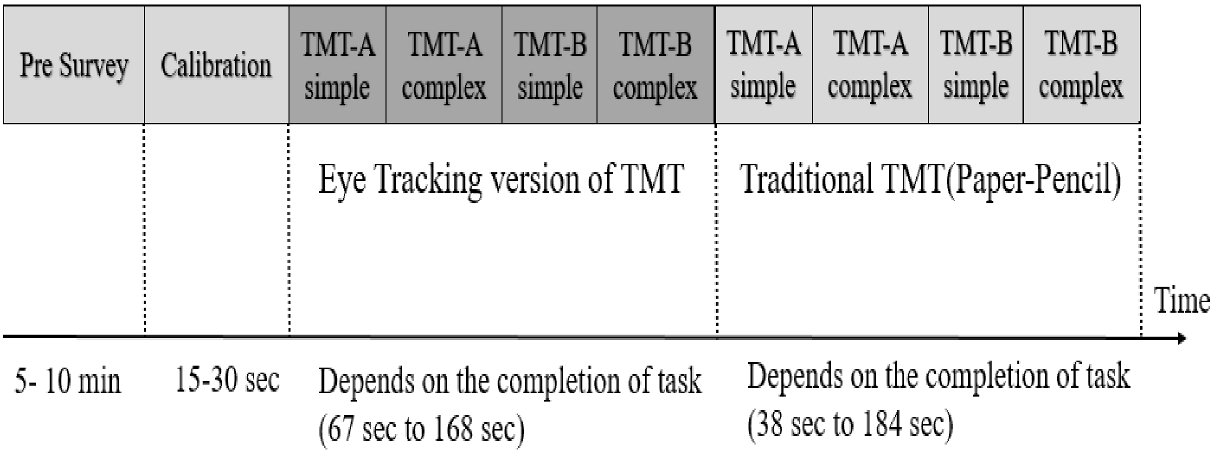


Figure 3.7: Experimental setup for ET_TMT_CI data collection

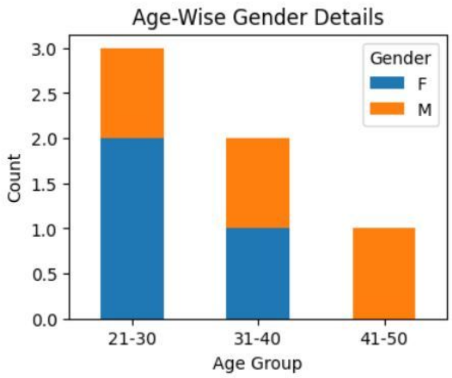


Figure 3.8: Demographic details of ET_Video_ES dataset

Figure 3.9: Experiment procedure of ET_Video_ES dataset

a 2-minute and 30 seconds. The calm [158] and stressful video scenes [159] were presented subsequently. The calm segment of the video was selected based on its ability to induce relaxation within 5 minutes, as per the recommendation of specialists in stress and anxiety therapy. The introductory video was not included in the analysis but served to acquaint the participants with the experimental setup. The calm video period was considered the baseline phase, with the expectation that the participants would experience a state of relaxation during this time. The sampling frequency of 60 Hz allowed us to capture 60 samples of raw eye gaze data per second. In total, a substantial dataset of observations was accumulated, consisting of 10 minutes or 600 seconds. This translates to a vast raw eye gaze data repository comprising 36,000 samples for a participant. The raw eye gaze data includes (X, Y) gaze coordinates, pupil diameter, and timestamp, as shown in Table 3.2.

The eye tracking data was collected using the SMI Redn Professional Eye Tracker (Company: SensoMotoric Instruments, Germany) with a sampling frequency of 60Hz and Experiment Center 3.7 software, which provides comprehensive tools for stimulus presentation and precise data collection.

GSR data was also collected simultaneously using a grove-GSR sensor, with two electrodes attached to two fingers of a hand as shown in Figure 3.10, [160]. It measures the electrical resistance of the subject's skin, and this information is then used to produce an output voltage, typically measured in millivolts (mV). The participant was given the wearable band and instructed to wear it in accordance with the guidelines. Arduino integrated development environment(IDE) software was used to collect GSR data. Eye tracking and GSR data were collected using the same machine, ensuring the system clock remained consistent. To synchronize the data collection, a software trigger was set to initiate eye tracking and GSR measurements simultaneously, resulting in timestamps that are accurately aligned across the datasets.

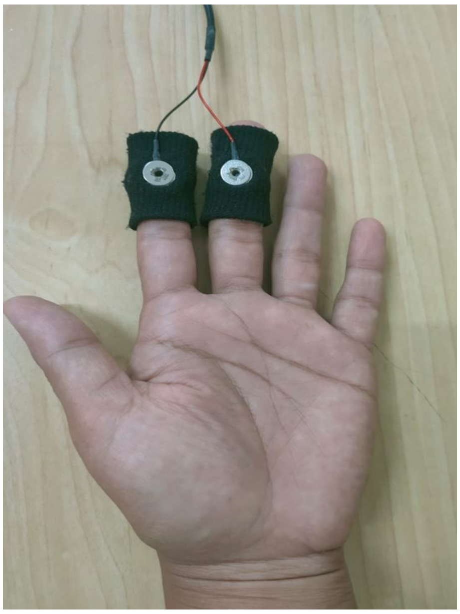


Figure 3.10: GSR sensor used for ET_Video_ES data collection

3.3.2 Eye Gaze Data Analysis

Four eye gaze datasets were created to study mental state parameters like cognitive load, impairment, and emotional arousal. The raw eye gaze data collected needs pre-processing and feature extraction tailored to each application for efficient results. Various visualization techniques are available to help understand the eye gaze data, which will be explained in the following section, along with feature extraction.

The raw data extracted from each dataset, as illustrated in Tables 3.2, 3.3, will undergo pre-processing. This process entails the removal of erroneous data, particularly instances where participants did not look at the screen, resulting in unacceptable data. Participants with a substantial amount of erroneous data will have their data discarded. Subsequently, after pre-processing, a diverse range of features will be extracted to suit the specific applications. The following section outlines the fundamental eye gaze features and the algorithms utilized for feature extraction.

Eye Gaze Features

Eye gaze features based on fixation, saccade, blink, and pupil diameter play a crucial role in eye tracking and can provide insights into a person's visual attention and cognitive processes. Here are details on each of these features:

1. Fixation: Fixation refers to the stable and sustained gaze of the eyes on a specific point or object in the visual field. During a fixation, the eyes remain relatively still. The duration of a fixation typically ranges from 100 to 500 milliseconds with a maximum dispersion of 100 pixels. Fixations suggest that the individual is processing or perceiving information from the location where they are focused. Researchers use fixations to identify what visual elements or areas attract a person's attention and for how long. An identified fixation event will have an X and Y coordinate.

2. Saccade: Saccades are rapid, involuntary eye movements that quickly shift the gaze from one point of interest to another in the visual field that lasts for 30 to 100 milliseconds. Their high velocity and short duration characterize saccades. They allow the eyes to reposition and explore the visual scene by moving from one fixation point to another. Saccades are instrumental in studying visual scanning patterns, eye movement coordination, and how attention is directed between different objects or points of interest. An identified saccade event will have two gaze coordinates (x1,y1) and (x2,y2), where (x1,y1)indicates the source point and (x2,y2) indicates the destination point.

In Figure 3.11, the scanpath depicts a combination of fixations and saccades.

Each circle within the scanpath represents a fixation, indicating that the person's gaze remained at a specific location for a fixed amount of time. The diameter of these circles reflects the fixation duration, signifying how long the individual focused their attention on that particular location. The lines connecting these fixations illustrate saccades, representing the rapid eye movements between fixations. This scanpath provides a visual representation of how the person's gaze moved and fixated during a specific task or observation.

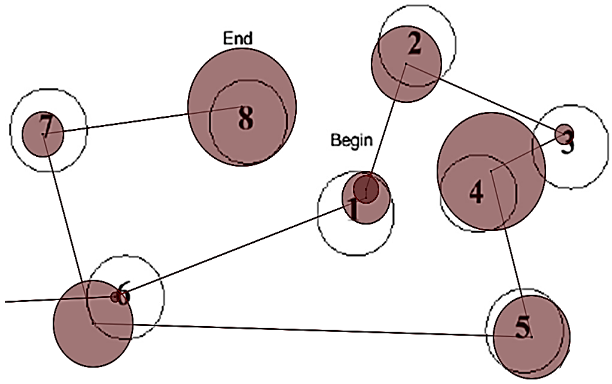


Figure 3.11: Scanpath

3. Blink: A blink is a brief eyelid closure temporarily obstructing vision. Blinks are rapid and typically last for about 100-400 milliseconds. People blink multiple times per minute, with an average blink rate of around 12-15 blinks per minute. Blink occurs when the pupil diameter is zero or when horizontal and vertical gaze positions are zero, as explained in Algorithm 3.1. When any of these conditions are satisfied, a blink event is recorded. Given the input of gaze coordinates and the minimum and maximum duration required to classify an event as a blink, the algorithm aims to identify instances of eye blinks and return the start time, end time, and duration for each detected blink. Blink data can be used to assess aspects of cognitive load, attention, and even aspects of emotional or physiological states.

4. Pupil Diameter: Pupil diameter refers to the size of the dark, central aperture of the eye, known as the pupil. Pupil diameter can change in response to different stimuli. Pupil diameter varies from 2 to 8 millimeters in adults, depending on lighting conditions and the level of arousal or cognitive effort. Pupils constrict to increased light and dilate in response to reduced light, arousal, or cognitive load. Pupil diameter serves as a measure of cognitive load, emotional state, and even changes in attention. It is valuable for understanding the psychological and physiological responses of individuals.

Algorithm 3.1 Blink Detection Algorithm

```
function DetectBlinks(GazeCoordinates, MinBlinkDuration)
    Initialize an empty list Blinks
    Initialize a duration timer
    Initialize a blink flag
    StartTime ← 0
    for all Coordinate in GazeCoordinates do
        if Coordinate.PupilDiameter = 0 then
            if blink flag is false then
                StartTime ← Coordinate.Time
                Set blink flag to true
            end if
        else
            if blink flag is true then
                EndTime ← Coordinate.Time
                Duration ← EndTime − StartTime
                if Duration ≥ MinBlinkDuration then
                    Add blink with StartTime, EndTime, and Duration to Blinks
                end if
                Set blink flag to false
            end if
        end if
    end for
    return Blinks
end function
```

The criteria for detecting fixation coordinates, saccade coordinates, and blink events are typically defined within specific ranges, but these conditions can be adjusted to some extent based on the particular requirements of different applications. For instance, in some scenarios, blinks lasting less than 70 milliseconds may be excluded from analysis, and there might not be a set maximum limit for blink duration. Similarly, the acceptable range of values for detecting fixation and saccade events can be customized to match the specific needs of each unique application. This flexibility allows for fine-tuning eye tracking parameters to best suit the demands of various contexts.

These eye gaze features are essential for analyzing visual attention, cognitive processes, and emotional responses during tasks, making eye tracking a versatile tool in various fields, including psychology, neuroscience, marketing, and human-computer interaction.

Algorithm 3.2 Fixation Detection Algorithm (I-DT) [161]

```
function DetectFixations(Points, Threshold, DurationThreshold)
    Initialize an empty list Fixations
    while Points is not empty do
        Initialize an empty list Window
        Initialize a duration timer
        Add the first point in Points to Window
        Remove the first point from Points
        while dispersion of Window ≤ Threshold do
            Find the dispersion of Window
            if duration of Window ≥ DurationThreshold then
                Note a fixation at the centroid of Window
                Add the centroid to Fixations
                Clear Window
            end if
            if Points is not empty then
                Add the first point from Points to Window
                Remove the first point from Points
            end if
        end while
    end while
    return Fixations
end function
```

Algorithm 3.3 Saccade Detection Algorithm [161]

```
function DetectSaccades(Points, Threshold, DurationThreshold)
    Initialize an empty list Saccades
    Initialize an empty list Fixations
    Fixations ← DetectFixations(Points, Threshold, DurationThreshold)
    NumFixations ← Length(Fixations)
    for i ← 1 to NumFixations − 1 do
        Duration ← GetDuration(Fixations[i], Fixations[i + 1])
        if Duration ≤ 100 milliseconds then
            Add saccade from Fixations[i] to Fixations[i+1] to Saccades
        end if
    end for
    return Saccades
end function
```

The device, an eye tracker, typically provides raw data, including gaze x and y coordinates, timestamps, and pupil diameter, as shown in Tables 3.2 and 3.3. The sampling frequency of the eye tracker determines the number of samples obtained in one second. For event detection and to extract additional features from this raw data, specialized algorithms such as Identification by Dispersion Threshold (I-DT) [161] or Velocity-Threshold Identification(I-VT) can be employed. These algorithms help identify specific events, like fixations, saccades and blinks, and provide a structured way to process and analyze the eye tracking data.

Alternatively, eye tracking software applications associated with the eye trackers, such as BeGaze and Ogama, can be utilized. These software tools often include features for event detection, data visualization, and analysis, making it easier for researchers and practitioners to work with eye tracking data efficiently.

The I-DT algorithm, as shown in Algorithm 3.2 and Algorithm 3.3, are designed for fixation and saccade event detection in eye tracking data [161]. It works by initializing a window over a sequence of data points and gradually expanding it to cover a duration threshold. The input to both the algorithms are the X and Y gaze coordinates of any eye(left/right) and the time. If the dispersion of data points within this window falls below a certain threshold, the points are labeled as a fixation, and their centroid is noted. Fixation coordinates are detected iteratively as the window moves through the data. This algorithm helps identify periods when the eyes are relatively still and focused on a single point. The movement from one fixation to another is considered as a saccade. The fixation algorithm provides the start time, end time, duration, and fixation coordinates. The saccade algorithm generates a saccade's start time, end time, duration, and starting and ending coordinates.

Visulization Techniques

Visualizations like scanpaths, heatmaps, and gaze plots are essential for visualizing eye tracking data. Eye tracking data visualization allows researchers and analysts to understand how individuals visually interact with various stimuli, such as images, websites, or scenes. It provides insights into where people look, the order in which they focus on different areas, and the duration of their fixations. Visualizations make it easier to communicate the results of eye tracking studies to a broader audience, including stakeholders, clients, or non-experts. A well-crafted visualization can convey complex information more effectively than raw data. Visualizations help in identifying recurring patterns in gaze behavior. Researchers can spot trends, common areas of interest, or anomalies in the data, leading to a better understanding of cognitive processes, user preferences, and potential usability issues. In usability testing and web design, visualizations help designers and developers assess where users focus their attention, helping optimize user interfaces for better user experiences.

1. Scanpath: A scanpath is a visual representation of the sequential order in which a person's gaze moves through a visual stimulus, as shown in Figure 3.11. It provides an insightful depiction of the path followed by the eyes as they transition between fixations and saccades, marking the areas of visual interest and the transitions between them. Scanpaths offer a valuable means to study and understand the flow of visual attention, allowing researchers to analyze how individuals process visual information, prioritize specific elements, and explore the intricacies of their cognitive processes. These sequences of fixations and saccades within a scanpath offer critical insights into the temporal dynamics of visual exploration and are fundamental in various fields, including eye tracking research, psychology, user experience design, and more.

During image scanning, individuals unknowingly fixate on certain locations and move to others, collectively forming a sequence of fixations and saccades known as a scanpath. These scanpath representations are characterized by numerical annotations denoting the order of fixations, typically depicted using a scanpath string. Each symbol in the scanpath string signifies either a fixation or dwell in an AOI (Area of Interest). For instance, the scanpath illustrated in Figure 3.11 can be represented as a scanpath string: 612345678. In this representation, each number within the circle corresponds to an AOI, and the sequence indicates the order of fixations. This detailed scanpath string unveils the temporal dynamics of visual exploration and finds applications in diverse fields such as eye tracking research, psychology, and user experience design.

2. Heatmap: In the realm of mental health monitoring, heat maps derived from eye tracking data offer valuable insights into the distribution of visual attention. These visualizations, as shown in Figure 3.12 play a pivotal role in understanding how individuals engage with various stimuli, such as therapeutic content, mental health assessment tools, or interventions. By providing a clear representation of where attention is concentrated, heat maps can assist mental health professionals and researchers in gauging the effectiveness of interventions or the impact of specific visual stimuli on patients' mental states. These visualizations help identify key areas of interest or "hotspots," shedding light on the content or elements that captivate a patient's attention during mental health assessments or therapeutic sessions. Consequently, mental health practitioners can tailor their approaches, optimizing the delivery of treatment and support based on the observed visual attention patterns. Heat maps in mental health monitoring thus serve as a powerful tool to enhance the understanding of patients' interactions with mental health resources and interventions, ultimately contributing to

Figure 3.12: Heatmap

more effective and personalized mental health care.

3. Area of Interest(AOI): Area of Interest are predefined subsections within a presented stimulus. These delineated regions serve as a valuable tool for measuring and comparing the performance or attention allocation to distinct areas within a video, image, website, or interface.

3.3.3 Multimodal Dataset Creation

A multimodal dataset has been created, incorporating diverse physiological measures, including ECG, GSR, PPG, respiratory signals, and eye tracking data. The subsequent sections provide the details of the data collection procedures for the generation of the multimodal dataset.

EmoRPhyE Dataset

EmoRPhyE (Emotion Recognition Using Physiological and Eye Tracking data) is a multimodal dataset curated to detect the emotional state of individuals as they view images with varying valence and arousal levels. Multimodal datasets include data from diverse physiological measures like ECG, PPG, GSR, respiratory signal, and eye tracking data. The data was collected from 30 students (14 males and 16 females) with an average age of 26 ± 5 years. The study utilized stimuli from the International Affective Picture System (IAPS) [162], a curated collection of emotionally charged images rated for arousal and valence. The demographic details are shown in Figure 3.13

Physiological signals were recorded using the CGX AIM physiological monitor, a portable system capturing raw ECG, respiratory, PPG, and GSR data at 500 Hz without filtering. The CGX AIM physiological monitor, positioned opposite the subject's dominant hand as shown in Figure 3.14, enables unrestricted movement

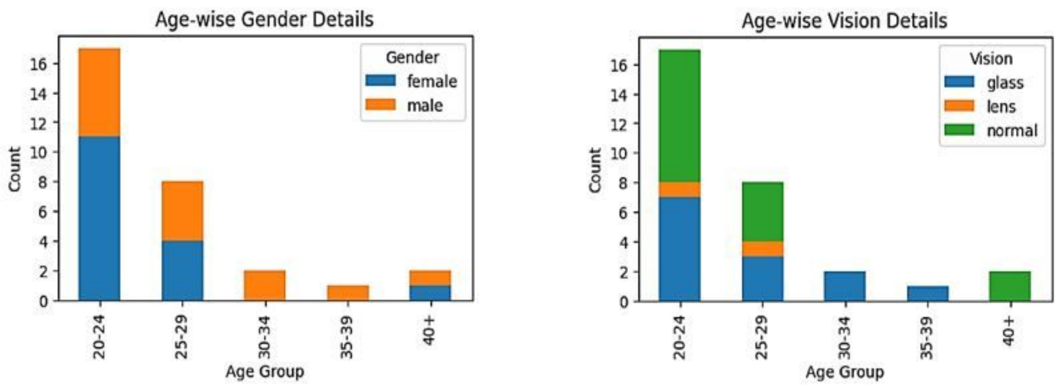


Figure 3.13: Demographic details of EmoRPhyE dataset

during data acquisition. This setup proves particularly beneficial for capturing signals like GSR and PPG, which necessitate electrode and sensor placement on the non-dominant hand. Bio-impedance-based pad sensors under the collarbones collected respiratory data and served for single-lead ECG recording. An additional ECG derivation used two adhesive electrodes on the upper left torso.

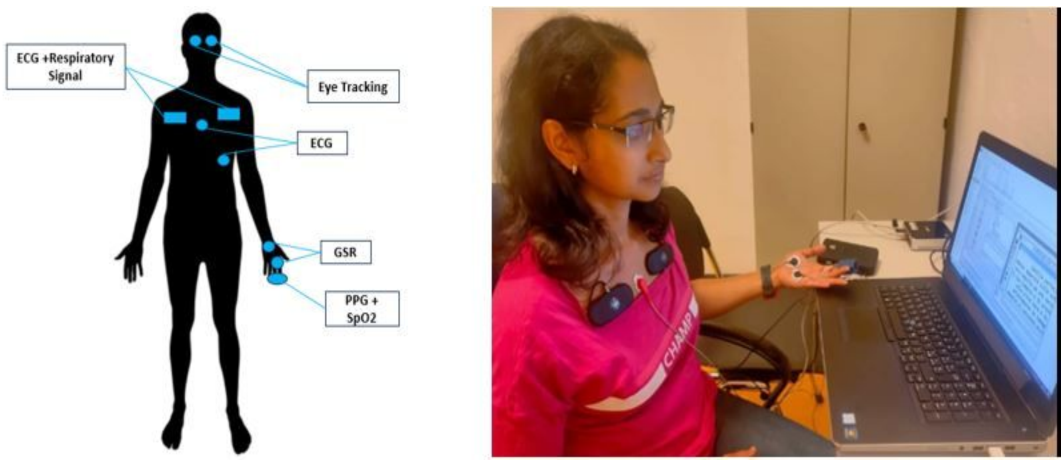


Figure 3.14: Placement of physiological sensors

PPG signal was obtained through a plethysmograph on the index finger of the non-dominant hand, where adhesive electrodes on the palm also recorded GSR. A strap electrode on the left ankle served as the common ground terminal for all sensors. Eye tracking data utilized the Tobii Pro X3-120 eye tracker, a remote system with binocular tracking capabilities. Operating at 120Hz for gaze data and 40Hz for pupil diameter, it effectively functioned within a range of 50 cm to 90 cm from the participant's eyes. The stimuli were presented to the participants using the E-prime version 3.0.3.80 software.

A total of 192 images were selected from the IAPS dataset for presentation. These images were organized into 48 groups, with each group containing four images sharing the same valence and arousal, as shown in Figure 3.15. After

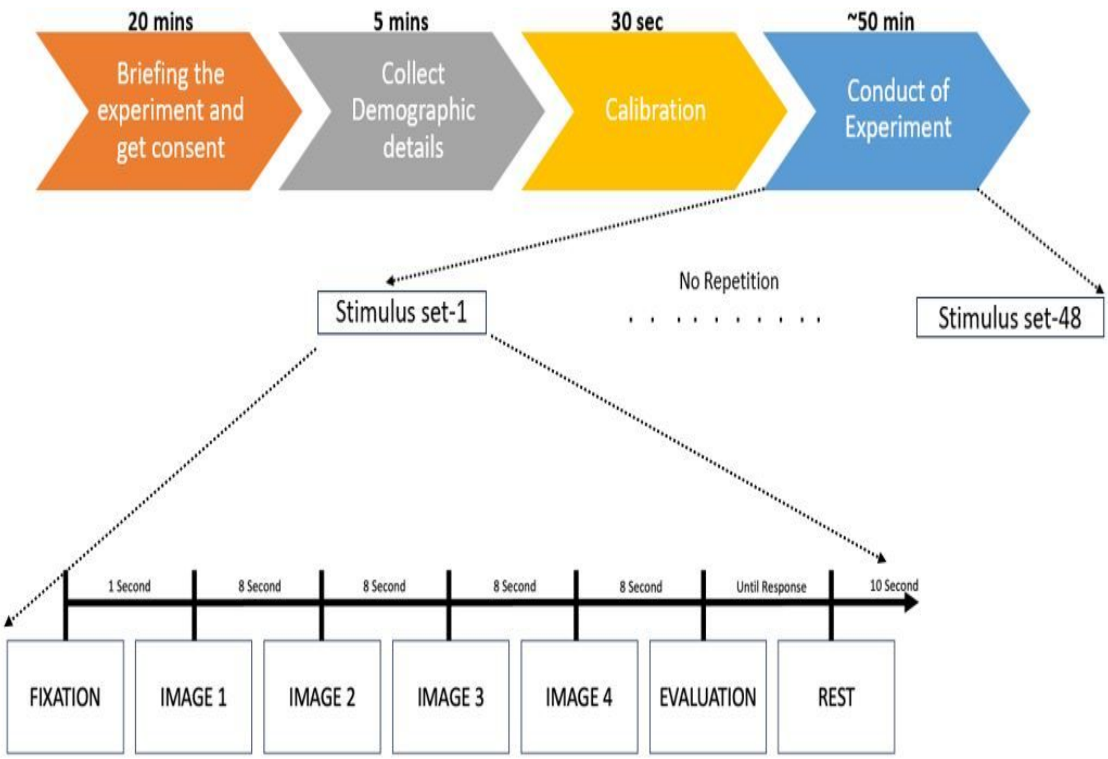


Figure 3.15: Experimental setup for EmoRPhyE data collection

displaying each set of four images, participants provided feedback on the valence and arousal using the Self-Assessment Manikin method (SAM) [163], rating on a scale of 1 to 9 as shown in Figure 3.16. Each image was presented for 8 seconds, followed by a 10-second blank screen. To avoid the repetition of emotional content, the presentation order of each group of images and the arrangement of images within each group were both randomized.

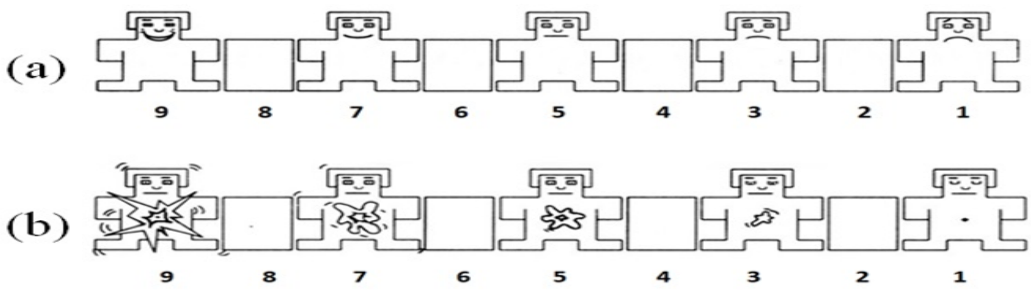


Figure 3.16: Rating scale based on the Self-Assessment Manikin (a) rating scale for valence and (b) rating scale for arousal

In the multimodal data collection process, synchronization was crucial to align physiological data, stimulus responses, and eye tracking data accurately. Utilizing the Lab Streaming Layer (LSL) time protocol ensured consistent timestamps between eye tracking data processed by E-prime and events managed by the E-prime software, such as stimulus presentation and subject responses. However, as physiological data was controlled by a separate laptop with its internal clock, synchronization was achieved through the LSL framework. This framework facilitated the exchange of synchronized data over the local network, addressing clock discrepancies between devices and ensuring precise alignment in this comprehensive, multimodal data acquisition. the synchronization scheme is shown Figure 3.17

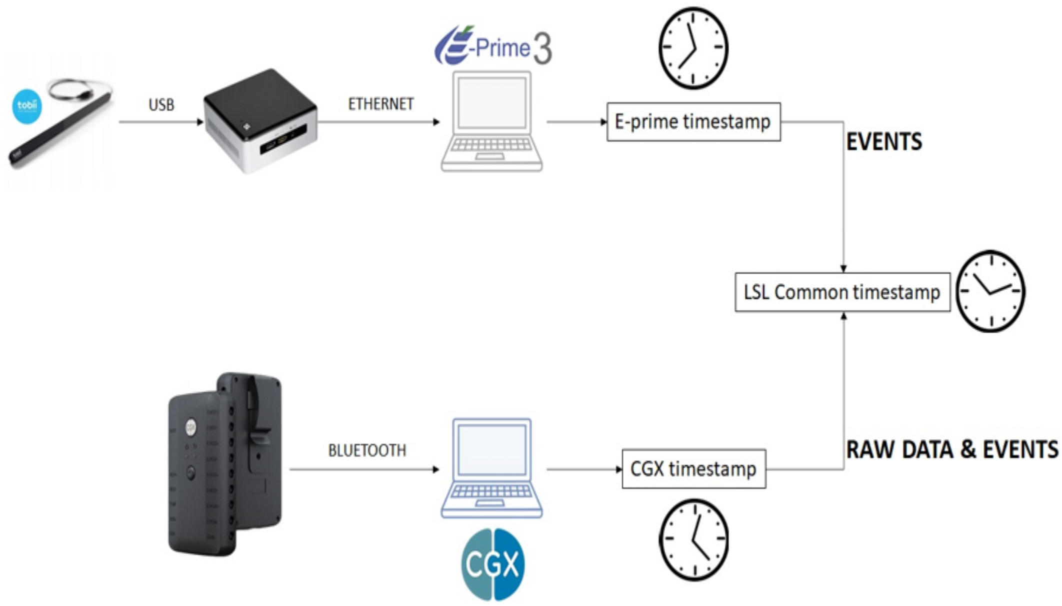


Figure 3.17: Synchronization scheme

EmoRPhyE contains raw data specific to each physiological signal. The ECG signal is a graphical representation of the electrical events occurring during the cardiac cycle. It is obtained by placing electrodes on the chest showing voltage changes reflecting heart muscle cell depolarization and repolarization. The ECG waveform consists of the P, QRS complex, and T waves. Two types of features, time and frequency domain, are commonly used in emotion assessment via ECG signals. Time domain features analyze the ECG signal's temporal patterns, while frequency domain features reveal information about energy distribution across frequency bands, though, in this study, only time domain features were employed due to the signal's short duration.

Peaks and valleys are identified from the raw respiratory signal. The breath rate series, extracted from the respiratory signal, is derived as the time differences between the peaks and valleys extracted for each window.

The raw EDA signal contains data on two distinct types of activity: tonic and phasic. The EDA signal contains information about tonic activity, representing gradual changes in skin conductivity due to factors like skin hydration and au-

Table 3.4: Raw eye gaze data from Tobi Pro Eye Tracker

Features	**Description**
RTTime	Timestamp
CursorX	Gaze X coordinate
CursorY	Gaze Y coordinate
PupilDiameterLeftEye	Left eye's Pupil diameter
PupilDiameterRightEye	Right eye's Pupil diameter
PupilValidityLeftEye	Left eye data validity indicator. True-1, False-0
PupilValidityRightEye	Right eye data validity indicator. True-1, False-0

tonomic regulation. Phasic activity in the EDA signal exhibits short-term peaks or fluctuations, reflecting rapid changes in skin conductivity caused by emotional responses and arousal mediated by the sympathetic nervous system.

The eye tracking data, collected using Tobii Pro X3-120 eye tracker, comprises two text files: one detailing the stimuli and participant feedback, and the other containing raw eye gaze data as shown in Table 3.4. This data is used as input for an emotion state detection model, which subsequently extracts features through various algorithms.

3.3.4 Summary of Datasets

Four eye tracking-based datasets, including one multimodal dataset aimed at understanding mental state parameters such as cognitive load, cognitive impairment, and emotional state, have been created. These datasets encompass raw eye gaze data (Tables 3.2 and 3.3, 3.4), as well as raw physiological signals, including ECG, PPG, GSR, and respiratory signals within the EmoRPhyE dataset. The subsequent feature extraction for each dataset is tailored to the specific application and is explained in Chapters 4 through 7. Figure 3.18 shows the raw data and the subsequent feature extraction from each dataset.

3.4 Software and the Hardware Used for the Study

3.4.1 Software

1. Experiment Center 3.7- stimulus presentation

2. BeGaze 3.7 - feature extraction

3. E Prime - stimulus presentation

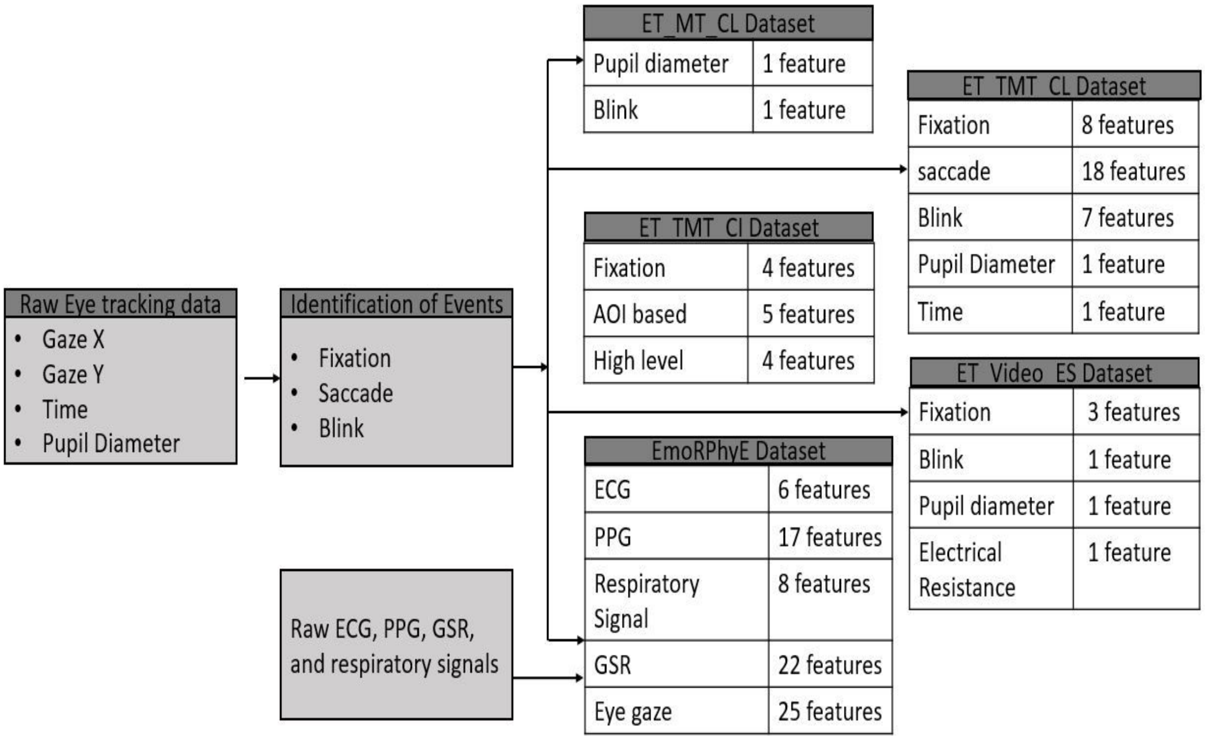


Figure 3.18: The raw eye tracking data and feature extraction in each dataset

3.4.2 Hardware

1. SMI REDn Eye Tracker with a sampling frequency of 60Hz.
2. VT3 Mini Eye tracker with a sampling frequency of 60Hz.
3. Tobii Pro X3 Eye tracker with a sampling frequency 120Hz for gaze and 40Hz for pupil diameter.
4. GSR sensor
5. necessary hardware and software required for collecting EEG, ECG, BVP, respiration rate, and GSR

This chapter has explained the first objective outlined in the study by elaborating on the creation of datasets aimed at addressing the scarcity identified in the literature. The following chapters will focus on the models developed as part of the second objective.

Chapter 4

Classification of Cognitive Load using ECL Models

4.1 Introduction to Cognitive Load

The total mental effort in a person's working memory is their cognitive load. The increase in the cognitive load of a person can have a negative impact on task completion. However, cognitive load is not the same for everyone. It varies from person to person based on age and gender. Cognitive load refers to the total amount of mental activity imposed on the working memory in an instant. Our working memory is very much limited. It will get overloaded while handling complex information. In addition, the problematic or confusing learning experience can cause cognitive load [44].

The mental state of an individual is linked to their cognitive load, which pertains to the mental effort and resources necessary for performing specific tasks or engaging in mental activities. Cognitive load can vary across a spectrum, influencing how a person feels and functions. This load is typically categorized into two primary states: low and high cognitive load, each associated with distinct psychological and functional aspects.

When cognitive load is low, individuals typically experience a sense of relaxation and ease. They often describe feeling comfortable and minimally stressed. Tasks requiring low cognitive load demand less mental effort and strain. People are more open to creative thinking, brainstorming, and exploring new ideas in this state. It's a mental state that encourages relaxation and creativity, conducive to problem-solving and innovative thinking.

On the other hand, high cognitive load tasks often result in increased stress and anxiety. Individuals may feel overwhelmed when the cognitive demands exceed their capacity to process information effectively. Creative thinking and problem-solving may suffer under the strain of high cognitive load as the mind becomes preoccupied with the complexity of the task. This state is marked by mental exhaustion and reduced decision-making capabilities, leading to decreased performance and an increased likelihood of errors.

It's essential to note that the impact of cognitive load on mental state can vary significantly from one person to another. Factors such as prior experience, expertise, and individual coping mechanisms play a crucial role in how cognitive load affects a person's mental state. Additionally, the duration of exposure to different cognitive load levels is significant; while short-term high cognitive load can be motivating, prolonged exposure to such stressors can lead to negative outcomes. Understanding one's cognitive load and developing strategies to manage it effectively is key to maintaining a healthy mental state and optimizing performance [164].

Monitoring cognitive load is of paramount importance, especially in industries where individuals are continually engaged in tasks over extended periods. High cognitive load can lead to performance reduction, stress, and errors, ultimately impacting an individual's mental health [44]. Machine learning (ML) techniques can be applied to extract meaningful patterns from this data [165], providing quantifiable assessments of cognitive load [44]. Eye tracking technology, offers valuable insights by correlating eye movements, such as fixations, saccades, pupil dilation, and blink with cognitive load [43]. These measurements are crucial for detecting cognitive load variations and potential mental fatigue, making eye tracking an essential tool for mental health monitoring and proactive intervention [99].

The study introduces two models, ECL-1 (Eye-Tracking Cognitive Load-1) and ECL-2 (Eye-Tracking Cognitive Load-2), developed for the detection and classification of cognitive load, a mental state parameter shedding light on the cognitive demands on an individual's mental faculties. These models offer valuable insights into understanding mental workload and the level of cognitive effort exerted by individuals. The following sections comprehensively explain the models, detailing their architectures, methodologies, and proficiency in discerning cognitive load. Section 4.2 introduces the ECL-1 model, detailing its methodologies and the statistical analysis conducted on this model. Subsequently, section 4.3 discusses the ECL-2 model, with subsection 4.3.1 focusing on its feature extraction process, specifically explaining the extraction of fixation, saccade, and blink-based features. Section 4.3.2 elaborates on the machine learning model employed, while section 4.3.3 delves into its performance evaluation. Additionally, section 4.3.4 examines data exploration and statistical analysis. Finally, section 4.4 concludes the research.

4.2 ECL-1 Model

An initial study was conducted with the ECL-1 model to understand the significance of certain features in the cognitive load classification. This was a preliminary investigation, which included a small population of 20 students. The objective was to identify the most important features contributing to cognitive load in a learning environment. This initial model helped to refine our methodology and identify key areas for further exploration.

The ECL-1 model, based on the ET_MT_CL dataset(explained in Chapter 3), is crucial for understanding cognitive load fluctuations, especially in simple and complex mathematical tasks. The model examined the effects of stressors on cognitive load in student participants. The raw data obtained from the dataset included timestamps, stimulus types, gaze details, and pupil diameter information. BeGaze 3.7 software was used to extract the features fixation duration, blink duration, pupil diameter from the raw data.

The experimental task entailed responding to a set of twenty mathematical questions characterized by diverse difficulty levels, classified based on the complexity of both operands and operators. Participants were allocated limited time to answer these questions, aiming to investigate the impact of cognitive load variations on stress levels. The utilization of mathematical stimuli with varying complexities served as a means to induce mental load in participants, and the subsequent analysis incorporated eye-tracking features to discern and understand the corresponding changes in cognitive load. Identifiable patterns in eye movement, pupil dilation, and blink are acknowledged as dependable signs of mental workload across all age groups [166,167], and this study focused on eye gaze measures, particularly pupil diameter and blink frequency.

The study compared stress levels and cognitive load between participants, Participant 1 and Participant 2. It was noted that Participant 2 experienced higher stress than Participant 1. The research findings corroborated existing literature, indicating a linear increase in pupil dilation with a higher working memory load.

The study observed how changes in cognitive load, induced by varying difficulty levels of mathematical questions, affected participants. Participant 1 displayed a minimal change in pupil diameter and blink frequency with simpler questions, but as question difficulty increased, both pupil diameter and cognitive load increased, leading to a reduction in blink frequency.

This experiment also allowed an assessment of each participant's expertise in solving mathematical problems. For Participant 1, the slight increase in pupil diameter and a steady number of blinks after an initial spike suggested increased cognitive load and active engagement in problem-solving, indicating expertise and interest in mathematics.

In contrast, Participant 2 exhibited a significant increase in pupil diameter and fluctuating blink frequency as question difficulty increased. This indicated an inability to concentrate, failure to answer within the allotted time, heightened stress, and a lack of expertise and interest in solving mathematical problems.

Statistical analysis, which focused on the association between paired samples, utilized a Pearson's product-moment correlation coefficient t-test to evaluate the significance of eye measurements across various difficulty levels of mathematical problems. The outcomes of this statistical analysis are presented in Table 4.1.

Table 4.1: ECL-1 model statistical analysis

Eye Measures	P value	Correlation	Remarks
Pupil Diameter	0.00001197	0.815154	Positive correlation
Blink Count	0.001385	-0.6647843	Negative Correlation

The statistical analysis showed a clear linear relationship between pupil diameter and cognitive load, which, in turn, indicates an increase in stress levels as cognitive load escalates. The positive correlation suggests that as the mathematical problems became more challenging, participants' pupils dilated more, reflecting their greater mental effort [49].

Conversely, there is a negative correlation between the number of blinks and cognitive load. This signifies that individuals tend to blink less frequently as cognitive load intensifies. It could be interpreted as an indication that higher cognitive loads lead to lapses in spontaneous blinking. The reduction in blink frequency, particularly when cognitive load is elevated, is an intriguing aspect that might signify a heightened state of concentration or cognitive strain during more complex problem-solving tasks. These findings provide valuable insights into how the eyes' physiological responses can shed light on the mental states of individuals when tackling tasks of varying complexity [49].

In the ECL-1 model, no specific computational models were used to classify cognitive load; rather, a statistical approach was taken to examine the significance of these eye gaze differences. The model's primary focus was to investigate the significance of differences in eye gaze features during mathematical tests, particularly when the difficulty level of the questions changed. This study exclusively utilized a statistical model to assess the significance of these differences. By employing t-tests, it was possible to efficiently demonstrate that there were indeed significant differences in the eye gaze features observed while participants engaged in mathematical problem-solving tasks.

4.3 ECL-2 Model

Based on the insights from ECL-1 model, ECL-2 model was designed to explore further the influence of various features on cognitive load in a larger population. This study aimed to extend the findings from ECL-1 model, contributing to a deeper understanding of cognitive load and its potential applications.

The second model, the ECL-2 model, is designed to classify mental states based on the cognitive load experienced by individuals based on the eye tracking data obtained from the ET_TMT_CL dataset. Utilizing eye tracking technology, this model provides an approach to understand the impact of cognitive load on an individual's mental state. The model extracted 35 features from the raw data obtained from ET_TMT_CL dataset and applied Random forest algorithm to classify cognitive load as low and high. A feature selection process is applied to enhance the model's accuracy, removing features with lower significance in classification and thereby refining the feature set used for cognitive load detection. This systematic approach ensures the model's effectiveness in assessing cognitive load based on eye-tracking data.

The TMT stimuli used for this study involve tasks with both simple and complex elements, and the progression from simpler to complex tasks within the TMT can induce cognitive load in participants. Figure 4.1 shows the ECL-2 model for cognitive detection. Participants were instructed to verbally announce visited numbers during the task to ensure adherence to traverse all numbers in ascending order without omissions. Failure to meet this requirement could elevate stress levels and increase cognitive load.

The following sections elaborate on the diverse features extracted from the ET_TMT_CL dataset, providing a comprehensive overview of the extraction methodologies employed. Furthermore, the machine learning algorithms used to implement the ECL-2 model are detailed, offering insights into the obtained results. The concluding segment encompasses a comparative analysis, highlighting the distinctions in performance between the ECL-2 model and existing models.

4.3.1 Feature Extraction

The model utilizes the dataset ET_TMT_CL and identifies the events fixation and saccade coordinates based on the IDT algorithm [161], explained in Chapter 3. The study was conducted on 100 participants. 35 eye gaze features were extracted for further analysis. The IDT algorithm was employed in the cognitive load detection application with a minimum fixation duration of 50 milliseconds and a maximum dispersion threshold of 100 pixels for fixation detection. Blink detection utilized a minimum duration threshold of 70 milliseconds. The model extracts derived

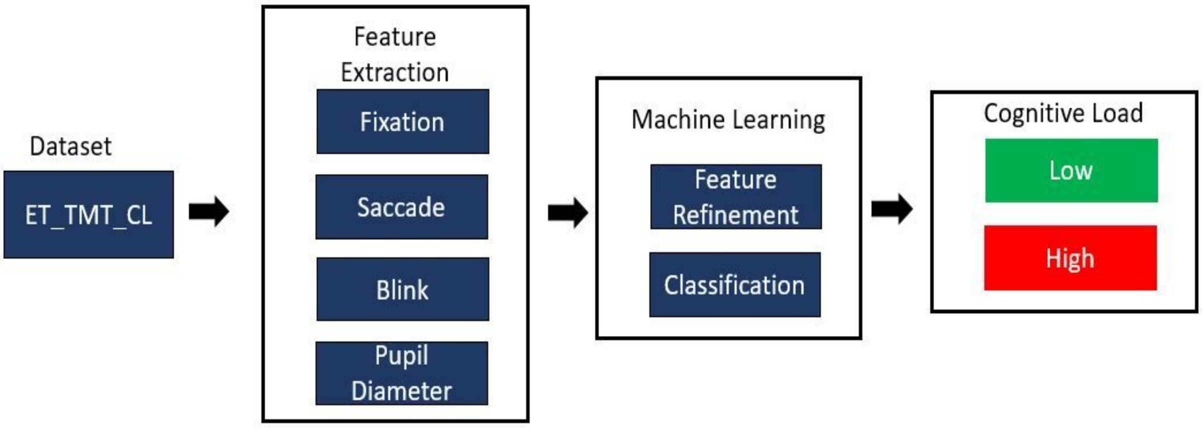


Figure 4.1: ECL-2 system model

features by leveraging these events, blink, and pupil diameter measurements.

In eye movement data analysis, fixations and blinks are mutually exclusive events. Fixation-related features at a specific time 't' exclude corresponding blink-related features, and vice versa. This distinction arises because blinks involve the temporary closure of the eyes, rendering fixations impossible during these brief periods.

So, based on the detected events like fixation coordinates and saccade coordinates and also based on the blink and pupil diameter, 35 features have been extracted. The extracted features and their descriptions are shown in Table 4.2.

The extracted features are:

Fixation based features:

Equations of fixation-based features are provided from Equation 4.1 to Equation 4.8.

first_dur indicates the first fixation duration.

$$\text{first\_dur} = \text{End Time of First Fixation} - \text{Start Time of First Fixation} \tag{4.1}$$

$$\text{fix\_freq} = \frac{\text{Total Number of Fixations}}{\text{Total Duration of Task in Seconds}} \tag{4.2}$$

$$\text{fix\_count} = \sum_{i=1}^{n} \text{Number of Fixations in Task } i \tag{4.3}$$

$$\text{time\_to\_first\_fix} = \text{Time at Onset of First Fixation} - \text{Start Time of Task} \tag{4.4}$$

Table 4.2: Eye tracking features extracted for the classification of cognitive load

Events	Feature Notations	Feature Description	Unit
Fixation	first_dur	First fixation duration	[ms]
	fix_freq	Fixation frequency	[count/s]
	fix_count	Fixation count	
	time_to_first_fix	Time to first fixation	[ms]
	min_fix_dur	Minimum fixation duration	[ms]
	max_fix_dur	Maximum fixation duration	[ms]
	tot_fix_dur	Total fixation duration	[ms]
	avg_fix_dur	Average fixation duration	[ms]
saccade	sac_freq	Saccade frequency	[count/s]
	count_sac	Saccade count	
	min_sac_dur	Minimum saccade duration	[ms]
	max_sac_dur	Maximum saccade duration	[ms]
	tot_sac_dur	Total saccade duration	[ms]
	avg_sac_dur	Average saccade duration	[ms]
	min_sac_amp	Minimum saccade amplitude	[°]
	max_sac_amp	Maximum saccade amplitude	[°]
	tot_sac_amp	Total saccade amplitude	[°]
	avg_sac_amp	Average saccade amplitude	[°]
	min_sac_vel	Minimum saccade velocity	[°/s]
	max_sac_vel	Maximum saccade velocity	[°/s]
	tot_sac_vel	Total saccade velocity	[°/s]
	avg_sac_vel	Average saccade velocity	[°/s]
	min_sac_lat	Minimum saccade latency	[°/s]
	max_sac_lat	Maximum saccade latency	[°/s]
	tot_sac_lat	Total saccade latency	[°/s]
	avg_sac_lat	Average saccade latency	[°/s]
Blink	blink_freq	Blink frequency	[count/s]
	blink_count	Blink count	
	time_to_first_blink	Time to first blink	[ms]
	min_blink_dur	Minimum blink duration(ms)	[ms]
	max_blink_dur	Maximum blink duration(ms)	[ms]
	tot_blink_dur	Total blink duration(ms)	[ms]
	avg_blink_dur	Average blink duration(ms)	[ms]
Pupil Diameter	avg_pupil	Average pupil diameter(mm)	[ms]
Time duration	tot_time	Time duration for entire task	[ms]

$$\text{min\_fix\_dur} = \min(\text{Duration of Fixation}_1, \ldots, \text{Duration of Fixation}_n) \tag{4.5}$$

$$\text{max\_fix\_dur} = \max(\text{Duration of Fixation}_1, \ldots, \text{Duration of Fixation}_n) \tag{4.6}$$

$$\text{tot\_fix\_dur} = \sum_{i=1}^{n} \text{Duration of Fixation}_i \tag{4.7}$$

$$\text{avg\_fix\_dur} = \frac{\sum_{i=1}^{n} \text{Duration of Fixation}_i}{\text{Number of Fixations in the Trial}} \tag{4.8}$$

Saccade based features:

Equations of saccade-based features are provided from Equation 4.9 to Equation 4.26.

$$\text{sac\_freq} = \frac{\text{Number of Saccades}}{\text{Total Duration of Task in Seconds}} \tag{4.9}$$

$$\text{count\_sac} = \sum_{i=1}^{n} \text{Number of Saccades in Trial } i \tag{4.10}$$

$$\text{min\_sac\_dur} = \min(\text{Duration of Saccade}_1, \ldots, \text{Duration of Saccade}_n) \tag{4.11}$$

$$\text{max\_sac\_dur} = \max(\text{Duration of Saccade}_1, \ldots, \text{Duration of Saccade}_n) \tag{4.12}$$

$$\text{tot\_sac\_dur} = \sum_{i=1}^{n} \text{Duration of Saccade}_i \tag{4.13}$$

$$\text{avg\_sac\_dur} = \frac{\sum_{i=1}^{n} \text{Duration of Saccade}_i}{\text{Number of Saccades in the Trial}} \tag{4.14}$$

$$\text{min\_sac\_amp} = \min(\text{Amplitude of Saccade}_1, \ldots, \text{Amplitude of Saccade}_n) \tag{4.15}$$

$$\text{max\_sac\_amp} = \max(\text{Amplitude of Saccade}_1, \ldots, \text{Amplitude of Saccade}_n) \tag{4.16}$$

$$\text{tot\_sac\_amp} = \sum_{i=1}^{n} \text{Amplitude of Saccade}_i \tag{4.17}$$

$$\text{avg\_sac\_amp} = \frac{\sum_{i=1}^{n} \text{Amplitude of Saccade}_i}{\text{Number of Saccades in the Trial}} \tag{4.18}$$

$$\text{min\_sac\_vel} = \min(\text{Velocity of Saccade}_1, \text{Velocity of Saccade}_2, \ldots, \text{Velocity of Saccade}_n) \tag{4.19}$$

$$\text{max\_sac\_vel} = \max(\text{Velocity of Saccade}_1, \text{Velocity of Saccade}_2, \ldots, \text{Velocity of Saccade}_n) \tag{4.20}$$

$$\text{tot\_sac\_vel} = \sum_{i=1}^{n} \text{Velocity of Saccade}_i \tag{4.21}$$

$$\text{avg\_sac\_vel} = \frac{\sum_{i=1}^{n} \text{Velocity of Saccade}_i}{\text{Number of Saccades in the Trial}} \tag{4.22}$$

$$\text{min\_sac\_lat} = \min(\text{Latency of Saccade}_1, \text{Latency of Saccade}_2, \ldots, \text{Latency of Saccade}_n) \tag{4.23}$$

$$\text{max\_sac\_lat} = \max(\text{Latency of Saccade}_1, \text{Latency of Saccade}_2, \ldots, \text{Latency of Saccade}_n) \tag{4.24}$$

$$\text{tot\_sac\_lat} = \sum_{i=1}^{n} \text{Latency of Saccade}_i \tag{4.25}$$

$$\text{avg\_sac\_lat} = \frac{\sum_{i=1}^{n} \text{Latency of Saccade}_i}{\text{Number of Saccades in the Trial}} \tag{4.26}$$

Blink based features:

Equations of blink based features are provided from Equation 4.27 to Equation 4.33.

$$\text{blink\_freq} = \frac{\text{Number of Blinks in the Trial}}{\text{Duration of the Trial in Seconds}} \tag{4.27}$$

$$\text{blink\_count} = \sum_{i=1}^{n} \text{Number of Blinks in Trial}_i \tag{4.28}$$

$$\text{time\_to\_first\_blink} = \text{Time at Onset of First Blink} - \text{Start Time of Trial} \tag{4.29}$$

$$\text{min\_blink\_dur} = \min(\text{Duration of Blink}_1, \text{Duration of Blink}_2, \ldots, \text{Duration of Blink}_n) \tag{4.30}$$

$$\text{max\_blink\_dur} = \max(\text{Duration of Blink}_1, \text{Duration of Blink}_2, \ldots, \text{Duration of Blink}_n) \tag{4.31}$$

$$\text{tot\_blink\_dur} = \sum_{i=1}^{n} \text{Duration of Blink}_i \tag{4.32}$$

$$\text{avg\_blink\_dur} = \frac{\sum_{i=1}^{n} \text{Duration of Blink}_i}{\text{Number of Blinks in the Trial}} \tag{4.33}$$

4. Pupil Diameter based feature: Equation of pupil diameter based feature is provided with Equation 4.34

$$\text{avg\_pupil} = \frac{\sum_{i=1}^{n} \text{Pupil Size}_i}{n} \tag{4.34}$$

5. Total Time Duration: Equation of Total time duration is provided with Equation 4.35

$$\text{tot\_time} = \text{End Time of Task} - \text{Start Time of Task} \tag{4.35}$$

The labeling of the dataset was carried out with the invaluable support of domain experts. The extracted features were assigned labels corresponding to the task's difficulty level. Data collected during the TMT tasks, specifically TMT A and B, were categorized into two distinct cognitive load levels: 'Low' for TMT A and B simple tasks and 'High' for TMT A and B complex tasks. This labeling was crucial for our analysis, as it allowed us to distinguish cognitive load levels in the dataset accurately.

Outlier detection is crucial for ensuring data quality and integrity, preventing the distortion of statistical analyses and model predictions. After feature extraction, we employed the Interquartile Range (IQR) method to identify and remove outliers from the dataset. The IQR, calculated as the difference between the third quartile (Q3) and the first quartile (Q1) of each feature (Equation (4.36)), served as the criterion. Data points exceeding the upper or lower threshold were deemed outliers and excluded from analysis. This process was executed separately for 'L' and 'H' classes, denoting low and high cognitive load. The resultant cleaned dataset, free of outliers, was then utilized for subsequent analysis.

$$\text{IQR} = Q3 - Q1 \tag{4.36}$$

Outliers were identified using the thresholds specified in the equations (4.37) and (4.38)

$$\text{Upper Threshold} = Q3 + 1.5 \times \text{IQR} \tag{4.37}$$

$$\text{Lower Threshold} = Q1 - 1.5 \times \text{IQR} \tag{4.38}$$

After removing outliers, Z-score normalization was applied to bring features to a common scale, optimizing machine learning algorithms. This involves transforming features with a mean of 0 and a standard deviation of 1 (Equation (4.39)). Z-score normalization scales and centers the data around zero, making it suitable for algorithms sensitive to feature distribution. This method retains information about data spread and ensures the effective contribution of all features to model training while referencing equations. (4.39).

$$\text{Zscore} = \frac{\text{Original Value} - \text{Mean}}{\text{Standard Deviation}} \tag{4.39}$$

4.3.2 Machine Learning Model

Upon extracting pertinent features from eye-tracking data, the Random Forest algorithm was applied to predict mental states based on cognitive load, effectively distinguishing between low and high cognitive load. This ensemble learning approach leveraged the **35 eye-tracking features**, providing accurate and interpretable classifications of cognitive load states.

In cognitive load detection based on 35 eye-tracking features, the Random Forest algorithm, comprising essential components, plays a pivotal role. Initially, the algorithm strategically selects a subset of features at each split point, mitigating overfitting and enhancing model precision. Subsequently, it constructs multiple decision trees from diverse data subsets, improving the model's overall accuracy by capturing various data features. The algorithm then integrates predictions from these distinct decision trees to yield a final prediction. The Random Forest algorithm is trained on the ET_MT_CL dataset. Once trained, it predicts individual mental states based on unique eye tracking data, providing accurate and interpretable classifications of low and high cognitive load states.

Feature importance analysis is crucial in machine learning, providing insights into the relative importance of individual features, especially in complex algorithms like Random Forest or Gradient Boosting, which lack inherent interpretability. In the mental state prediction model for cognitive load, initially considering 35 features yielded a 90% accuracy. However, feature importance analysis, illustrated in Figure 4.2, offered valuable insights into the relevance of these features, enhancing the understanding of the model's predictions. These plots serve as indispensable tools, helping discern the contribution of features, facilitating feature selection, and improving model interpretability. This information is vital for effectively communicating results in cognitive load assessment.

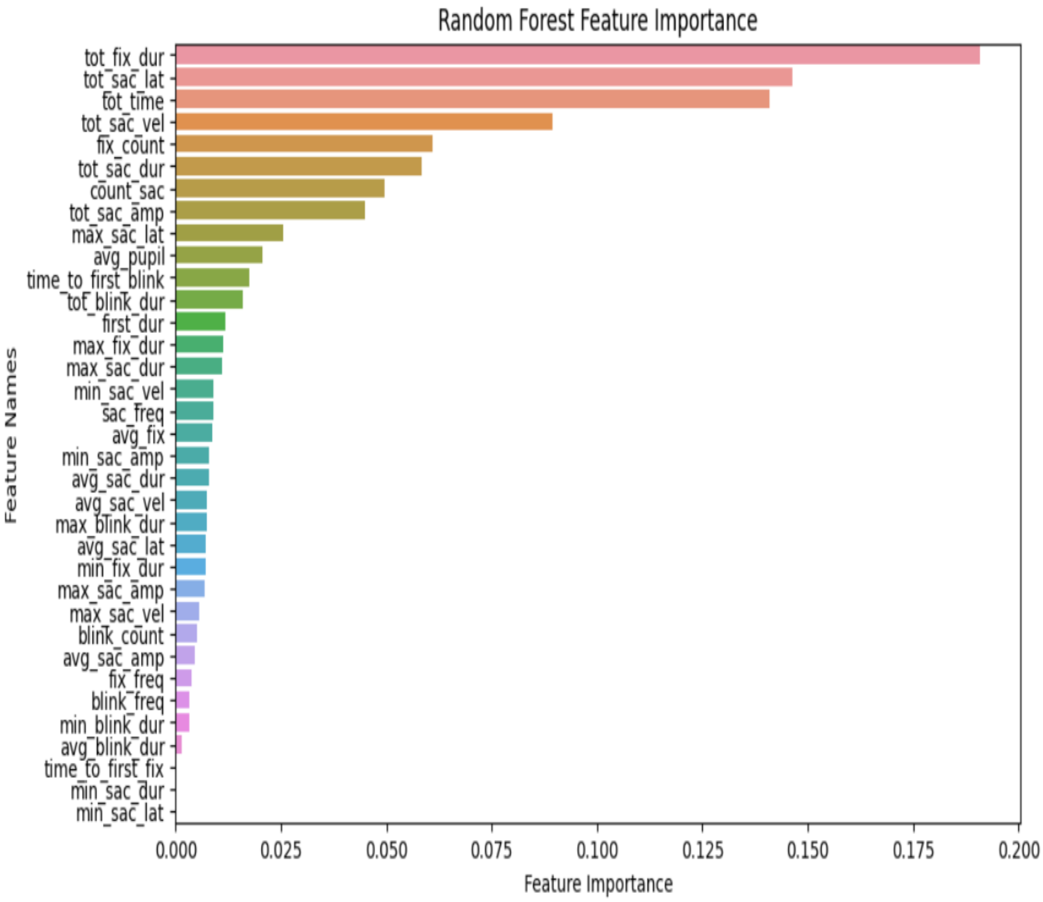


Figure 4.2: Feature importance plot based on Random Forest algorithm

4.3.3 Performance Evaluation of ECL-2 Model

The ECL-2 model, utilizing the Random Forest algorithm for classifying cognitive load as low and high, underwent a comparative analysis with various machine learning algorithms. These included Logistic Regression(LR), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN), and Naive Bayes (NB).

Random Forest is an ensemble learning method which is used for both classification and regression problems. It is based on the concept of bagging, which is a type of ensemble learning technique. Random Forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

The Logistic Regression algorithm is crucial for binary classification tasks, where it effectively models the probability of an instance belonging to a specific category. This is particularly advantageous in scenarios with categorical dependent variables. The logistic function, integral to its inference, transforms a linear combination of features into a range between 0 and 1, representing probabilities. The formula for Logistic Regression is expressed as (4.40)

$$P(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + \ldots + b_n X_n)}} \tag{4.40}$$

Support Vector Machine (SVM) proves invaluable in high-dimensional spaces, demonstrating versatility in handling both linear and non-linear classification challenges. SVM's inference involves seeking the hyperplane that best separates data points into distinct classes, expressed by the formula (4.41)

$$f(x) = \langle w, x \rangle + b \tag{4.41}$$

Decision Trees are renowned for their intuitive interpretation and ability to capture complex relationships in data, particularly robustness to outliers. Decision Trees' inference involves recursively splitting the data based on features, with decision rules at each node determined by conditions on features.

K-Nearest Neighbors (KNN), effective in scenarios where the decision boundary is not well-defined, relies on the similarity between instances. The class of a data point is determined by the majority class among its k nearest neighbors.

Naive Bayes, a probabilistic classifier based on Bayes' theorem, offers computational efficiency and efficacy across various applications. The inference for Naive Bayes is expressed through Bayes' theorem (4.42)

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \tag{4.42}$$

These algorithms play distinct roles in classifying cognitive load based on 35 eye gaze features, each leveraging its unique characteristics for effective model performance. Each model underwent a rigorous evaluation employing a comprehensive set of performance metrics to assess their effectiveness in cognitive load classification. The comparison of each model based on the performance measures is shown in Table 4.3

The accuracy, precision, recall, F1 Score used in the performance analysis of the ECL-2 model are calculated based on the formula (4.43) to (4.46), respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \times 100 \tag{4.43}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{4.44}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{4.45}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{4.46}$$

Table 4.3: Performance evaluation of ECL-2 model for the classification of cognitive load

ML Algorithms	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F1 Score (%)**
LR	90	88	91	89
RF	90	88	91	89
SVM	88	88	88	88
DT	88	90	84	87
KNN	88	88	88	88
NB	87	85	88	86

Notably, among the various models, it became evident that Random Forest consistently outperformed the others, demonstrating superior accuracy, precision, recall, and F1 score in distinguishing between low and high cognitive load. This marked superiority in Random Forest's performance underscores its effectiveness in accurately predicting mental states based on cognitive load levels, particularly in distinguishing between low and high cognitive load, utilizing the extracted eye-tracking features. All models' confusion matrices are presented in Figure 4.3, visually representing their classification performance.

In refining the mental state prediction model for cognitive load, thoughtful feature removal based on importance analysis significantly enhanced model performance. Careful consideration was given to optimizing comprehension, reducing overfitting risks, and enhancing training efficiency. The removal of less critical features, such as "min_sac_lat", "min_sac_due", and "time_to_first_fix" resulted in a streamlined model with improved accuracy, achieving 94%. Notably, by utilizing feature importance, ECL-2 achieved an impressive accuracy of 94%, and the model's efficiency was further highlighted as it utilized only 32 features instead of the original 35. This reduction in features underscores the significance of feature importance analysis in streamlining the model while maintaining high accuracy, emphasizing its pivotal role in optimizing the cognitive load classification process.

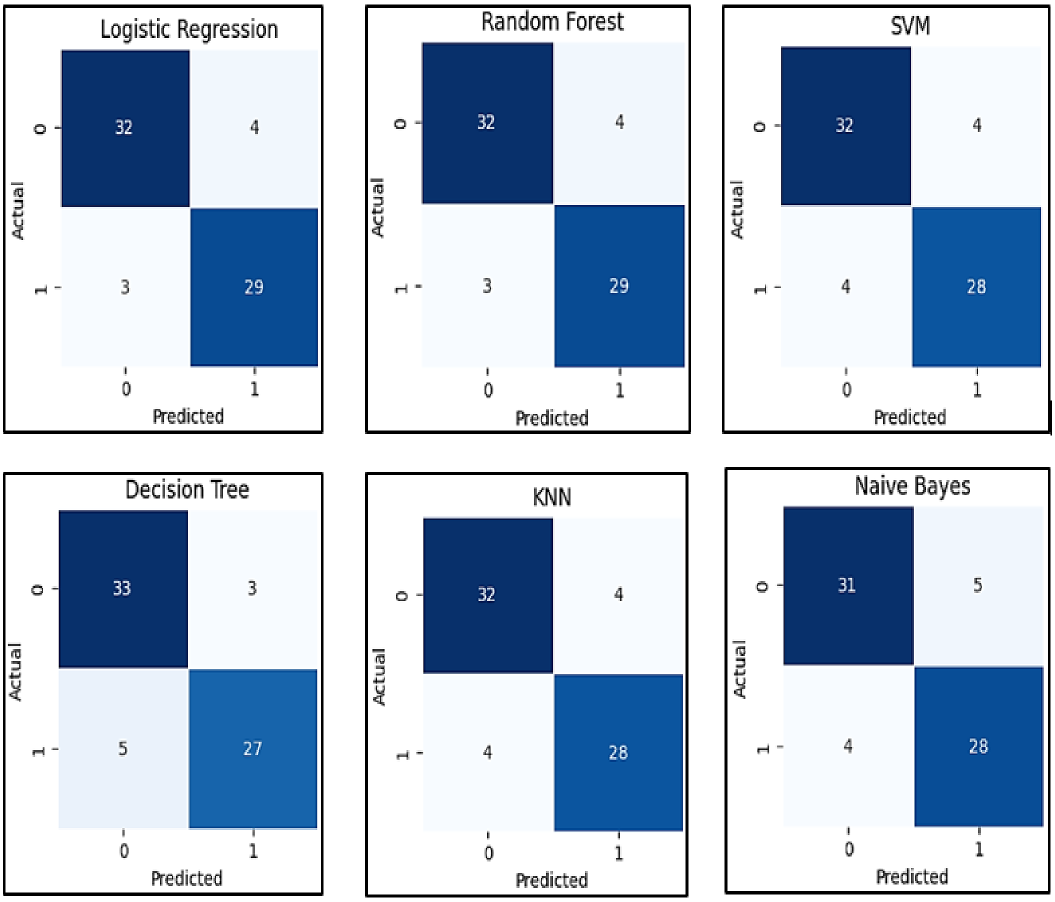


Figure 4.3: Confusion matrix for machine learning algorithm comparison in cognitive load classification

The detailed performance analysis, as presented in Table 4.4, solidifies the efficacy of the ECL-2 model in comparison to alternative machine learning approaches. Feature importance analysis and removal were pivotal in achieving a more accurate and interpretable model.

4.3.4 Data Exploration and Statistical Analysis

A comprehensive statistical analysis was performed to understand the disparities between data collected during the simple and complex tasks of TMT. This analysis included three key steps:

1. **Boxplot Analysis:** Boxplot analysis visually represents data distribution for each cognitive load level, providing insights into central tendency, spread, and potential outliers within each group. Figure 4.4 illustrates essential features that play a crucial role in highlighting substantial differences between 'Low' and 'High' cognitive load conditions. These visual representations offer a clear understanding of variations in central tendency and data spread within each condition, pivotal for comprehending the impact of cognitive load on the investigation.

Table 4.4: Analysis based on feature importance using Random Forest

Number of Features	**Features removed**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
35	NIL	90	88	91	89
32	**min_sac_lat min_sac_dur time_to_first_fix**	**94**	**94**	**94**	**94**
30	min_sac_lat min_sac_dur time_to_first_fix blink_freq avg_blink_dur	93	91	94	92
29	min_sac_lat min_sac_dur time_to_first_fix blink_freq avg_blink_dur blink_count	90	88	91	89
29	min_sac_lat min_sac_dur time_to_first_fix blink_freq avg_blink_dur max_blink_dur	91	91	91	91

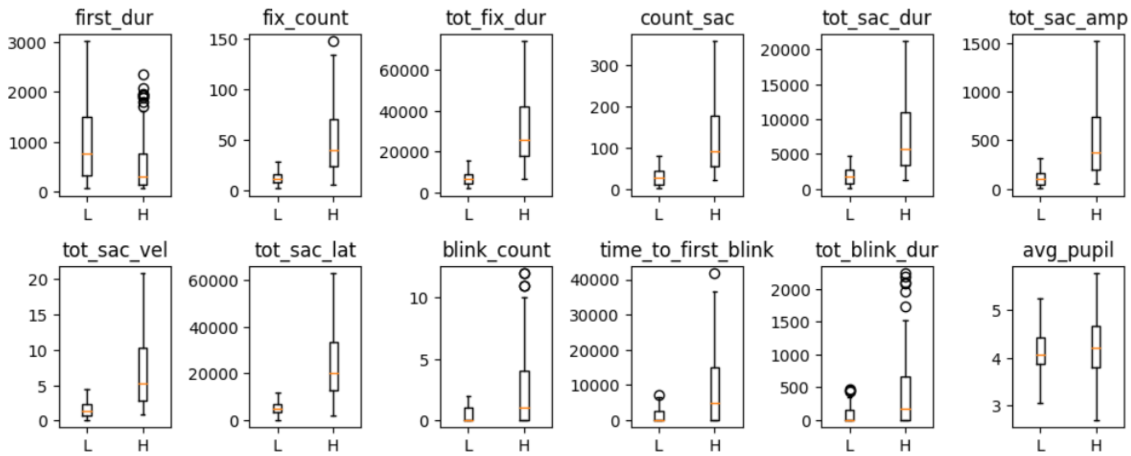


Figure 4.4: Boxplot analysis. ’L’: cognitive load low, ’H’: cognitive load high

A significant increase in cognitive load is observed during the complex TMT task compared to the low cognitive load condition. This insight sheds light on the dynamics of cognitive resource allocation during task execution. The demanding nature of the complex TMT task, requiring cognitive functions like working memory and sustained attention, leads to heightened cognitive engagement. Individuals allocate more mental resources to meet these demands, resulting in a noticeable upswing in cognitive load, as evidenced by the boxplot analysis. This increase in cognitive load is central to the research, directly influencing various facets of task performance.

2. **Welch Two-Sample t-Test:** Following boxplot analysis, the Welch two-sample t-test assessed 35 eye gaze features, chosen for comparing 'TMT' simple and complex tasks with potentially unequal variances or sample sizes (Table 4.5). The primary aim was to determine whether the means of 'Low' and 'High' cognitive load groups exhibited significant differences, testing the null hypothesis of no substantial impact of cognitive load on these features.

The null hypothesis (H0) was established as follows: there is no significant difference between the eye gaze features when comparing the 'Low' and 'High' cognitive load groups. This hypothesis suggested that cognitive load levels did not substantially influence these features.

The obtained p-values were predominantly below 0.001, signaling significant differences in most eye gaze features between 'Low' and 'High' cognitive load groups. Rejecting the null hypothesis, the findings suggest a notable difference in eye gaze features during TMT simple and complex tasks. These results underscore the increasing cognitive load during the complex TMT task, emphasizing the importance of these eye gaze features as indicators of cognitive workload in this research.

3. **one-way analysis of variance (ANOVA) Test:** To validate our findings and explore differences across groups, a one-way analysis of variance (ANOVA) test was applied (Table 4.6). Significant results suggest differences in cognitive load levels between 'TMT' simple and complex tasks. Many eye gaze features exhibited highly significant p-values below 0.001, 0.01, or 0.05, affirming statistical significance in the ANOVA test.

The F-statistic in an ANOVA test determines if there are significant differences between group means. A high F-statistic indicates significant differences between group means, leading to rejection of the null hypothesis; a low F-statistic suggests insignificant differences and retention of the null hypothesis. Based on the values obtained and displayed in Table 4.6, significant differences in features are confirmed.

The Eye gaze features analyzed with ANOVA displayed significant differences between 'TMT' simple and complex tasks, affirming distinct cognitive load levels. Variances in fixation, saccade, blink, pupil, and total time behaviors underscore the dynamic cognitive resource allocation during these tasks, providing valuable insights into cognitive load dynamics.

The feature importance plot as shown in Figure 4.2 indicated that certain fea-

Table 4.5: Comparison of cognitive load classification results using Welch two sample t-test

Feature	Cognitive Load:Low Mean(SD)	Cognitive Load:High Mean(SD)	P-value
first_dur	1074.13 (987.2)	1077.24 (1734.21)	0.98
fix_freq	1.74 (0.92)	1.58 (0.88)	0.1
fix_count	13.32 (7.16)	51.92 (38.43)	**p<0.001**
time_to_first_fix	40.32 (102.25)	103.06 (972.33)	0.4
min_fix_dur	154.8 (196.67)	87.11 (111.52)	**p<0.001**
max_fix_dur	2053.06 (1235.09)	3797.22 (2383.98)	**p<0.001**
tot_fix_dur	7531.67 (3485.73)	30977.81 (17475.26)	**p<0.001**
avg_fix	745.09 (540.01)	844.31 (690.3)	0.14
sac_freq	3.5 (1.73)	3.59 (1.58)	0.62
count_sac	29.52 (18.56)	124.36 (91.73)	**p<0.001**
min_sac_dur	49.42 (2.82)	49.51 (2.54)	0.74
max_sac_dur	96.58 (8.55)	99.71 (1.28)	**p<0.001**
tot_sac_dur	1837.49 (1140.08)	7850.27 (5802.17)	**p<0.001**
avg_sac_dur	62.98 (5.86)	63.09 (2.64)	0.83
min_sac_amp	0.73 (0.52)	0.43 (0.3)	**p<0.001**
max_sac_amp	16.18 (8.63)	20.21 (8.25)	**p<0.001**
tot_sac_amp	107.69 (70.89)	503.13 (400.91)	**p<0.001**
avg_sac_amp	3.75 (1.22)	3.87 (0.75)	0.26
min_sac_vel	0.01 (0.01)	0.01 (0.01)	p<0.001
max_sac_vel	0.19 (0.09)	0.24 (0.08)	**p<0.001**
tot_sac_vel	1.57 (1.01)	7.31 (5.7)	**p<0.001**
avg_sac_vel	0.05 (0.01)	0.06 (0.01)	0.05
min_sac_lat	16.83 (1.8)	16.64 (0.0)	0.16
max_sac_lat	1373.66 (950.62)	3401.42 (3534.74)	**p<0.001**
tot_sac_lat	5310.2 (3275.84)	24536.6 (14679.49)	**p<0.001**
avg_sac_lat	229.16 (174.52)	283.27 (243.74)	**p<0.05**
blink_freq	0.11 (0.23)	0.11 (0.18)	0.99
blink_count	0.87 (1.64)	4.11 (7.01)	**p<0.001**
time_to_first_blink	1686.7 (2672.08)	10163.42 (12476.15)	**p<0.001**
min_blink_dur	62.23 (117.94)	92.03 (132.71)	**p<0.05**
max_blink_dur	92.57 (169.76)	533.21 (2426.41)	**p<0.05**
tot_blink_dur	150.4 (319.33)	1060.71 (3072.21)	**p<0.001**
avg_blink_dur	73.31 (127.3)	212.93 (650.6)	**p<0.01**
avg_pupil	4.09 (0.64)	4.3 (0.75)	**p<0.01**
tot_time	8290.22 (3537.59)	34390.36 (18471.89)	**p<0.001**

tures were less significant in determining the model's performance. Upon removing these features and reclassifying, the model demonstrated improved accuracy, as shown in Table 4.4. Subsequent Welch two-sample t-tests and ANOVA tests revealed that many eye gaze features exhibited significant differences, with p-values under 0.05. Interestingly, some features that initially displayed lower significance

Table 4.6: Comparison of cognitive load classification results using one-way ANOVA test

Feature	F-statistic	P-value
first_dur	0	0.98
fix_freq	2.75	0.1
fix_count	165.7	**p<0.001**
time_to_first_fix	0.7	0.4
min_fix_dur	15.17	**p<0.001**
max_fix_dur	71.65	**p<0.001**
tot_fix_dur	294.24	**p<0.001**
avg_fix	2.17	0.14
sac_freq	0.25	0.62
count_sac	174.51	**p<0.001**
min_sac_dur	0.11	0.74
max_sac_dur	22.2	**p<0.001**
tot_sac_dur	175.74	**p<0.001**
avg_sac_dur	0.05	0.83
min_sac_amp	42.54	**p<0.001**
max_sac_amp	19.32	**p<0.001**
tot_sac_amp	160.36	**p<0.001**
avg_sac_amp	1.29	0.26
min_sac_vel	37.74	**p<0.001**
max_sac_vel	23.91	**p<0.001**
tot_sac_vel	167.4	**p<0.001**
avg_sac_vel	3.72	0.05
min_sac_lat	2	0.16
max_sac_lat	52.15	**p<0.001**
tot_sac_lat	277.72	**p<0.001**
avg_sac_lat	5.53	**p<0.05**
blink_freq	0	0.99
blink_count	34.32	**p<0.001**
time_to_first_blink	75.02	**p<0.001**
min_blink_dur	4.78	**p<0.05**
max_blink_dur	5.58	**p<0.05**
tot_blink_dur	14.77	**p<0.001**
avg_blink_dur	7.54	**p<0.01**
avg_pupil	7.97	**p<0.01**
tot_time	327.33	**p<0.001**

in the feature importance plot and were subsequently removed for classification did not show a significant difference in these statistical analyses. This suggests their removal allowed other features to play a more prominent role in the model's classification accuracy.

Table 4.7: Comparison with the existing CL detection models

Model	Stimuli	Features	Algorithm	Observation
[40]	Military aviation simulator	Pupil dilation, fixation and saccade count	ANOVA, t- test	Estimating pilot's cognitive load
[168]	Driving simulator	pupil size, blink rate and fixation time.	Statistical analysis	Detection of cognitive load of drivers
[42]	Coding problems	Fixation, Saccades, blinks	NB, RF MLP, SVM, KNN, LR, Decision Tree	Prediction of stressful technical interview settings
[41]	Difficult and easy mental calculations	inter-trial and intra-trial changes in pupil diameter	ANCOVA	Observed significant difference in eye measures during easy and difficult tasks.
ECL-2	TMT	35 features based on fixation, saccade, blink, pupil diameter and time	RF	Classified into low and high CL

4.4 Discussion

A comparison of the ECL-2 model with the existing CL detection models is shown in Table 4.7. The ECL-2 model, designed to detect cognitive load, particularly focused on the stimuli from the Trail Making Test (TMT) tasks, sets itself apart from existing models that have utilized various stimuli like military aviation simulators, driving simulators, coding problems, and mental calculations. With the capacity to extract 35 features based on fixation, saccade, blink, pupil diameter, and time, the ECL-2 model employs the Random Forest algorithm for classification, distinguishing it from other models that use a range of algorithms. Moreover, while the ECL-2 model classifies subjects into low and high cognitive load categories, other models have different objectives, such as estimating a pilot's cognitive load, detecting the cognitive load of drivers, predicting stressful technical interview settings, and observing significant differences in eye measures during easy and difficult tasks.

4.5 Conclusion

The ECL models offer robust frameworks for assessing the impact of cognitive load on an individual's mental state. ECL-1 model focuses on mathematical task complexity, revealing correlations between eye gaze features like pupil diameter and blink frequency with cognitive load. This approach enhances our comprehension of cognitive load dynamics, shedding light on the intricate relationship between task complexity and cognitive load variations.

The ECL-2 model successfully classifies low and high cognitive load states, holding significant implications for psychology and cognitive science. Eye-tracking data offers a detailed understanding of how cognitive load influences mental states, from relaxation to heightened stress. The systematic approach to data collection and feature extraction, combined with machine learning algorithms, enhances accuracy in predicting cognitive load levels. Using the ET_TMT_CL dataset with TMT tasks as stimuli, participants' vocal feedback ensured task compliance, with deviations leading to increased cognitive load. Applying various machine learning algorithms, the Random Forest consistently outperformed others, accurately distinguishing between low and high cognitive load based on extracted eye-tracking features. The Welch two-sample t-tests and ANOVA tests revealed significant differences among eye gaze features, contributing to a better understanding of their role in cognitive load classification.

The ECL models enhance our understanding of the dynamic relationship between cognitive load and cognitive states. Its application promises to optimize performance and cognitive assessment in various contexts, offering valuable insights into the impact of cognitive load on mental well-being and efficiency.

This chapter focused on classifying mental state parameters using eye tracking measures, specifically addressing cognitive load using ECL-1 and ECL-2 models. The next chapter will focus on another mental state parameter, cognitive impairment.

Chapter 5

Classification of Cognitive Impairment using ETMT Model

5.1 Introduction to Cognitive Impairment

In the modern era, people's lives have undergone significant transformations, leading to challenges in memory recall, comprehension of new information, retention of details, sustained attention, and sound judgment, all of which can impact their daily functioning [14]. These subtle alterations in cognitive functions can profoundly affect an individual's behavior. When individuals struggle with memory, learning, and focus on their tasks, it may indicate a decline in cognitive abilities or cognitive impairment [169].

Cognitive impairment is recognized as a highly expensive condition, factoring in the costs associated with medications and nursing care facilities [105]. Unfortunately, cognitive impairment is considered incurable [45, 170]. Nevertheless, its progression can be slowed down through timely diagnosis and appropriate care. There is a substantial increase in the number of people affected by this condition [107], making the rapid growth of dementia cases a significant public health concern. Early detection and prompt intervention can effectively mitigate the advancement of cognitive impairment [106].

While cognitive impairment is often observed in individuals over the age of 65, it is not confined to a specific age group. Other risk factors include a family history of the condition, brain injuries, exposure to harmful substances, and brain irradiation [108]. Additionally, education level, as well as the presence of other medical conditions, can contribute to cognitive impairment. Some medications, vitamin deficiencies, depression, and various health issues can also lead to mild cognitive impairment (MCI) [171]. The spectrum of cognitive impairment ranges from mild to severe, with the transitional stage from subtle cognitive abnormalities to early-stage dementia termed as MCI [172, 173]. Individuals with mild cognitive impairments exhibit slight changes in their cognitive functions but can still manage their daily activities. In contrast, severe impairment can result in an inability to communicate effectively, understand the significance of things, and live independently. Cognitive impairment can affect various aspects, including mental flexibility, concentration, visual attention, and focused attention.

Traditional dementia detection relies on tests like MMSE, MoCA, and TMT, known for accuracy but criticized for being time-consuming and stressful, especially for patients with writing challenges [35, 109, 111]. Eye tracking technology, recognized for assessing cognitive and neurological conditions [174], offers a less intrusive method for monitoring eye movements, providing a promising approach for evaluating cognitive decline [175].

ETMT model leverages fuzzy inference systems to assess an individual's mental state based on their cognitive impairment. The study's primary objective is to extract **novel high-level features** from eye-tracking data, using fuzzy rules to detect various deficits contributing to cognitive impairments. The model goes beyond by offering detailed scores that shed light on visual search speed, focused attention, and an overall **cognitive impairment score**. These scores provide valuable insights into an individual's mental state.

Acknowledging its significant role in identifying cognitive issues, we propose the development of the Eye Tracking-Based Trail Making Test (ETMT) as a screening tool to support healthcare professionals [108]. This enables extracting a broader range of features, promising more comprehensive inferences in detecting early signs of cognitive decline and guiding targeted interventions.

The ETMT model's architecture is outlined in section 5.2. It further elaborates on the feature extraction process in section 5.2.1, the visual search speed fuzzy inference system in section 5.2.3, the focused attention fuzzy inference system in section 5.2.4, and the adaptive neuro-fuzzy inference system (ANFIS) in section 5.2.5. The results of the model's operation are analyzed in section 5.3, followed by a discussion of the model's various achievements in comparison to state-of-the-art models in section 5.4. Finally, the conclusion is presented in section 5.5.

5.2 ETMT Model Architecture

The ETMT model's system architecture is depicted in Figure 5.1. Eye-tracking dataset ET_TMT_CI is used for the ETMT model, and the data collection procedure was explained in chapter 3. The raw eye tracking data collected during the experiments underwent comprehensive data analysis. Gaze features were meticulously extracted from the eye tracking data, shedding light on participants' gaze behavior, attentional focus, and visual search speed. These features provided quantifiable insights into cognitive processes that traditional assessments could not easily capture. The extracted features were then fed into the ETMT model to produce three scores: visual search speed, focused attention, and overall cognitive impairment. Two separate fuzzy inference systems, FIS-1 and FIS-2 were employed to calculate visual search speed and focused attention scores, while an Adaptive Neuro-Fuzzy Inference System (ANFIS) based on all the extracted features determined the overall cognitive impairment score.

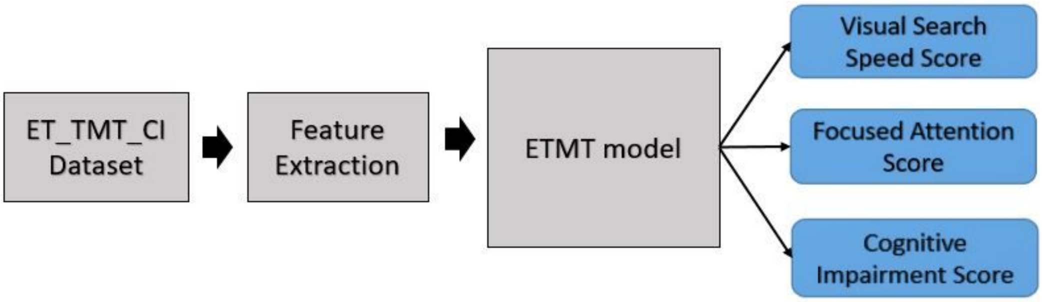


Figure 5.1: System architecture of ETMT model.

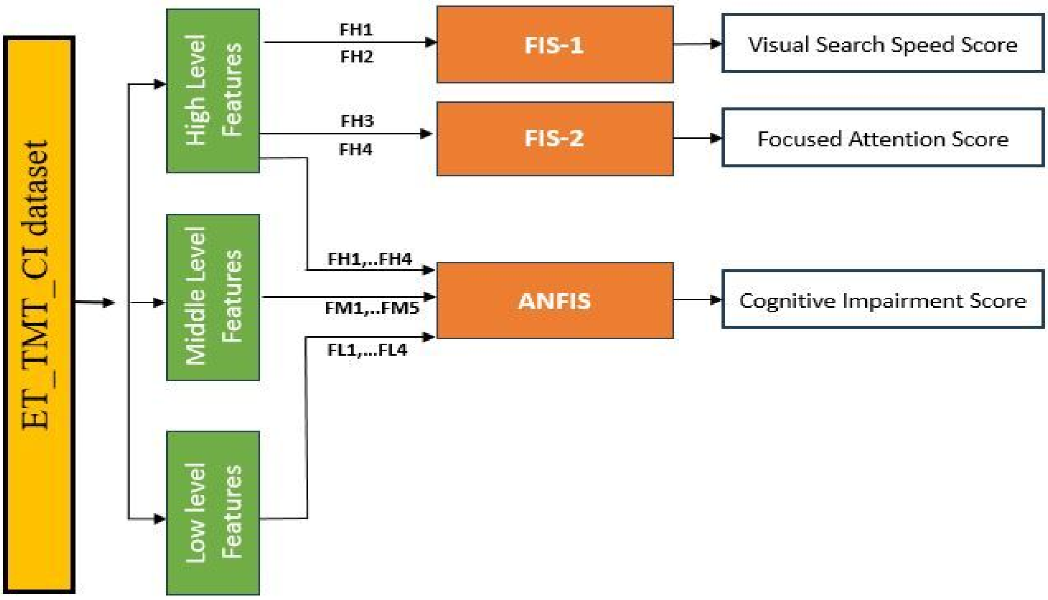


Figure 5.2: Detailed architecture of ETMT model.

The cognitive impairment score provides insight into a patient's overall cognitive functioning. However, the visual search speed and focused attention scores obtained through our ETMT model can offer psychologists a clear understanding of the patient's cognitive processing and enable tailored treatment approaches. These scores serve as sensitive markers, pinpointing the exact cognitive deficit of the patient. Different cognitive impairments may manifest different types of deficits, and the detailed scores can facilitate early diagnosis and guide psychologists to initiate appropriate treatment that helps delay the advancement of the disease. The detailed architecture of the ETMT model is shown in Figure 5.2. The following section explains each module of the ETMT model.

5.2.1 Feature Extraction

Raw data in the ET_TMT_CI included timestamps, stimulus types, gaze details, and pupil diameter information collected from 31 participants. The Low-level and middle-level features were extracted using the BeGaze 3.7 software and are listed in Table 5.1. Low-level features primarily focused on fixation-based metrics (FL1 to FL4). Middle-level features (FM1 to FM5) were extracted based on user-defined Areas of Interest (AOIs) corresponding to each number and alphabet in the stimulus. AOIs are sub-regions used to extract specific metrics. A total of 13 features were extracted from the ET_TMT_CI dataset.

Table 5.1: Low-level, Middle-level and High-level features extracted from ET_TMT_CI dataset

Type	Feature	Description
Low-level	Fixation Count(FL1)	Total number of fixation points
Low-level	Fixation Time [ms](FL2)	Sum of the time duration of each fixation point
Low-level	Fixation Time [%](FL3)	Percentage of fixation time with respect to total time
Low-level	Fixation Duration Average[ms](FL4)	Average of all the fixation durations concerning the trial
Middle-level	Dwell time(FM1)	Duration of gaze on a specific AOI
Middle-level	Dwell Time [%](FM2)	Percentage of Dwell time with respect to total time
Middle-level	Glance(FM3)	Count of fixation points in an AOI
Middle-level	Revisits(FM4)	Count of repeated gazes at a specific spot or AOI.
Middle-level	First Fixation duration[ms](FM5)	Time duration of first fixation in an AOI
High-Level	Scanpath Score(FH1)	Score based on how closely each participant's eye movement pattern aligns with the expected scanpath
High-Level	Total Time(FH2)	Total completion time
High-Level	Error Rate(FH3)	Rate of mistakes during TMT task
High-Level	Inattentional Blindness(FH4)	Fail to perceive unexpected stimuli or events

The middle-level features are calculated based on the formula 5.1 to 5.5. The middle-level feature dwell time(FM1) is a crucial metric in eye tracking, representing the cumulative time spent looking at a specific AOI. It encompasses the total duration of fixations and saccades within the AOI [176]. The dwell time serves as a reliable indicator of interest, with longer duration exceeding 500ms suggesting a higher level of interest in the AOI, while shorter duration below 100 ms suggest limited engagement and processing. The feature FM2, dwell time percentage, is the ratio of time spent on a specific area of interest to the total observation time, expressed as a percentage. Glance (FM3) is the total count of fixation points within an Area of Interest (AOI), representing the number of times the eyes fixate on that specific region. Revisits(FM4) quantify how often a participant returns their gaze to a specific area, revealing sustained interest and highlighting significant focal points in the observed content or environment. The first fixation duration(FM5) represents the time spent looking at a specific Area of Interest (AOI) during the initial gaze, providing insight into what initially captures attention in a scene.

$$\text{Dwell Time} = \sum \text{Time(Fixation Points in AOI)} \tag{5.1}$$

$$\text{Dwell Time \%} = \left(\frac{\sum \text{Time(Fixation Points in AOI)}}{\text{Total Time}} \right) \times 100\% \tag{5.2}$$

$$\text{Glance} = \sum \text{fixation points in AOI} \tag{5.3}$$

$$\text{Revisits} = \sum \text{revisits to the AOI} \tag{5.4}$$

$$\text{First fixation duration} = \text{time(first fixation in AOI)} \tag{5.5}$$

High-level features (FH1 to FH4) were derived from low-level and middle-level features using specific algorithms detailed in Algorithms 5.1, 5.2, and 5.3. The feature scanpath score (FH1) was extracted based on Algorithm 5.1, which generates a scanpath string. A scanpath is a visualization that shows the sequence of fixations and saccades in the order in which it is visited. The scanpath string specifies the sequence of fixations visited in order. The Levenshtein distance was calculated between each participant's scanpath string, and the expected scanpath string [14]. The expected scanpath string is the expected order of viewing the AOIs in the stimulus. The calculated Levenshtein distance is considered the scanpath score [177]. The overall amount of time required to complete the entire task is the feature total time (FH2). The feature error rate (FH3) is the count of missed targets. While viewing the AOIs in the specified order, there can be fixations outside the AOIs and wrong fixations. So, the total count of those errors indicates the error rate. Inattentional blindness (FH4) occurs when a person fails to notice something completely visible to them [178]. It indicates the attentional abilities and working memory of a person [179]. Lower working memory is observed in people with inattentional blindness [180].

Algorithm 5.1 Scanpath String Generation

```
Define the AOIs
Get the fixations
i ← 1
j ← 1
N1 ← count_fixation
N2 ← count_AOIs
while i ≠N1 do
    while j ≠N2 do
        if Fixation falls within AOI then
            Print AOI Name
        end if
    end while
end while
```

Algorithm 5.2 Error Rate

```
Define the AOIs
Get the fixations
error_rate ← 1
i ← 1
j ← 1
N1 ← count_fixation
N2 ← count_AOIs
while i ≠N2 do
    while j ≠N1 do
        if Fixation not falls within AOI then
            error_rate+=1
        end if
    end while
end while
Print error_rate
```

5.2.2 Visual Search Speed Fuzzy Inference System (FIS-1)

The specific score on the visual search speed of the participant can give inference to the exact deficit of those participants. The TMT follows a visually guided task where the participants need to visually attend each AOI in a specific order

Algorithm 5.3 Inattentional Blindness

Require: Scanpath String
Ensure: Inattentional Blindness Score

```
function CalculateInattentionalBlindness(scanpath)
    count ← 0
    i ← 0
    N ← length(scanpath)
    while i < N do
        if repetition of a pattern in the scanpath starting from index i then
            count ← count + 1
            i ← i + length(pattern)                 ▷ Skip the repeated pattern
        else
            i ← i + 1
        end if
    end while
    return count
end function
```

[181]. The visual search pattern of an individual may differ from each other [182]. The visual search speed score indicates the efficiency in visual search and speed of completion. Gaze patterns of an individual are the indicators of their visual behavior [183,184]. An individual's cognitive state and activities can be discovered based on their visual behavior [185].

The visual search speed fuzzy inference system(FIS) considers the high-level features FH1 and FH2 as the inputs and generates the score based on the generated rules [186]. The FH1 is generated based on the comparison with the scanpath, which followed the correct path in completing the task [187]. FH2 is the time required to complete the entire task. The completion time indicates the speed and focuses on identifying each AOI. The Mamdani model for visual search speed FIS is shown in Figure 5.3. The parameters of visual search speed FIS are given in Table 5.2.

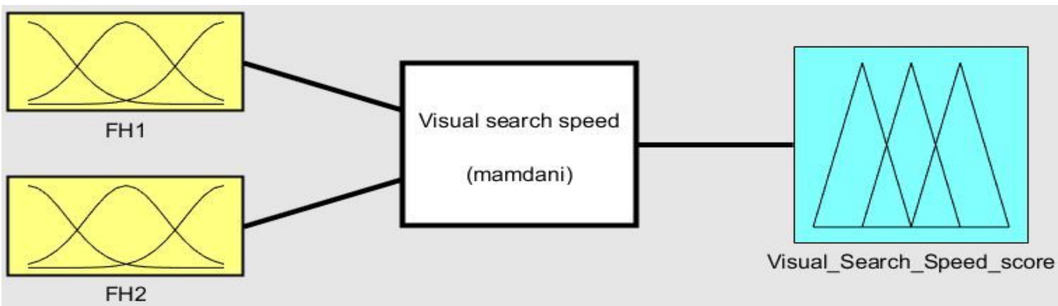


Figure 5.3: Visual Search Speed Fuzzy Inference System (FIS-1), submodule within the ETMT Model.

By utilizing this visual search speed score, the fuzzy inference system can pro-

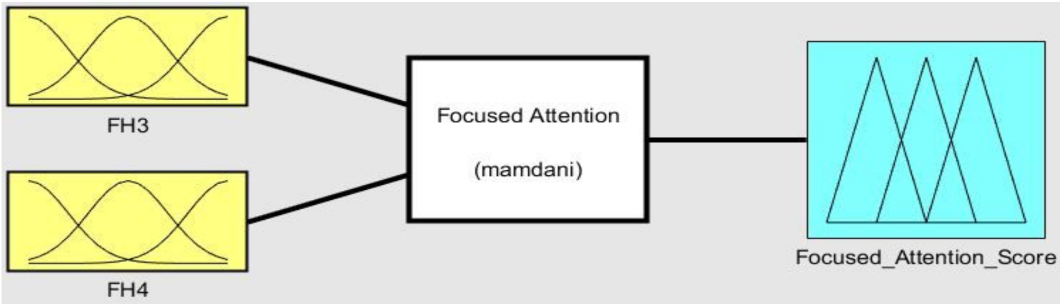


Figure 5.4: Focused Attention Fuzzy Inference system (FIS-2), submodule within the ETMT Model.

vide valuable insights into various deficits, particularly those related to visual attention and processing speed. An individual's visual attentional efficiency and effectiveness may be reflected in their visual search speed score. The person with deficits in visual attention may exhibit slower scanpath scores (FH1) compared to the reference path [188]. This could indicate difficulties in properly attending to and processing relevant visual stimuli within the task. Processing speed refers to how individuals perceive, analyze, and respond to visual information. In the fuzzy inference system context, people with processing speed deficits might exhibit longer total completion times (FH2) than those without deficits. A slower completion time indicates difficulty in processing visual information effectively and quickly enough to make decisions. Visual Search Speed FIS provides an indication of deficits in processing speed and visual attention impairments.

5.2.3 Focused Attention Fuzzy Inference System (FIS-2)

Focused attention, the ability to concentrate on a task without distractions, is a critical cognitive function. On the other hand, sustained attention involves maintaining a consistent behavioral response during an ongoing activity. The Focused Attention Fuzzy Inference System (FIS) utilizes high-level features FH3 and FH4 as inputs to generate a score based on predefined rules. Specifically, FH3, which represents the error rate, and FH4, indicative of inattentional blindness, play key roles in detecting focused attention. Figure 5.4 depicts the Mamdani model for the Focused Attention FIS, while Table 5.2 provides the parameters.

The Focused Attention FIS's output score holds valuable information concerning various deficits, including motor impairment, challenges in attentional disengagement, memory deficits, neuropsychological impairments, and issues related to executive functioning. An elevated error rate, one of the input parameters, indirectly suggests motor impairment as it reflects difficulties in executing motor tasks accurately.

Table 5.2: Parameters of Visual Search Speed FIS and Focused Attention FIS.

Parameter	Description
Fuzzy structure	Mamdani
Membership function	Trapezoidal
Number of membership functions for each input	3
Number of inputs	2
Number of outputs	1
Rules generated	9

Attentional disengagement, which pertains to the ability to shift focus from one task or stimulus to another, can be linked to higher levels of inattentional blindness. This may signify a deficit in attentional disengagement, where individuals struggle to divert their focus from the primary task to attend to unexpected stimuli.

Memory deficits, associated with challenges in storing and retrieving information, can be inferred from the Focused Attention FIS. Higher error rates and increased inattentional blindness might indicate difficulties in managing task-related information, indicating potential memory deficits. Neuropsychological impairment, a spectrum of cognitive deficits resulting from neurological conditions or brain injuries, can be suggested by a lower focused attention score, reflecting performance on the task in terms of error rate and inattentional blindness inputs.

Executive functioning, encompassing cognitive processes responsible for goal-directed behaviors like planning, problem-solving, and cognitive flexibility, can also be indirectly assessed using the Focused Attention FIS. Elevated error rates and increased inattentional blindness may indicate challenges in effective executive functioning. In summary, the Focused Attention FIS provides insights into motor impairment, attentional disengagement, memory deficits, neuropsychological impairments, and executive functioning [189].

5.2.4 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The overall cognitive impairment score was generated by considering all extracted features. The combined low-level, middle-level, and high-level features underwent clustering using the K-Means algorithm with k set to 3. This clustering process involved using all features from the 31 participants as input, grouping the samples into three clusters based on data similarities. Subsequently, these clusters were assigned labels based on expert knowledge.

The next step involved inputting the labeled clustered data into the ANFIS. ANFIS is a hybrid model that combines the strengths of an adaptive artificial neural network (ANN) and an FIS [190, 191]. This hybrid nature offers advantages

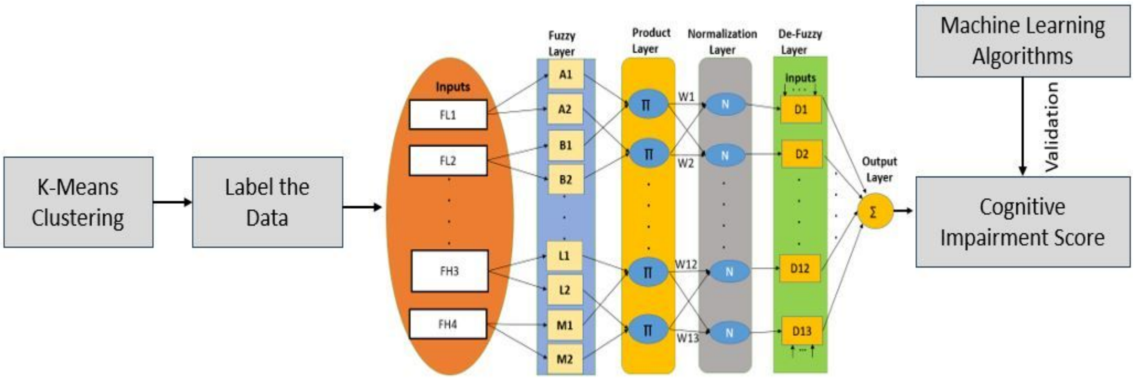


Figure 5.5: Architecture of ANFIS for overall cognitive impairment score, sub-module within ETMT Model.

from both approaches. ANFIS follows the Takagi-Sugeno fuzzy model, generating fuzzy IF-THEN rules based on input-output relationships. It employs a hybrid learning algorithm that merges the backpropagation technique with the least squares approach.

ANFIS comprises five layers: the fuzzy layer for fuzzifying inputs, the product layer to calculate rule firing strength through multiplication, the normalization layer for normalizing fuzzy strengths, the de-fuzzy layer to perform defuzzification of inputs, and the output layer, a single node that sums incoming signals.

The detection of the overall cognitive impairment score involved providing the ANFIS model with the labeled low-level, middle-level, and high-level features from all 31 participants as input. The ANFIS architecture for this purpose is illustrated in Figure 5.5, and the training model's parameters are detailed in Table 5.3. [190, 191]

Table 5.3: Parameters of ANFIS for Overall Cognitive Impairment Score.

Parameter	Description
Fuzzy structure	Sugeno
Membership function	Gaussian
Number of membership function	3
Number of inputs	13
Number of outputs	1
Optimization Method	Hybrid
Training number of epoch	30
Training samples	75%
Testing samples	25%

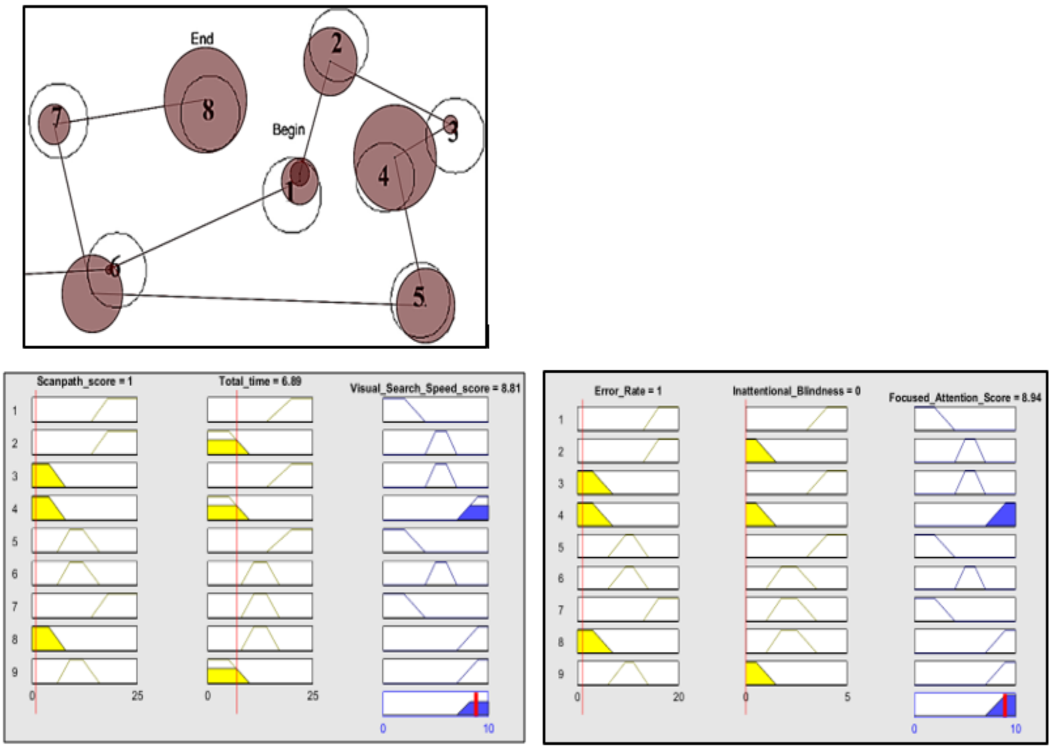


Figure 5.6: (a) Scanpath; (b) Visual Search Speed FIS; (c) Focused Attention FIS of
Participant 1 with High Visual Search Speed score and Focused Attention score.

5.3 Result Analysis

ETMT serves as a screening tool for assessing an individual's mental state, primarily focusing on their cognitive impairment. The system generates scores based on various cognitive parameters, providing valuable insights into the individual's cognitive well-being. In the results analysis, we first examine the ETMT's capacity to generate scores by extracting novel features from eye-tracking data, offering valuable insights into cognitive impairment. Subsequently, we compare the ET_TMT_CI dataset with traditional TMT datasets, assessing its potential as an effective screening tool for cognitive health assessment. The following subsection, 5.3.1, focuses on the analysis of the generation of scores based on the cognitive impairment of a person, and 5.3.2 focuses on the comparison of the ET_MT_CI dataset with the traditional paper-pencil TMT dataset

5.3.1 Generation of Scores

The ETMT model employs two fuzzy inference systems, FIS-1 and FIS-2, to generate individual scores for visual search speed and focused attention. Subsequently, an overall cognitive impairment score is computed through an ANFIS model, con-

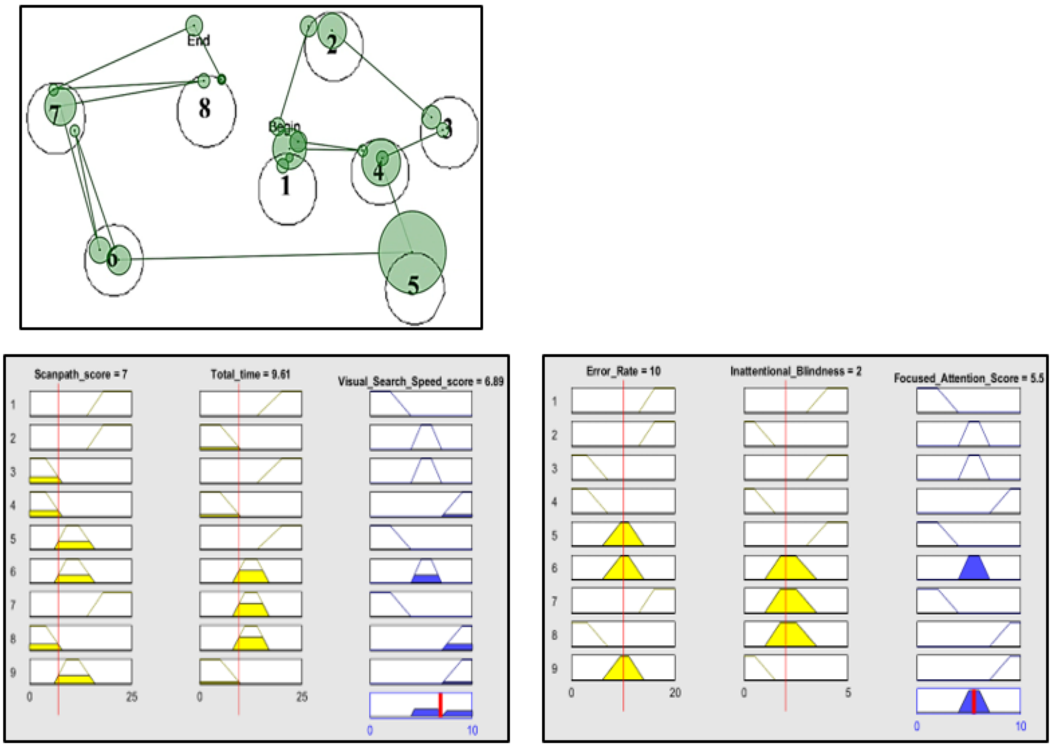


Figure 5.7: (a) Scanpath (b) Visual Search Speed FIS (c) Focused Attention FIS of
Participant 2 with Medium Visual Search Speed score and Focused Attention score.

sidering all the relevant extracted features. The fuzzy inference systems are measured on a scale of 1 to 10, and the ranges were categorized as scores from 1 to 4 as low, 5 to 7 as medium, and 8 to 10 as high. Since it is fuzzy based system there is a degree of overlap at the boundaries.

The FIS-1, with the features, scanpath score and total time, assesses visual search speed, offering insights into processing speed and visual attention impairments. Meanwhile, the FIS-2, relying on the features error rate and intentional blindness, evaluates focused attention, providing a detailed understanding of deficits in motor function, attentional disengagement, memory, neuropsychological function, and executive functioning.

Examining the results for Participant 1, it was observed that they received a low cognitive impairment score. This score indicates strong cognitive health. Participant 1 demonstrated quick visual search speed and excellent focused attention, with a lower error rate and a complete absence of inattentional blindness. As illustrated in Figure 5.6, this outcome signifies their sound cognitive state [192].

In contrast, Participant 2's results, displayed in Figure 5.7, indicate a moderate cognitive impairment score. Participants with such scores often exhibit average visual search speed and reasonable, focused attention. In this case, Participant 2 displayed a medium error rate and signs of inattentional blindness. Notably, the participant repeatedly visited certain areas (AOIs 6 and 7) without detecting them, a characteristic of inattentional blindness. Individuals with more pronounced cognitive impairments may be particularly susceptible to inattentional blindness due to difficulties maintaining attention and processing unexpected stimuli even within their visual field. These impairments can stem from deficits in various cognitive functions, including attention, working memory, and executive functions.

The overall cognitive impairment score is generated by the ANFIS model, which considers all extracted features. A split of 25% for testing and 75% for model training was used. The ANFIS model underwent training for a range of epochs, from 20 to 50, with the analysis revealing that an epoch of 30 produced a lower Root Mean Square Error (RMSE). The error statistics pertaining to the ANFIS model can be found in Table 5.4.

Table 5.4: Error statistics of ANFIS model for cognitive impairment score.

Error Statistics	Epoch = 20	Epoch = 30	Epoch = 40	Epoch = 50
RMSE	0.6702	0.3581	0.6504	0.8041
Error Mean	-0.0851	0.1262	-0.0988	0.1835
Error STD	0.7107	0.3583	0.6873	0.8369

Each individual receives a comprehensive evaluation comprising a detailed visual search speed score, a focused attention score, an overall cognitive impairment score, and specific indications of deficits. Figure 5.8 illustrates the comprehensive score for Participant 1. The accompanying scanpath, created during the TMT task, is represented by the string '6 1 2 3 4 5 6 7 8,' resulting in a scanpath score of 1. This scanpath, characterized by a solitary task error and the absence of inattentional blindness, signifies a high level of focused attention. Such characteristics contribute to elevated visual search speed and focused attention scores, ultimately yielding a lower cognitive impairment score. Further scrutiny of the visual search speed score reveals that Participant 1 exhibits minimal deficits in visual attention and processing speed. Meanwhile, the focused attention score provides insight into potential deficits in motor skills, attentional disengagement, memory, neuropsychological functions, and executive functioning. In Participant 1's case, these deficits are minimal, suggesting limited impairment.

Turning to Figure 5.9, it presents the score for Participant 2. This participant demonstrates moderate cognitive impairment, reflected in their visual search speed and focused attention scores. Participant 2's scanpath exhibits an irregular and disorganized pattern compared to that of Participant 1, resulting in a lower visual search speed. Multiple task errors and potential instances of inattentional blindness indicate a reduced level of focused attention compared to Participant 1.

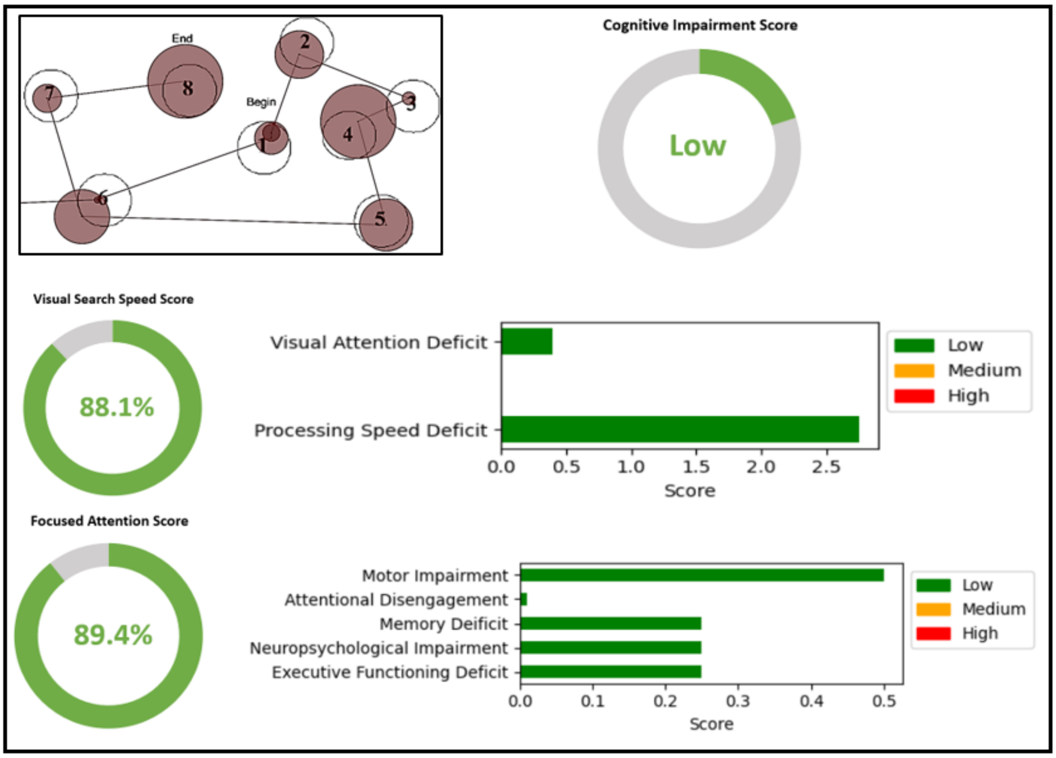


Figure 5.8: Detailed score generated by ETMT model for Participant 1 with low cognitive impairment.

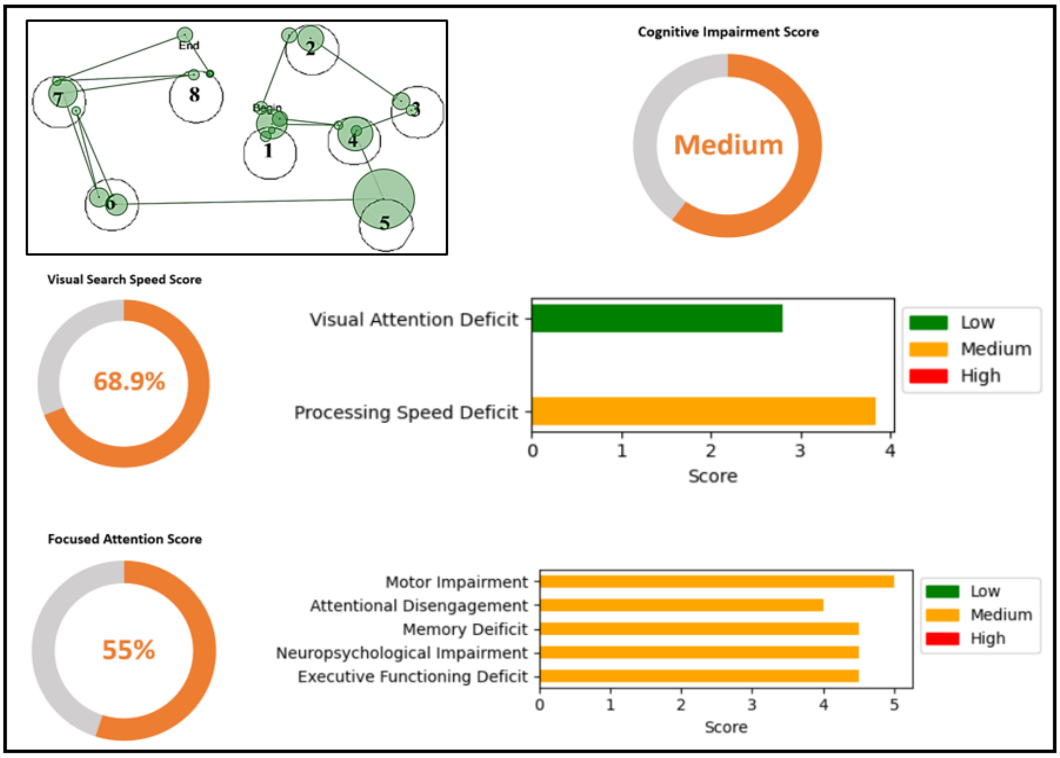
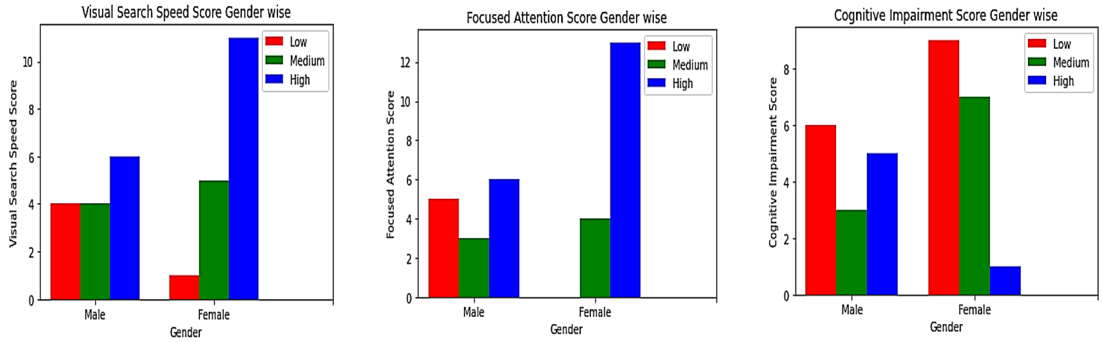


Figure 5.9: Detailed score generated by ETMT model for Participant 2 with medium cognitive impairment.

Participant 2 does exhibit a deficit in visual attention but, notably, possesses a moderate deficit in processing speed. The focused attention score suggests moderate impairment across various domains. In comparison to Participant 1, Participant 2 experiences a moderate level of cognitive impairment.

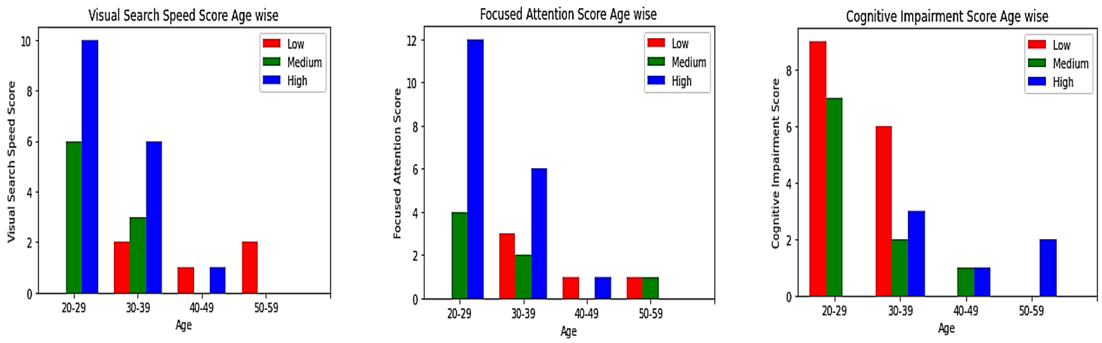


(a) Visual Search Speed Score

(b) Focused Attention Score

(c) Overall Cognitive Impairment Score

Figure 5.10: Gender-wise analysis of (a) Visual Search Speed score; (b) Focused Attention score (c) overall cognitive impairment score.



(a) Visual Search Speed Score

(b) Focused Attention Score

(c) Overall Cognitive Impairment Score

Figure 5.11: Age-wise analysis of (a) Visual Search Speed score; (b) Focused Attention score (c) overall cognitive impairment score.

In the paper-pencil based TMT, scores are generated based on the time it takes for an individual to complete the tasks. Participant 1 achieved a score of 6, indicating completion in 6 seconds, while participant 2 achieved a score of 9, indicating completion in 9 seconds. These paper-pencil-based TMT scores are highly correlated with their respective scores derived using the ETMT model.

The analysis considered each generated score in relation to the participant's age and gender, with the results displayed in Figure 5.10 and Figure 5.11. Notably, younger participants tended to exhibit high or medium visual search speed scores and focused attention scores, resulting in lower or medium cognitive impairment scores within the same age group. In contrast, the older age group showed lower visual search speed scores and predominantly low to medium-focused attention scores, which correlated with higher cognitive impairment scores in this age category. Gender-wise, a significant number of female participants demonstrated higher visual search speed scores than their male counterparts. Surprisingly, no female participants were observed with lower focused attention scores, and only one female participant received a high cognitive impairment score.

Further analysis of participant scores revealed a compelling pattern: those with high visual search speed and focused attention scores consistently obtained significantly lower cognitive impairment scores. In particular, participants with high visual search speed and focused attention scores predominantly fell into the low cognitive impairment score category and vice versa. This distribution of participants in each score is visually represented in Figure 5.12. The figure illustrates that eight participants with high visual search speed scores and eleven with high focused attention scores primarily belong to the low cognitive impairment score category.

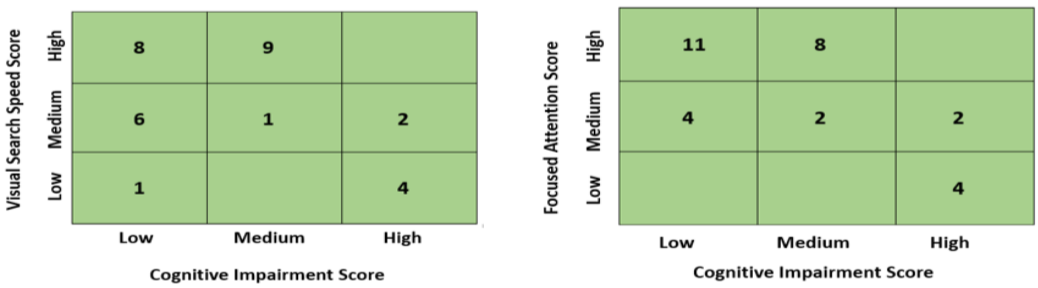


Figure 5.12: (a) Distribution of participants in cognitive impairment and Visual Search Speed scores (b) Distribution of participants in cognitive impairment and Focused Attention scores.

Moving to Figure 5.13, 5.14, and 5.15, these figures provide a visual representation of the distribution of each feature relative to cognitive impairment scores. The boxplot analysis of eye tracking features in these figures reveals distinct variations among groups with low, medium, and high cognitive impairment scores. These observed patterns strongly support the notion that eye tracking features can effectively capture meaningful differences in eye movement behavior associated with varying levels of cognitive impairments. This suggests the promising potential of eye tracking measures as valuable tools for assessing and understanding cognitive impairments.

The validation of data clustering and labeling was carried out using a range of machine learning algorithms, including decision trees, linear discriminant analysis (LDA), neural networks, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Naive Bayes. This validation process involved the allocation of 75%

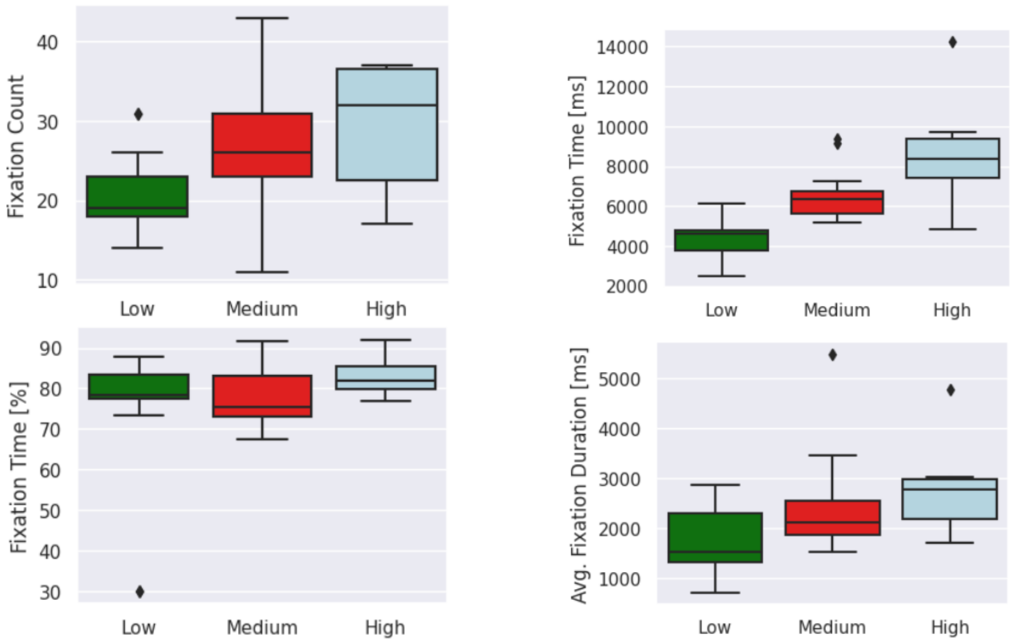


Figure 5.13: Boxplot of the low-level features based on cognitive impairment score

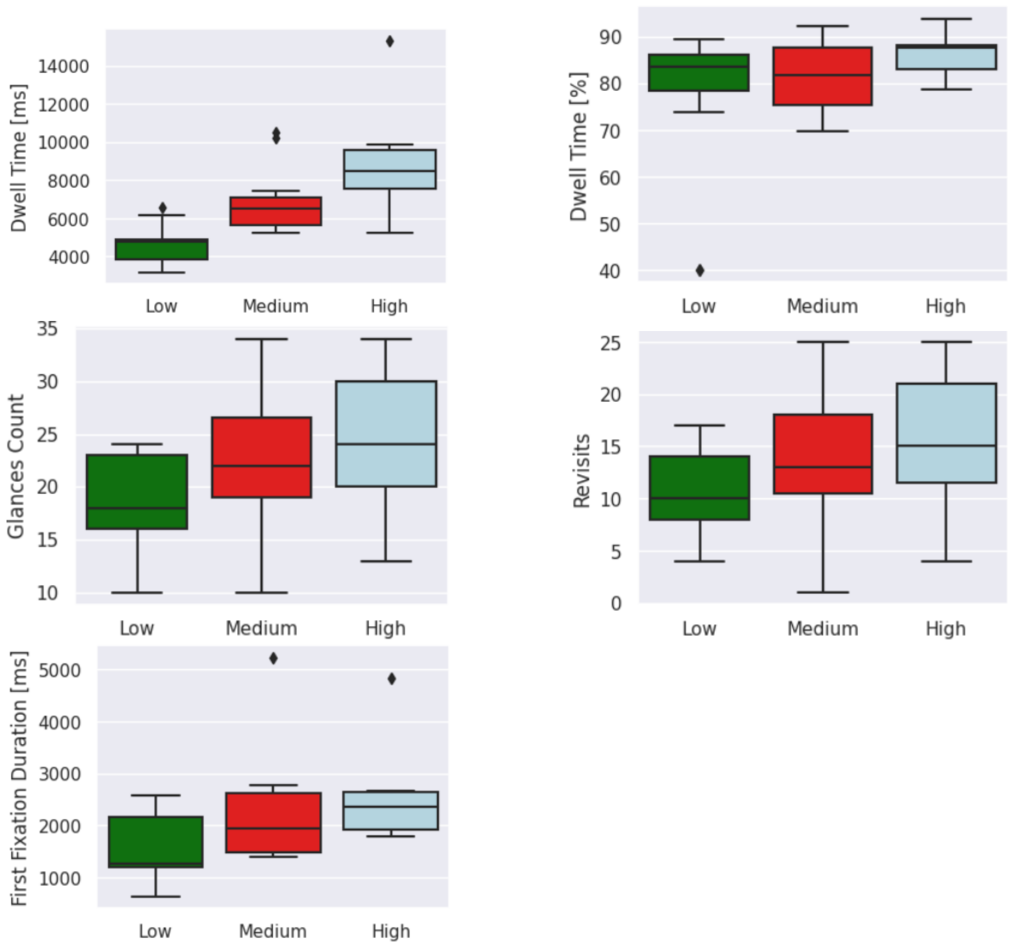


Figure 5.14: Boxplot of the middle level features based on cognitive impairment score

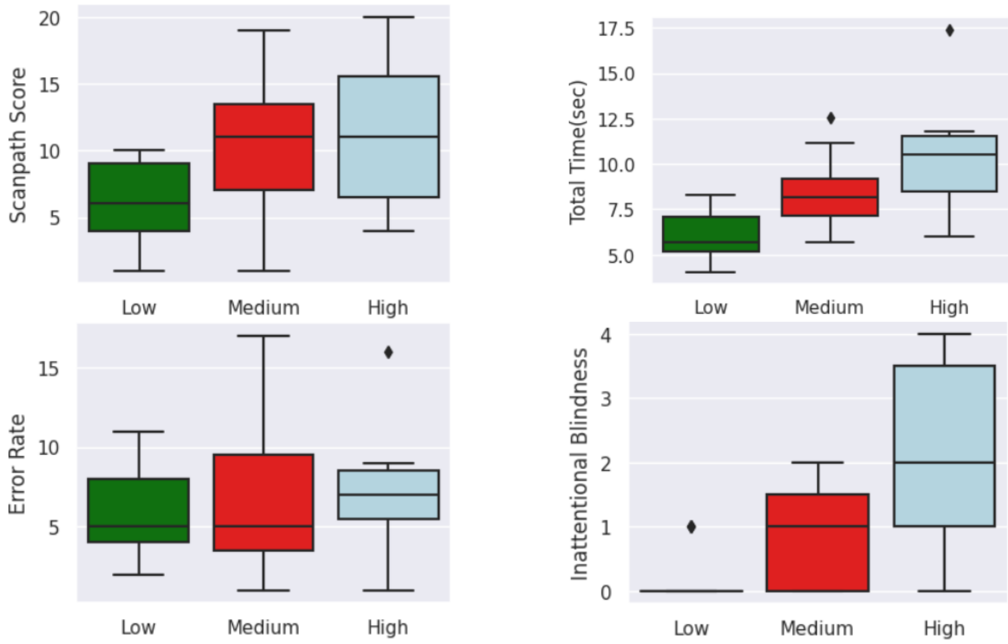


Figure 5.15: Boxplot of the high-level features based on cognitive impairment score

Table 5.5: Validation of classification based on cognitive impairment score

Machine Learning Algorithm	Testing accuracy
Decision Tree	100%
Linear Discriminant	75%
Neural Network	87.5%
KNN	87.5%
SVM	75%
Naive Bayes	87.5%

of the dataset for training purposes, with the remaining 25% dedicated to testing. The analysis was conducted using the Matlab toolbox 'classification learner,' and the results are presented in Table 5.5.

5.3.2 Comparison of ET_MT_CI Dataset with Traditional Paper-Pencil TMT Dataset

Data collection for the ETMT model included eye tracking and traditional TMT methods. The ETMT model extracted crucial features aiding cognitive impairment detection, while traditional TMT relied on completion time as a primary feature for the same purpose. A comparison between ETMT and the traditional TMT revealed a correlation in completion times. The minimum time taken by a participant to complete the entire ETMT test was 67 seconds, with a maximum of 168 seconds. In contrast, the traditional TMT had a minimum completion time of 38 seconds and a maximum of 184 seconds. Table 5.6 displays the time duration

Table 5.6: Time duration of ETMT and Traditional TMT

		ETMT	Traditional TMT
TMT-A simple	min time(ms)	4042	3800
	max time(ms)	17402.8	10200
TMT-A complex	min time(ms)	16821.8	12000
	max time(ms)	60461.3	52000
TMT-B simple	min time(ms)	3654.9	3500
	max time(ms)	17202.6	19900
TMT-B complex	min time(ms)	33438	18200
	max time(ms)	108670.7	102100
Entire test duration	min time(ms)	66619.1	37500
	max time(ms)	167504.9	184200

Table 5.7: Correlation of eye tracking TMT with the traditional method of TMT.

Stimulus	Correlation Score	P value
TMT A Simple	0.727	0.003
TMT A Complex	0.769	0.001
TMT B Simple	0.734	0.002
TMT B Complex	0.725	0.003

for each task in both ETMT and traditional TMT data collection. Importantly, participants did not experience stress during the experiment due to the simplicity and brevity of the tests.

The correlation between completion times in the traditional and eye-tracking versions of the TMT tasks was analyzed. The analysis revealed a notable correlation between the time of completion in the eye tracking-based and traditional versions of the TMT, as indicated by the correlation score and the p-value. The specific correlation values and p-values are provided in Table 5.7. This statistical analysis confirms a linear relationship between the completion times of the eye tracking and traditional TMT versions, providing robust support for the ETMT model's validity.

5.4 Discussion

The ETMT model, as an eye-tracking adaptation of the TMT, is well-suited to address impairments associated with focused attention and visual search speed by providing precise scores for these specific deficits. In the following sections, the benefits of the ETMT model will be projected in comparison with state-of-the-art models. Section 5.4.1 will detail the comparison with standard tests, while Section 5.4.2 will explain the comparison with the traditional paper-pencil-based TMT.

5.4.1 Comparison with Standard Tests

In contrast to standard tests like ECAS, MMSE, MoCA, and ADAS-CoG, which primarily focus on identifying specific diseases by pinpointing particular impairments associated with those diseases, the ETMT model offers a more comprehensive approach. The impairments that can be identified with standard tests are shown in Table 5.8. ETMT has the potential to provide a more detailed understanding of various impairments than other tests. It can reveal information about an individual's cognitive performance, specifically their ability to quickly search, assess, and interpret visual input without succumbing to distractions. This assessment also offers insights into a person's level of mental flexibility, focused attention, and processing speed, all of which fall within the realm of executive functioning. A significant decline in these skills could serve as an indicator of cognitive impairment.

Table 5.8: Cognitive impairments and associated standard tests

Tests	Motor impairment	visual attention	Attentional disengagement	Memory	Neuropsychological impairment	Social cognition	Executive functioning	Processing speed
ETMT [15]	✓	✓	✓	✓	✓	X	✓	✓
ECAS [57, 123, 124]	✓	✓	X	✓	X	✓	X	X
MMSE [35]	X	✓	✓	✓	X	X	X	X
ADAS-Cog [193]	X	✓	✓	✓	✓	X	✓	X
SCOPA-COG [129]	✓	✓	✓	✓	✓	X	✓	X
PD-CRS [129]	✓	X	✓	✓	✓	X	X	X
MoCA [130–132]	X	✓	✓	✓	X	X	✓	X
MACFIMS [57]	X	✓	✓	✓	X	X	✓	✓
MRI and CT scan [57]	✓	X	X	✓	✓	X	X	X

The significance of features such as the scanpath comparison score, the total time required for task completion, error rate, and inattentional blindness is that ETMT contributes to identifying cognitive impairment in individuals.

The scanpath comparison score aids in assessing the coherence and efficiency of eye movements during visual tasks, offering valuable insights into attentional focus and cognitive processing. The total time needed to complete a task reflects processing speed and efficiency, which can be compromised in individuals with cognitive impairment. The error rate provides valuable information about the accuracy of cognitive processing, with higher error rates suggesting potential cognitive deficits. Lastly, inattentional blindness, which pertains to the failure to notice salient stimuli, can indicate attentional impairments and cognitive challenges. These features play a crucial role in aiding psychologists in identifying cognitive impairment and developing suitable intervention strategies.

Compared to standard cognitive assessments, the utilization of eye tracking technology can offer a more accurate and precise means of assessing cognitive impairment. As shown in Table 5.9, it is a low-cost alternative, making it more accessible for researchers and patients. It does not require specialized training for assessment, which further contributes to its accessibility. Additionally, it becomes possible to detect cognitive impairment in its earlier stages by applying computational techniques to the retrieved eye gaze features. This is complemented by the shorter test duration, which does not overburden the participants. Finally, the detailed scores generated by eye tracking technology can also address lower motor neuron atrophy. The ETMT model, which combines eye tracking with the TMT, has the potential to revolutionize the conventional approach to cognitive impairment detection.

Table 5.9: Strengths of standard tests

Tests	Low cost	No specialized training to administer	Easy to detect in early stages	shorter time duration for the tests	Address lower motor neuron atrophy.	Does not stress the participant
ETMT [15]	✓	✓	✓	✓	✓	✓
Traditional TMT [109]	✓	X	✓	✓	✓	✓
ECAS [57, 123, 124]	✓	X	X	X	X	✓
MMSE [35]	✓	✓	X	✓	✓	✓
ADAS-Cog [193]	✓	✓	✓	X	✓	X
SCOPA-COG [129]	✓	✓	X	✓	✓	✓
PD-CRS [129]	✓	✓	X	✓	✓	✓
MoCA [130–132]	✓	✓	X	✓	✓	✓
MACFIMS [57]	✓	X	✓	X	✓	✓
MRI and CT scan [57]	X	X	✓	X	✓	✓

5.4.2 Comparison with Traditional TMT

As presented in Table 5.9, the ETMT model is compared with its traditional paper-pencil-based TMT to assess their respective strengths. ETMT highlights several strengths of the ETMT model, an eye-tracking version of the TMT, which can overcome many limitations of the conventional TMT method. The primary strength of the ETMT model is its ability to be administered without a trained evaluator, making it more accessible and efficient. This is particularly important in clinical and residential settings where there may be a shortage of mental health professionals, as highlighted by a study conducted by the Observer Research Foundation (ORF) in 2021 [52]. Notably, the ETMT model utilizes computational techniques to automatically generate scores, eliminating the need for the manual intervention typically required in traditional methods for score determination or final result assessment. The ETMT model also inherits the advantages of the traditional TMT, particularly in terms of completion time as shown in Table 5.7

Compared to the traditional TMT, an additional strength of the ETMT model is its ability to capture a broader range of features for evaluation. While the traditional TMT could only capture the total time to complete the task, the ETMT model could extract 13 distinct features, clearly indicating cognitive functioning. Specifically, the ETMT model can derive high-level features labeled as FH1 to

FH4, which are crucial in identifying cognitive impairment. These capabilities are invaluable in assisting healthcare professionals with more comprehensive and accurate cognitive health assessments.

It is worth noting that the ETMT model is currently limited to features derived from eye tracking during the TMT for detecting cognitive impairment. It does not encompass additional characteristics based on facial expressions or observations of other physiological signals. In the future, the ETMT model could be further enhanced by incorporating additional provisions to identify and assess various types of impairments, ultimately enabling the generation of comprehensive recommendations tailored to specific findings.

5.5 Conclusion

ETMT represents a valuable tool designed to assist healthcare professionals in assessing an individual's mental state by evaluating their cognitive impairment, particularly using the TMT. The primary innovation of ETMT lies in its ability to extract crucial high-level features to generate cognitive impairment scores. These high-level features, which include the error rate, scanpath comparison score, total time, and inattentional blindness score, are employed to calculate three distinct scores for screening cognitive impairment. In addition to an overall cognitive impairment score, ETMT provides detailed assessments of visual search speed and focused attention, allowing for a more precise understanding of a patient's deficits.

This tool comprehensively evaluates an individual's cognitive abilities, particularly visual perception and attentional processes. Psychologists can gain in-depth insights into an individual's cognitive functioning by incorporating advanced computational techniques and eye-tracking technology. This, in turn, leads to more accurate diagnoses, personalized treatment plans, and enhanced patient care. The integration of eye-tracking enriches practitioners' feedback during the administration of the TMT, enabling the identification of cognitive aspects that may require attention.

The ETMT model serves as a cost-effective screening tool for identifying cognitive impairments in a comfortable setting, allowing healthcare resources to be effectively allocated to patients with more severe conditions.

It's important to note that the ETMT model had limitations, particularly in the distribution of samples across different age groups. Furthermore, as the study did not involve actual patients, the results may not be directly applicable to clinical diagnosis. Nevertheless, the study provides a comprehensive understanding of cognitive impairment by offering multiple scores to assess various deficits.

In the future, there are opportunities to enhance ETMT by integrating it with eye tracker software as a plugin for automated cognitive impairment screening. This could include incorporating saccadic and blink features to better understand different cognitive impairment deficits. Collecting data from actual patients, accompanied by additional physiological measures as reference points, will further enhance the model's effectiveness. Additionally, future iterations of ETMT could encompass a broader range of impairments, providing scores and recommendations tailored to specific cognitive deficits or conditions.

This chapter explained one of the mental state parameters, cognitive impairment, and introduced the ETMT model for its classification. The next chapter will explain the PredictEYE model, which assesses an individual's mental state based on their emotional state while watching calm and stressful videos.

Chapter 6

Classification of Emotional State using PredictEYE Model

6.1 Introduction to Emotional State

Understanding emotional state is important in monitoring mental health, as it serves as a key indicator of an individual's emotional well-being and can provide valuable insights into their psychological state. Several factors, including depression, anxiety, stress, and significant life changes, can significantly impact one's emotional health [49]. Understanding these emotional variations is beneficial in the mental health assessment and support.

Emotions, thoughts, and behaviors are connected to mental health, and any change in emotional states can indicate the underlying mental health issues. Signs and symptoms like feeling depressed, sad, upset, or reduced attention span can indicate emotional variations. These changes are always connected with the mental health of an individual and it is critical to understand those changes.

Eye tracking plays a major role in understanding those changes that is related with emotional state. It can provide more quantifiable data that helps to understand the mental state based on emotional state of a person. By analyzing various eye gaze measures like pupil diameter, fixation, saccade and blink frequency, researchers could gain valuable insights to emotional response of an individual [59]. There may be lot of variation while viewing pleasant, unpleasant and neutral images or videos and that variations are helpful in understanding the emotional state detection. When a person is emotionally upset, the sympathetic nervous system can be activated and there can be variations in eye measures while viewing pleasant and unpleasant images [60].

In the study of emotions' influence on eye behavior, it has been observed that different emotional stimuli impact visual attention [47]. For instance, negative images can lead to more extensive and faster saccades, indicating visual agitation, discomfort, and avoidance behavior. Conversely, positive images have been associated with a strong center bias in latitude and shorter Y dispersion of fixation. These variations in eye behavior, such as pupil diameter, blink frequency, saccadic angle, gaze patterns, and fixation duration, are directly related to emotional reactions and can provide insights into a person's emotional state.

A personalized model is important in understanding the emotional state of a person using eye tracking technology. By tailoring emotional state detection algorithms to individual characteristics and responses, a personalized model can enhance the accuracy and specificity of mental health assessments. This personalized approach considers unique emotional expressions, making the analysis more refined and pertinent to each individual.

In line with this personalized approach, PredictEYE is proposed, a model utilizing deep learning techniques to accurately predict an individual's emotional state based on their responses to calm and stressful videos, identifying specific triggering scenes within a video in real time.

6.1.1 Personalized Model

Remote [194, 195] and personalized [19, 196, 197] monitoring are two approaches to track health and wellness data that have gained significant attention recently. Remote monitoring involves technology to monitor an individual's health and wellness remotely. Remote monitoring in the healthcare sector could empower physicians to deliver high-quality care by keeping patients safe and healthy [195]. With multiple sensing systems, physicians can track patients' health effectively, monitor remotely, and provide immediate care [195].

Personalized monitoring involves tracking health and wellness data specific to an individual. It includes tracking dietary intake, sleep patterns, and other vital parameters that can impact health [197, 198]. Individuals can learn more about their health and wellness by keeping track of this data. They can use this information to make wise decisions about changing their lifestyles in ways that can enhance their general welfare. On the other hand, A non-personalized model is trained on a larger dataset that is not specific to any individual or group. A non-personalized model aims to provide recommendations or generic predictions that are applicable to a larger population.

Developing a personalized model involves data collection, feature engineering, model selection, training, validation, deployment, and monitoring. Relevant features are identified, and an appropriate model is chosen for the research question. The model is trained, validated, deployed, and continuously monitored and updated for optimal performance in a real-world setting.

Many researchers have built systems based on machine learning [83, 85–89], and deep learning [19, 83, 90–93] approaches. The deep learning-based time series personalized model has the advantage of capturing patterns and changes over time, leading to more accurate and personalized predictions.

Section 6.2 introduces the personalized model, PredictEYE, and explains the features extracted for this model in section 6.2.1, time series prediction using LSTM in section 6.2.2, the process of emotional state prediction with the Random Forest algorithm in section 6.2.3, and emotional state prediction with GSR data in section 6.2.4. Section 6.3 explains the model's comprehensive analysis of its results, including the analysis based on data exploration and statistical analysis in section 6.3.1 and the Performance Evaluation of PredictEYE in 6.3.2, along with validation with GSR in section 6.3.3. PredictEYE's comparison with other state-of-the-art models is discussed in section 6.4 and finally concluded in section 6.5.

6.2 PredictEYE Model

Introduces the innovative PredictEYE model, showing Figure 6.1 designed to accurately predict an individual's mental state based on their emotional response to calm and stressful videos. PredictEYE utilizes deep learning techniques, specifically a univariate Long Short-Term Memory (LSTM) model and the Random Forest algorithm, to forecast future sequences of eye gaze data and simultaneously identify the specific scenes within a video that trigger particular emotional states [36, 199].

This novel approach distinguishes PredictEYE from conventional models that rely solely on physiological measures to assess emotional states. By analyzing time-series eye gaze data, PredictEYE can effectively predict future sequences, offering valuable insights into how an individual's visual attention evolves over time. To validate the model's effectiveness, it is assessed using GSR data as a benchmark, demonstrating its capability to capture and predict emotional states accurately. This integrated method is particularly powerful in understanding how emotional states correspond with an individual's visual engagement while watching relaxing and anxiety-inducing videos.

6.2.1 Feature Extraction

PredictEYE utilizes the ET_Video_ES dataset, employing fixation and saccade detection along with corresponding feature extraction conducted through BeGaze

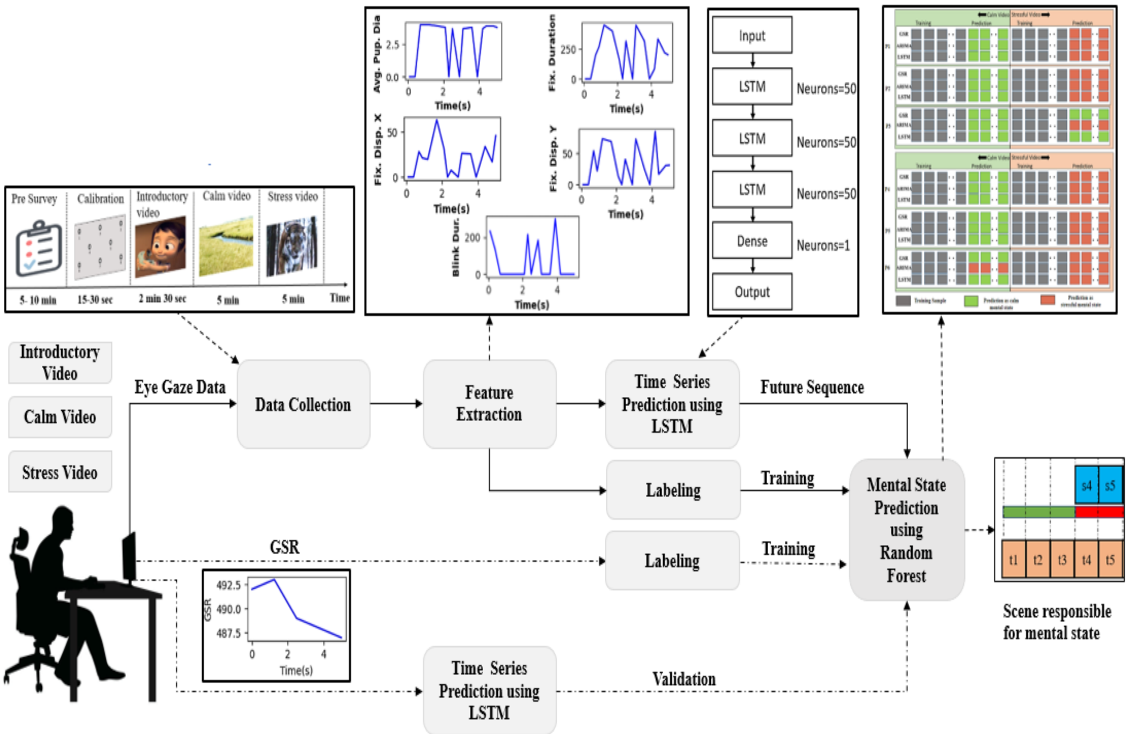


Figure 6.1: Architecture of PredictEYE, personalized time series model for emotional state prediction.
Avg Pup.Dia - average pupil diameter, Fix. Duration -fixation duration, Fix. Disp. X - fixation dispersion X, Fix. Disp. Y - fixation dispersion Y and Blink Dur. - blink duration.

3.7 software, which utilizes the identification by dispersion threshold (IDT) algorithm for precise event detection, including fixations and blinks [161]. The IDT algorithm identifies fixations by analyzing the dispersion and duration thresholds of the X and Y coordinates in the eye gaze data. Fixations are clusters of gaze points with low dispersion within a specific time frame. A minimum duration of 80 ms and a maximum dispersion of 100 pixels were set for fixation detection, while a 70 ms duration was considered for blink detection. Features such as Fixation Duration, Fixation Dispersion in the X and Y axes, Pupil Diameter, and Blink Duration were extracted from these events.

Fixation duration specifies how long a person is focusing on a specific object or any particular area in terms of milliseconds (ms). Fixation dispersion in the X and Y axes is the spatial variation in gaze patterns during fixation and that helps to understand attention distribution. Pupil diameter measures the size of a person's pupils measured in millimeters (mm) and reflects changes in cognitive functions or arousal. Blink duration, recorded in milliseconds, indicates eye closure during blinks and can provide insights into cognitive workload, fatigue, or attention.

This fixation, blink-based features, and pupil diameter were input for the

LSTM model. The data obtained while participants watched calm and stress-inducing videos were labeled and used to predict their emotional state through a Random Forest algorithm.

A threshold approach was employed for labeling, using the calm video period as the baseline phase. Any deviation from this baseline, determined by the threshold value, was labeled as 'stressful' or 'calm,' assuming that deviations from the baseline represented stress indicators.

The plot of these extracted features for Participant 1 while viewing the stress video is illustrated in Figure 6.1 as part of the feature extraction process. Each feature is plotted against its starting time.

6.2.2 Time series prediction using LSTM

The collected eye tracking time series data was employed in an LSTM (Long Short-Term Memory) model to predict new time series values using a sequence-to-sequence regression approach. The training data was initially normalized to have a mean of zero and a variance of one, preparing the model's predictors and responses. The LSTM network, depicted in Figure 6.1, consisted of three LSTM layers, each with 50 hidden units to store and manipulate temporal information. These LSTM layers were followed by a Dense output layer with a single output unit, responsible for predicting the subsequent time series value based on the input sequence of window-size time steps.

The window size, which determines the number of consecutive data points considered for calculations and predictions, was a crucial hyperparameter. Initially set to 60 to align with the data's sampling frequency of 60 Hz, it was later reduced to 10, representing a 1/6th of a second interval. This adjustment allowed for capturing finer changes in gaze behavior and improved prediction accuracy for future eye movements and reactions [200]. A window size that is too small in time series analysis can lead to issues like information loss and overfitting. Extensive experimentation confirmed that a value of 10 struck a balance between information capture and overfitting prevention.

In the LSTM model, input samples are structured sequentially, with each sample containing feature values spanning from time step 1 to 10, as illustrated in Figure 6.2. Subsequently, the following sample encompasses values from time steps 2 to 11, and this pattern continues. The model's training objective is to predict the next value based on the preceding ten values within each sequence. This approach employs a sliding window mechanism to establish input-output pairs, fostering the development of the LSTM model's predictive capabilities.

The activation function used in the LSTM layer was ReLU (Rectified Lin-

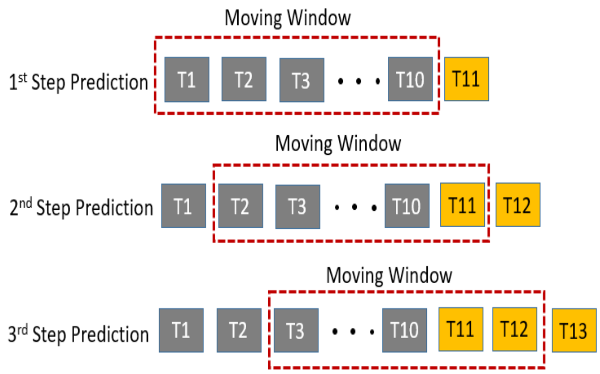


Figure 6.2: Time series prediction

ear Unit), a popular choice in deep learning models. The network parameters were optimized using the efficient Adam optimizer [201]. The mean squared error (MSE) loss function was employed for training, and the network was trained for 100 epochs with a batch size of 16. A gradient threshold of 1 was set to prevent gradient explosions. The initial learning rate was 0.005, with a reduction factor of 0.2 after every 125 epochs, ensuring smooth convergence and avoiding local minima.

During training, the model processed input sequences with a window-size of time steps and one feature, learning to predict the subsequent value in the time series. After training, the model was evaluated on a test set using the same window size, and the MSE loss was computed. Subsequently, the model generated a future sequence by iteratively predicting the next value in the time series based on the last window-size time steps of the training data. The LSTM network's training enabled it to capture dependencies and patterns in the input time series data, facilitating accurate predictions of future eye movements [200].

The input data for the task consists of time series eye tracking features such as Fixation Duration, Fixation Dispersion in the X and Y axes, Pupil Diameter, and Blink Duration. These features undergo preprocessing and are subsequently fed into a univariate Long Short-Term Memory (LSTM) model. LSTM, a recurrent neural network (RNN), is a well-regarded method for time series forecasting [202]. Within the LSTM architecture, memory cells equipped with input, forget, and output gates play a pivotal role. The input gate regulates the flow of input activation, the forget gate determines the duration of value retention within the cell, and the output gate governs the flow of cell activation into other networks. The primary function of the LSTM model is to discern patterns and dependencies within the eye gaze input data. By processing the time series eye tracking data, the model becomes proficient at capturing the temporal dynamics in the dataset, leveraging this knowledge to make predictions about future values of each eye tracking feature.

Time series model performance is typically assessed by comparing predicted values of eye tracking features to the actual target values of eye tracking features, employing various metrics, including:

1. Mean Absolute Error (MAE): Calculated as the average of the absolute differences between predicted and actual values, MAE is a positively valued metric. A lower MAE indicates greater accuracy, signifying that the model's predictions closely align with the true values. MAE is widely utilized in time series forecasting due to its interpretability.

2. Residual Sum of Squares (RSS): This metric measures the sum of squared residuals, where a lower value suggests closer alignment between the model's predictions and the true values. RSS provides insights into the model's overall fit, making it suitable for model performance comparisons.

3. Mean Squared Error (MSE): MSE is computed as the average of squared errors, resulting in a positively valued metric. A lower MSE indicates heightened accuracy, though it is less interpretable compared to MAE.

4. Root Mean Squared Error (RMSE): Derived from the square root of MSE, RMSE is positively valued, and a lower value signifies increased accuracy. RMSE is preferred for its interpretability and comparability across models.

5. Mean Absolute Percentage Error (MAPE): MAPE offers a relative measure of accuracy by calculating the average percentage difference between predicted and actual values. In time series forecasting, it's commonly used for straightforward accuracy assessment. A lower MAPE indicates improved accuracy, indicating closer proximity to actual values.

6. Mean Error (ME): Representing the average difference between actual and predicted values, ME helps assess bias. A value of zero indicates an unbiased model. While less common, ME can provide insights into error direction.

7. Mean Percentage Error (MPE): MPE computes the average percentage of actual values differing from predictions, offering information about the direction and magnitude of errors. A value of zero signifies an unbiased model.

These metrics facilitate the evaluation of the LSTM model's performance in time series forecasting, enabling the selection of the most suitable model for the given task. These error statistics are calculated based on the formula (6.1) to (6.7).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6.1}$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6.2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6.4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{6.5}$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{6.6}$$

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right) \times 100 \tag{6.7}$$

In the given formulas, y_i represents the actual eye gaze features, and $\hat{y}_i$ represents the predicted values of eye gaze features.

Comparing models using these metrics provides valuable insights into their performance, aiding researchers in identifying the most suitable models for a given task.

6.2.3 Emotional State Prediction with Random Forest Algorithm

The Random Forest algorithm is trained using labeled eye tracking data and leverages predicted sequences from the LSTM model for emotional state classification. As an ensemble learning method, Random Forest combines multiple decision trees to formulate predictions. In our study, the Random Forest algorithm serves as the classification tool for emotional states. Each decision tree within the Random Forest is trained with a distinct subset of the data, and the collective forecasts from all decision trees contribute to the final prediction.

PredictEYE, through the analysis of eye-tracking data, can also discern the specific areas within a video that an individual was focusing on during various video segments. These findings are then correlated with the individual's self-reported emotional state. This information provides valuable insights into the factors influencing an individual's emotional state, offering the potential for more targeted interventions or treatments.

The performance of the classification model is evaluated using multiple criteria, encompassing accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic (ROC) curve. These metrics assess the model's proficiency in correctly assigning instances to their respective classes. Accuracy is the ratio of correctly predicted instances to the total instances. Precision gauges the proportion of true positives (correctly identified cases) among all instances classified as positive, while recall measures the fraction of true positives correctly identified among all positive instances. The F1 score combines precision and recall to evaluate the model's performance comprehensively. The ROC curve graphically delineates the trade-off between the true positive rate (TPR) and false positive rate (FPR) of a classifier, with a higher Area Under the Curve (AUC) signifying the classifier's effectiveness in distinguishing between positive and negative examples.

The significance of accuracy, precision, recall, the F1 score, and the ROC curve stems from their capacity to assess a classification model's performance from various angles. While accuracy offers a holistic measure of the model's overall performance, precision and recall focus on its capability to recognize instances within a specific class accurately. The F1 score, as a harmonization of precision and recall, comprehensively evaluates the model's effectiveness.

The accuracy, precision, recall, F1 Score, TPR, and FPR used in the performance analysis of the PredictEYE model are calculated based on the formula (4.43) to (4.46) and (6.8) to (6.9) respectively.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6.8}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{6.9}$$

where TP is True Positive, TN is True Negative, FP is False positive, and FN is False Negative.

6.2.4 Emotional State Prediction with GSR Data

A non-invasive physiological measure, GSR is commonly utilized to gauge sympathetic nervous system activity, making it a valuable asset in studies exploring emotional and cognitive processing [28]. GSR stands out due to its ease of measurement and applicability across diverse settings, rendering it a popular choice for investigating emotional states. Notably, GSR exhibits an advantage over other physiological measures, such as heart rate or EEG, as it is less susceptible to motion artifacts and operates with a slower response time [203]. Given these attributes, GSR serves as a valuable tool for validating models designed to predict emotional states [204].

The validation process of PredictEYE's performance involved the collection of GSR data from participants while they viewed videos inducing calm and stressful states. Simultaneously, the eye movements of participants were tracked. Subsequently, an LSTM model was trained on this combined dataset to forecast future sequences of GSR data. In parallel, a Random Forest model was trained to predict participants' emotional states based on the anticipated GSR sequences. The Random Forest's predictions, driven by GSR data, were instrumental in corroborating the emotional state predictions derived from eye tracking data.

6.3 Result Analysis

PredictEYE is a personalized time series regression model meticulously crafted for the purpose of predicting an individual's emotional state based on their emotional state while watching calm and stressful videos. This model also concurrently identifies the scenes responsible for eliciting that emotional state. The evaluation of PredictEYE's efficacy comprises a two-fold analysis encompassing data exploration and statistical scrutiny, coupled with the assessment of its performance in emotional state estimation. The initial phase of the analysis involves data exploration and statistical assessment, focusing on the thorough examination of collected data. During this stage, efforts are concentrated on comprehending the dataset's distribution and variability. The aim is to unveil underlying patterns, trends, and correlations between diverse variables within the dataset, as this knowledge forms the bedrock for developing precise and dependable predictive models.

The next section focuses on testing how well PredictEYE can predict future eye movements to measure its accuracy. The final step is evaluating how well the model can estimate someone's emotions from their eye movements. This phase trains the model to recognize different emotions and helps us understand how well PredictEYE can predict someone's mood.

The following section elaborates on the data exploration and statistical analysis in Section 6.3.1. Following this, the performance evaluation of PredictEYE is discussed in Section 6.3.2, where the effectiveness of LSTM and Random Forest in estimating the emotional state is analyzed. Validation with GSR is further explained in Section 6.3.3.

6.3.1 Data Exploration and Statistical Analysis

The analysis of the mean eye-tracking measurements taken while participants watched both relaxing and stressful videos is presented in Figure 6.3. Within this analysis, 'P1-C' and 'P1-S' denote Participant 1's data during calm and stressful video-watching periods, respectively. The study revealed that during the stressful video-watching sessions, there was an increase in the mean pupil diameter and fixation dispersion in both the X and Y axes, whereas the mean fixation duration and blink duration decreased in comparison to the calm video-watching sessions. These findings suggest that eye-tracking features hold the potential for detecting changes in an individual's emotional state, particularly during tasks that induce stress. Furthermore, it was observed that stress or emotional upset triggers the activity of sweat glands, resulting in increased moisture secretion and decreased resistivity, which sensors can measure. Most participants exhibited decreased GSR values during the stress video-watching sessions, indicating physiological responses associated with increased arousal. It is important to note that GSR provides a more direct measurement of arousal, reflecting physiological responses, whereas eye-tracking data indirectly captures changes in visual attention and gaze behavior. Even subtle variations in eye-tracking data are valuable as they shed light on how individuals emotionally respond to different stimuli.

Participant 3 exhibited a significant difference in the measured variables compared to the other participants. Notably, there was no significant difference in this participant's body resistivity between stress and calm video-watching sessions, indicating that they did not experience a substantial increase in stress, as confirmed by the ground truth data. The eye-tracking measures aligned with this observation, showing variations in specific features for this participant in contrast to others.

To determine whether there were significant differences in eye-tracking data between participants while watching calm and stressful videos, a Welch two-sample t-test was conducted on all features. This statistical test is employed to compare the means of two independent groups when the variances of these groups are unequal. The goal was to investigate the emotional state of individuals after watching calm and stressful videos. While each video had a duration of 5 minutes, the analysis focused on the final minute of each video. This statistical test aimed to assess potential disparities in eye movement patterns between the two types of videos, providing insights into how the content influenced the emotional state of the participants during observation.

The null hypothesis states that there was no significant difference between the two sets of data, while the alternative hypothesis suggested a notable difference

Table 6.1: Results of Welch two sample t-test on classification of emotional state

Participant	Feature Name	P-Value	Calm (mean)	Calm (SD)	Stressful (mean)	Stressful (SD)	Segment length
P1	APD (mm)	**<0.0001**	2.36	1.73	3.23	1.79	169
	FD (ms)	0.91	204.44	199.80	202.00	163.91	169
	FDX (pixels)	**0.0001**	16.79	19.42	26.55	23.68	169
	FDY (pixels)	0.75	33.09	31.34	34.19	27.75	169
	BD (ms)	**0.0009**	67.96	97.60	35.33	68.16	169
	GSR (mV)	**<0.0001**	484.07	7.23	419.66	13.62	60
P2	APD (mm)	**<0.0001**	0.65	0.93	1.38	1.42	166
	FD (ms)	0.15	66.98	115.33	86.05	114.97	166
	FDX (pixels)	**<0.0001**	12.24	21.54	25.85	30.94	166
	FDY(pixels)	**0.01**	19.72	30.19	29.25	32.49	166
	BD (ms)	**<0.0001**	297.70	544.14	90.45	193.49	166
	GSR (mV)	**<0.0001**	406.19	4.36	399.33	8.20	60
P3	APD (mm)	**0.03**	1.69	1.17	1.32	1.38	111
	FD (ms)	0.60	289.57	320.19	265.15	397.83	111
	FDX (pixels)	0.50	27.42	28.72	24.84	30.37	111
	FdY (pixels)	**0.04**	29.54	28.58	22.03	26.91	111
	BD (ms)	0.70	293.43	1535.25	356.22	805.79	111
	GSR (mV)	0.25	486.45	1.394538	485.35	3.950683	60
P4	APD (mm)	**<0.0001**	3.61	1.09	4.67	1.25	140
	FD (ms)	0.44	336.99	220.02	314.94	256.04	140
	FDX (pixels)	**<0.0001**	26.56	14.78	35.35	18.96	140
	FDY (pixels)	**0.03**	29.54	28.58	21.42	26.58	140
	BD (ms)	0.41	15.10	52.24	10.51	40.26	140
	GSR (mV)	**<0.0001**	407.71	14.08	389.49	12.62	60
P5	APD (mm)	**<0.0001**	4.09	0.85	4.59	0.79	135
	FD (ms)	**<0.0001**	519.74	532.02	310.33	262.33	135
	FDX (pixels)	**0.002**	26.43	15.10	32.79	17.85	135
	FDY (pixels)	0.22	18.39	11.03	16.83	10.30	135
	BD (ms)	0.49	13.19	77.76	7.77	47.35	135
	GSR (mV)	**<0.0001**	489.98	2.06	483.40	3.55	60
P6	APD(mm)	**<0.0001**	2.20	1.62	3.60	1.16	150
	FD (ms)	0.31	388.23	542.96	324.16	279.60	150
	FDX (pixels)	**<0.0001**	20.52	22.81	48.12	31.40	150
	FDY (pixels)	**0.04**	27.05	27.30	35.14	27.66	150
	BD (ms)	**0.03**	124.75	314.60	40.55	190.61	150
	GSR (mV)	**<0.0001**	403	14.01	386.47	7.12	60

APD − Avg Pupil Diameter, FD − Fixation Duration, FDX − Fixation Dispersion X,
FDY − Fixation Dispersion Y, BD − Blink Duration

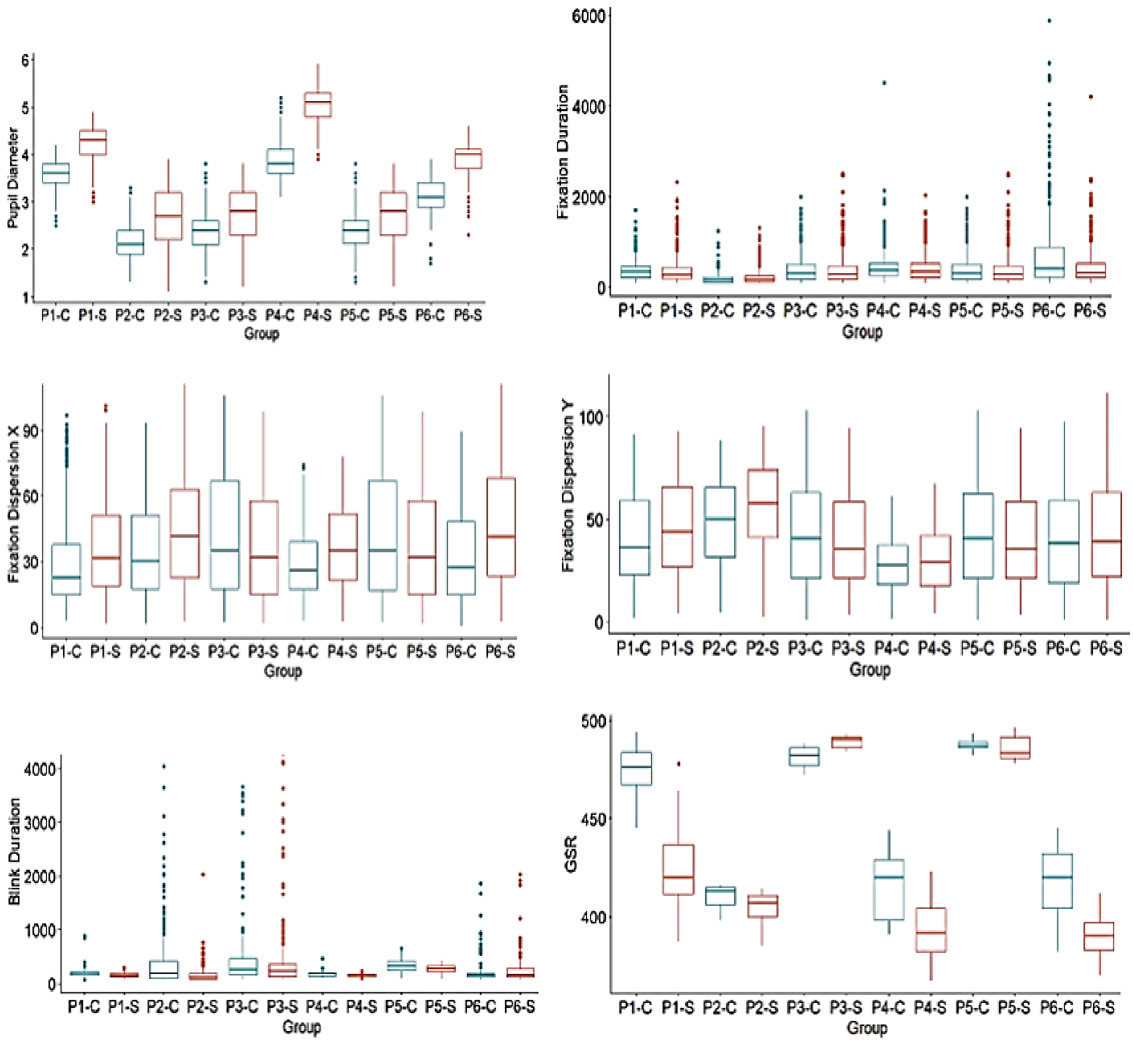


Figure 6.3: Boxplot of the eye tracking measures
P1-C indicates Participant 1 while watching the calm video, and P1-S indicates Participant 1 while watching the stressful video. Likewise, other abbreviations can be expanded.

between the calm and stress data. The test was conducted to determine whether the observed feature differences between the two video types were likely due to chance or were indicative of a significant distinction between the two states. The P-values obtained from the Welch Two-sample t-test for all the features are presented in Table 6.1.

Based on the results of the Welch Two Sample t-test, it can be concluded that there is a significant difference between the calm and stressful video data for most features, except for Participant 3. This implies that the null hypothesis can be rejected in favor of the alternative hypothesis, indicating that for most analyzed features, there were substantial differences in data after watching the stressful video compared to the calm video, with the exception of Participant 3.

The P-values for most participants being less than 0.05 for most features sug-

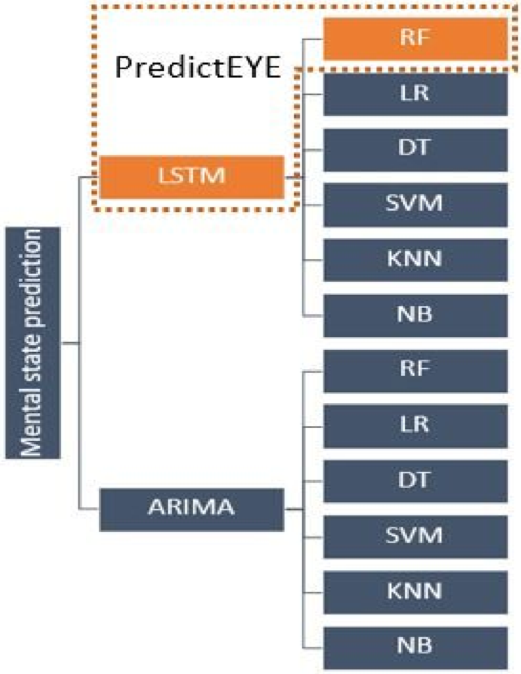


Figure 6.4: Determining PredictEYE through comparison with other models

gest that the differences between the calm and stressful video data were indeed significant and not merely a matter of chance. However, for Participant 3, the P-values for the features fixation duration, fixation dispersion in the X-axis, and blink duration at the end of calm and stressful video-watching were not less than 0.05. This indicates no significant differences were observed in these features during this participant's calm and stressful video-watching sessions. Additionally, the ground truth GSR values for Participant 3 at the end of calm and stressful video-watching also had P-values that were not less than 0.05, suggesting no significant difference in GSR between these two video-watching sessions.

6.3.2 Performance Evaluation of PredictEYE

In the data collection phase, 216,000 data samples were acquired from all the participants within a 10-minute time frame. On average, each participant contributed approximately 36,000 data samples under observation during calm and stressful video sessions. These gathered data samples played a vital role in making predictions about the participants' responses, offering a rich dataset for subsequent analysis. Standard performance metrics such as Mean Error, Mean Absolute Error, Mean Percentage Error, Mean Absolute Percentage Error, Mean Squared Error, and Root Mean Squared Error were employed to gauge the accuracy of the predicted data sequences. This allowed for a comprehensive analysis of prediction accuracy and error statistics using various performance measures.

The determination of the PredictEYE model, a strategic combination of LSTM

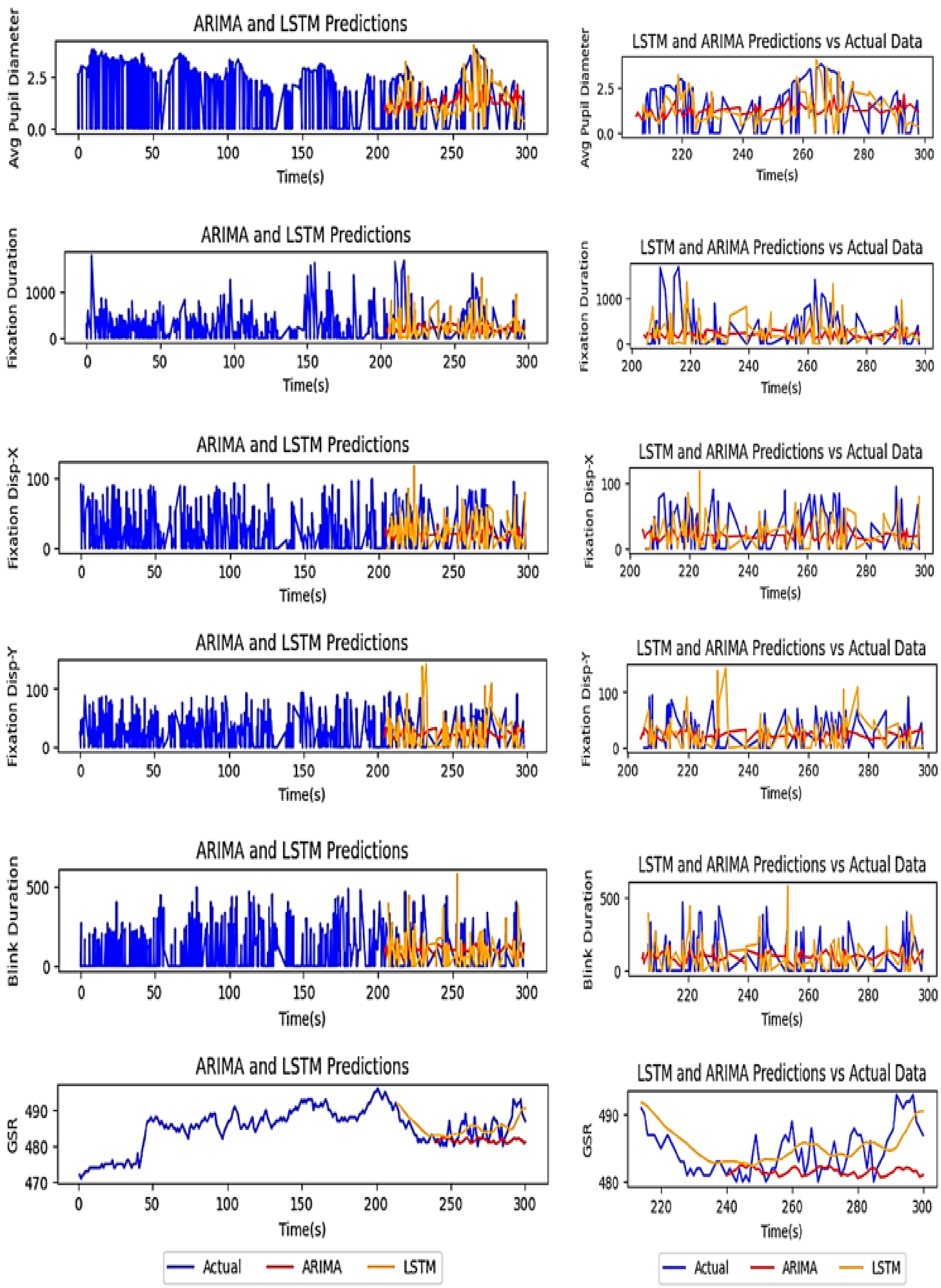


Figure 6.5: Forecasting data sequence of Participant-3 after watching stressful video using LSTM and ARIMA models in PredictEYE. Fixation Disp-X is Fixation Dispersion X and Fixation Disp-Y is Fixation Dispersion Y

and Random Forest (RF) models, was grounded in a rigorous comparison with alternative model combinations, as illustrated in Figure 6.4. The evaluation process commenced by contrasting the time series predictions of LSTM and ARIMA with

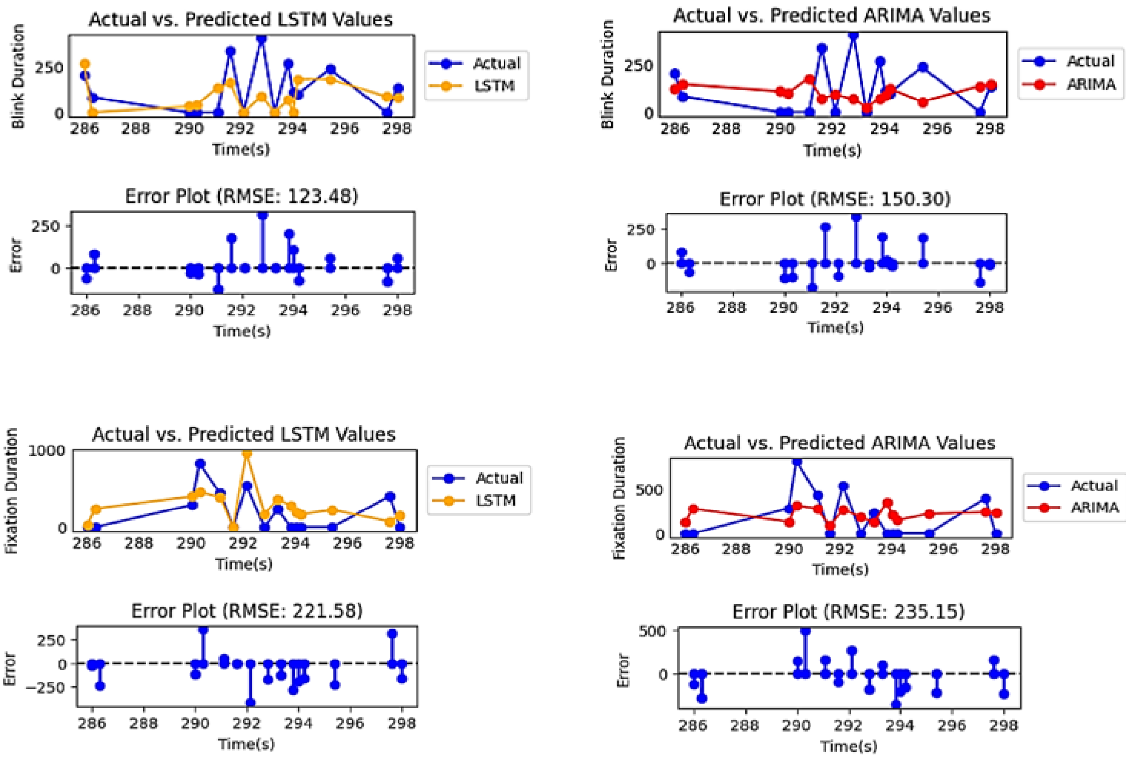


Figure 6.6: Comparison between actual and predicted values of LSTM and ARIMA for a short interval

the goal of discerning the unique attributes that ultimately shaped the composition of the PredictEYE model. Following this, the emotional state predictions based on Random Forest were examined across a range of machine learning algorithms, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K Nearest Neighbor (KNN), and Naive Bayes (NB). The subsequent section extensively compares the PredictEYE model with these different models, emphasizing its performance and effectiveness in predicting emotional states.

Performance Evaluation of LSTM with ARIMA based on Error statistics

The ARIMA model is a specialized time series analysis algorithm to uncover trends, seasonality, and cyclic patterns within the data [205]. Prior to model fitting, an exploratory data analysis is conducted, involving the examination of autocorrelations between current and past values across different time lags. This analysis helps determine the appropriate ARIMA model for the dataset, with the ACF and PACF plots assisting in understanding the model's autoregressive, moving average, and integrated components. The selected ARIMA model can then predict subsequent values within the eye tracking features.

To assess the prediction accuracy, the sequences predicted by the LSTM and ARIMA models at various intervals during the viewing of both calm and stressful videos were compared to the actual values. Figure 6.5 illustrates the results for the stressful video scenario, with 80% of the data used for model training and the remaining 20% for prediction utilizing both LSTM and ARIMA methods. This figure provides insights into the performance and predictive capabilities of these models within the dataset.

Table 6.2: Error statistics based on the prediction while watching calm video

Model	Features	Performance measures						
		MAE	**RSS**	**MSE**	**RMSE**	**MAPE**	**ME**	**MPE**
LSTM	**APD**	0.34	10.55	0.19	0.42	0.19	0.05	-0.01
	FD	139.66	1689383.93	29638.31	172.16	71.77	29.41	51.25
	FDX	14.86	17837.80	312.94	17.69	133.27	8.38	121.79
	FDY	21.2	33705.17	591.31	24.31	60.84	-5.79	23.72
	BD	41.81	131132.49	2731.92	52.26	34.46	22.09	26.05
	GSR	1.42	21.17	4.23	2.05	0.30	0.98	0.21
ARIMA	**APD**	0.73	38.82	0.57	0.76	19.84	-0.73	-19.84
	FD	171.86	4406649	73884.91	255.19	1.05	-51.89	0.22
	FDX	15.33	23562.35	351.67	18.75	123.89	4.87	105.62
	FDY	21.76	41707.75	622.50	24.95	56.42	-9.16	12.54
	BD	50.25	136372.19	3588.74	59.90	42.26	36.42	36.74
	GSR	5.2	386.18	43.31	5.76	0.14	-4.9	-0.01

APD − Avg Pupil Diameter, FD − Fixation Duration, FDX − Fixation Dispersion X,
FDY − Fixation Dispersion Y, BD − Blink Duration

Furthermore, Figure 6.6 offers a comparison between actual and predicted values for the 'blink duration' and 'fixation duration' features based on the LSTM and ARIMA predictions, focusing on a shorter time interval. This comparison reveals that fixations exclusively occur during the absence of blinks. Notably, the decrease in RMSE in the LSTM model, as depicted in Figure 6.6, holds significant implications for model performance assessment. A lower RMSE signifies that the LSTM model yields predictions closer to the actual values, indicating its efficacy in capturing and predicting subtle variations in eye movement patterns. This demonstrates the LSTM model's usefulness in accurately comprehending and analyzing rapid changes in gaze behavior.

Comparative analysis of error statistics for the LSTM and ARIMA models is conducted based on performance metrics, encompassing mean absolute error (MAE), residual sum of squares (RSS), mean squared error (MSE), root mean

Table 6.3: Error statistics based on the prediction while watching stress video

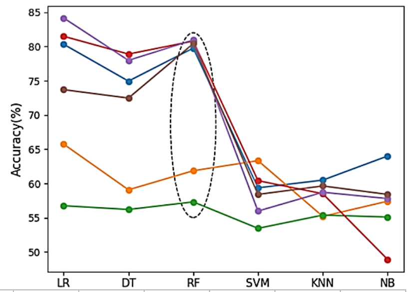
Model	Features	Performance measures						
		MAE	**RSS**	**MSE**	**RMSE**	**MAPE**	**ME**	**MPE**
LSTM	**PD**	0.22	4.44	0.06	0.25	5.56	0.08	2.36
	FD	102	883989.32	15785.52	125.64	54.21	24.21	36.6
	FDX	16.49	23618.86	421.76	20.53	96.23	-0.30	67.91
	FDY	16.79	24897.53	440.61	20.25	1.27	-0.78	0.37
	BD	41.62	28188.9	2818.89	53.09	23.42	-2.97	6.14
	GSR	3.83	250.43	28.92	4.52	0.10	-1.61	-0.01
ARIMA	**PD**	0.38	13.69	0.31	0.45	0.24	-0.05	-0.01
	FD	171.86	4406649	73884.91	255.19	1.05	-51.89	0.22
	FDX	17.38	29049.29	440.14	20.97	109.35	5.75	91.88
	FDY	17.83	30656.09	559.01	22.28	0.73	-1.49	-0.23
	BD	45.66	65777.52	3288.87	57.34	35.83	8.3	20.3
	GSR	9.26	1168.63	106.23	10.3	2.09	-9.26	-2.09

APD – Avg Pupil Diameter, FD – Fixation Duration, FDX – Fixation Dispersion X, FDY – Fixation Dispersion Y, BD – Blink Duration

squared error (RMSE), mean absolute percentage error (MAPE), mean error (ME), and mean percentage error (MPE). The summarized results are presented in Tables 6.2 and 6.3, facilitating an extensive model comparison and aiding in the determination of the better-suited model for the given dataset.

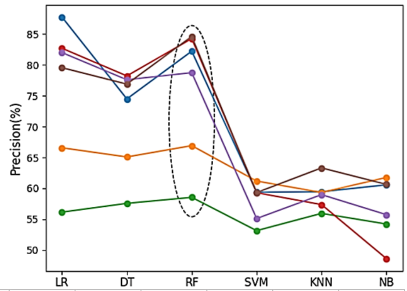
Upon scrutinizing the compiled error statistics for each feature across various performance metrics, it becomes evident that the LSTM model exhibits lower values than the ARIMA models. These reduced values across all evaluation measures in the LSTM model suggest its superior ability to capture data patterns and provide more precise predictions. Consequently, based on the error statistics analysis and diverse performance metrics, it can be deduced that the LSTM model surpasses the ARIMA model in terms of accuracy and predictive capabilities. The LSTM model's proficiency in capturing intricate patterns and long-term dependencies within time series data likely accounts for its superior performance over the ARIMA model.

In the context of the PredictEYE model, the selection of the LSTM model is rationalized by its aptitude for handling sequential data and capturing extended dependencies. While ARIMA serves as a robust tool for time series forecasting, the complexity and non-linearity inherent in eye gaze sequence data may render it less suitable for the task. Consequently, it is reasonable to infer that the LSTM model's adoption would result in enhanced accuracy and predictive performance, particularly for forecasting eye gaze sequences within the PredictEYE model [206].



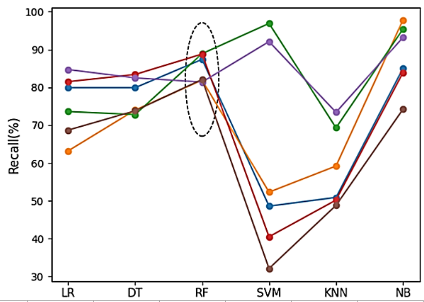
	LR	DT	RF	SVM	KNN	NB
P1	80.3	75.0	79.8	59.4	60.5	64.0
P2	65.7	59.1	61.9	63.3	55.2	57.4
P3	56.7	56.2	57.3	53.4	55.4	55.1
P4	81.5	78.9	80.8	60.4	58.5	48.9
P5	84.2	78.0	81.0	56.0	58.7	57.8
P6	73.7	72.5	80.4	58.4	59.6	58.4

(a) Accuracy



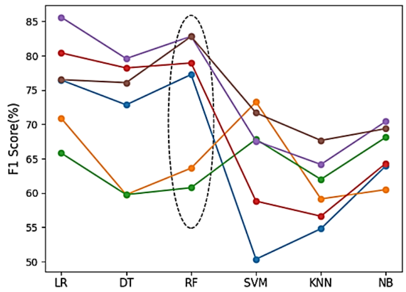
	LR	DT	RF	SVM	KNN	NB
P1	87.8	74.5	82.2	59.4	59.4	60.6
P2	66.6	65.1	66.9	61.2	59.3	61.8
P3	56.1	57.6	58.5	53.1	55.9	54.2
P4	82.7	78.2	84.2	59.3	57.3	48.6
P5	82.0	77.6	78.7	55.1	59.0	55.7
P6	79.5	76.9	84.6	59.3	63.3	60.6

(b) Precision



	LR	DT	RF	SVM	KNN	NB
P1	79.9	79.9	87.4	48.6	50.9	85.0
P2	63.2	74.0	81.9	52.3	59.2	97.8
P3	73.6	72.8	88.9	96.9	69.3	95.4
P4	81.5	83.4	88.8	40.5	50.2	83.9
P5	84.7	82.5	81.4	92.1	73.4	93.2
P6	68.7	73.8	82.1	32.1	48.8	74.2

(c) Recall



	LR	DT	RF	SVM	KNN	NB
P1	76.4	72.8	77.2	50.3	54.8	63.9
P2	70.9	59.7	63.6	73.2	59.1	60.5
P3	65.8	59.7	60.8	67.8	62.0	68.1
P4	80.4	78.2	78.9	58.9	56.6	64.3
P5	85.6	79.6	82.8	67.6	64.1	70.4
P6	76.5	76.1	82.8	71.7	67.6	69.4

(d) F1 Score

Figure 6.7: Analysis of classification model based on prediction with LSTM
LR – Logistic Regression, DT – Decision Tree, RF – Random Forest, SVM – Support Vector Machine, KNN – K Nearest Neighbor, NB – Naive Bayes

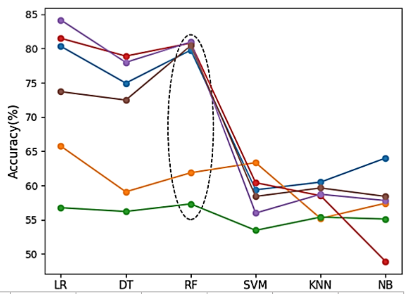
Performance Evaluation of PredictEYE in Estimating the Emotional State

This study extensively evaluated six distinct machine learning algorithms, specifically Logistic Regression, Decision Tree, Random Forest, SVM, KNN, and Naive Bayes, to assess their performance in predicting participants' emotional states. These predictions were based on the outputs of ARIMA and LSTM models. A notable trend emerged after conducting a thorough comparative analysis of these machine-learning models. It was observed that the machine learning algorithms consistently delivered superior results when utilizing the predictions generated by the LSTM model as opposed to those from the ARIMA model.

In assessing the classification algorithms' performance, various metrics were employed, including accuracy, precision, recall, F1 score, and the ROC curve.

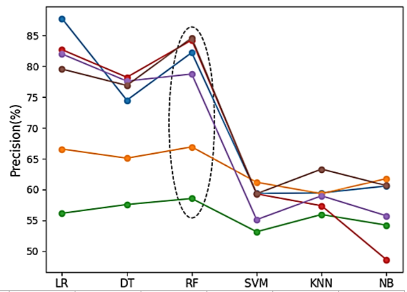
These metrics played a pivotal role in identifying the top-performing algorithm. Accuracy, for instance, quantifies the proportion of correctly predicted emotional states (calm or stressful) within the dataset. On the other hand, precision delineates the percentage of correctly predicted instances of a specific emotional state out of all the instances forecasted as that emotional state. Similarly, recall represents the ratio of correctly predicted instances of a specific emotional state relative to all actual instances of that emotional state in the dataset. Finally, the F1 score serves as a harmonious balance between precision and recall, characterizing the performance of the model for each emotional state.

The results, as presented in Figures 6.7 and 6.8, underscore the significance of the time series predictions of the LSTM model. Specifically, the prediction of emotional states relying on LSTM predictions exhibited higher levels of accuracy in comparison to ARIMA-based predictions. Moreover, an examination of the performance metrics revealed that all classification algorithms consistently outperformed the LSTM model in contrast to ARIMA-based predictions.



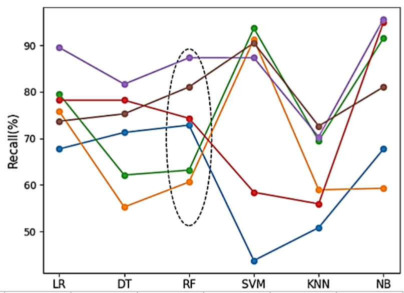
	LR	DT	RF	SVM	KNN	NB
P1	80.3	75.0	79.8	59.4	60.5	64.0
P2	65.7	59.1	61.9	63.3	55.2	57.4
P3	56.7	56.2	57.3	53.4	55.4	55.1
P4	81.5	78.9	80.8	60.4	58.5	48.9
P5	84.2	78.0	81.0	56.0	58.7	57.8
P6	73.7	72.5	80.4	58.4	59.6	58.4

(a) Accuracy



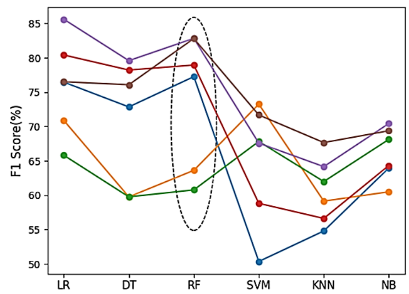
	LR	DT	RF	SVM	KNN	NB
P1	87.8	74.5	82.2	59.4	59.4	60.6
P2	66.6	65.1	66.9	61.2	59.3	61.8
P3	56.1	57.6	58.5	53.1	55.9	54.2
P4	82.7	78.2	84.2	59.3	57.3	48.6
P5	82.0	77.6	78.7	55.1	59.0	55.7
P6	79.5	76.9	84.6	59.3	63.3	60.6

(b) Precision



	LR	DT	RF	SVM	KNN	NB
P1	67.7	71.3	72.8	43.7	50.8	67.7
P2	75.8	55.2	60.6	91.2	58.9	59.3
P3	79.5	62.1	63.2	93.7	69.5	91.6
P4	78.2	78.2	74.3	58.4	55.9	95.0
P5	89.5	81.7	87.3	87.3	70.3	95.6
P6	73.7	75.3	81.1	90.5	72.6	81.1

(c) Recall



	LR	DT	RF	SVM	KNN	NB
P1	76.4	72.8	77.2	50.3	54.8	63.9
P2	70.9	59.7	63.6	73.2	59.1	60.5
P3	65.8	59.7	60.8	67.8	62.0	68.1
P4	80.4	78.2	78.9	58.9	56.6	64.3
P5	85.6	79.6	82.8	67.6	64.1	70.4
P6	76.5	76.1	82.8	71.7	67.6	69.4

(d) F1 Score

Figure 6.8: Analysis of classification model based on prediction with ARIMA
LR – Logistic Regression, DT – Decision Tree, RF – Random Forest, SVM – Support Vector Machine, KNN – K Nearest Neighbor, NB – Naive Bayes

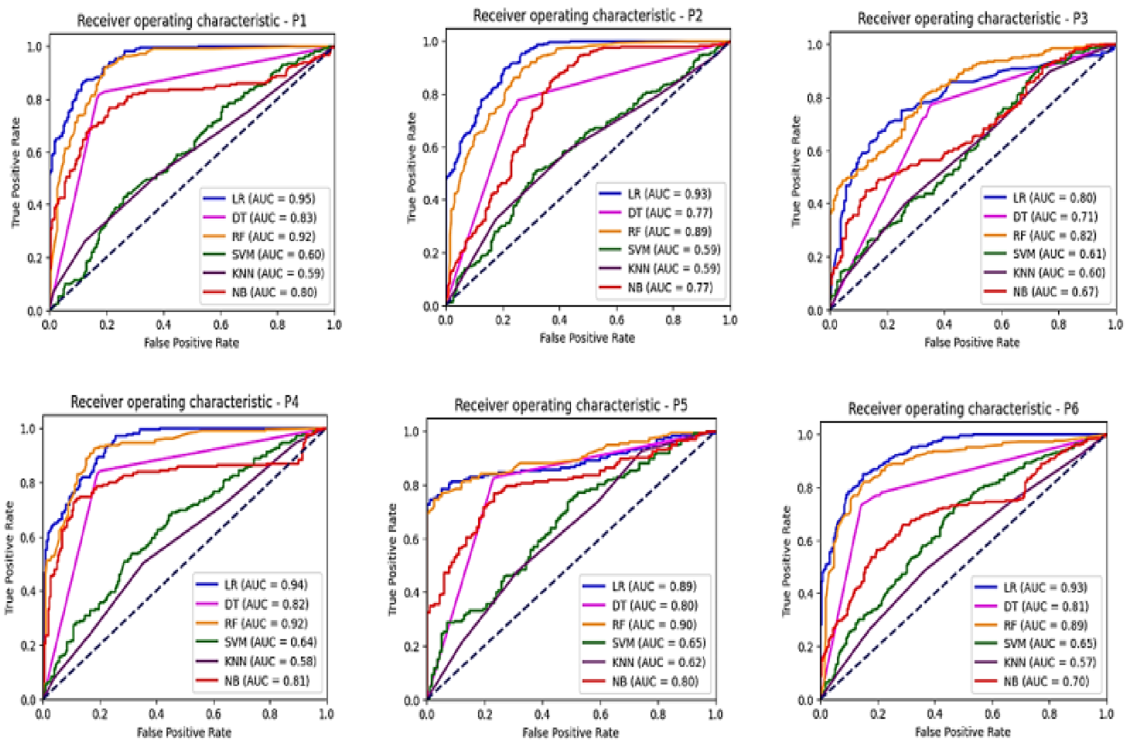


Figure 6.9: ROC curves based on classification algorithms applied after the prediction with LSTM

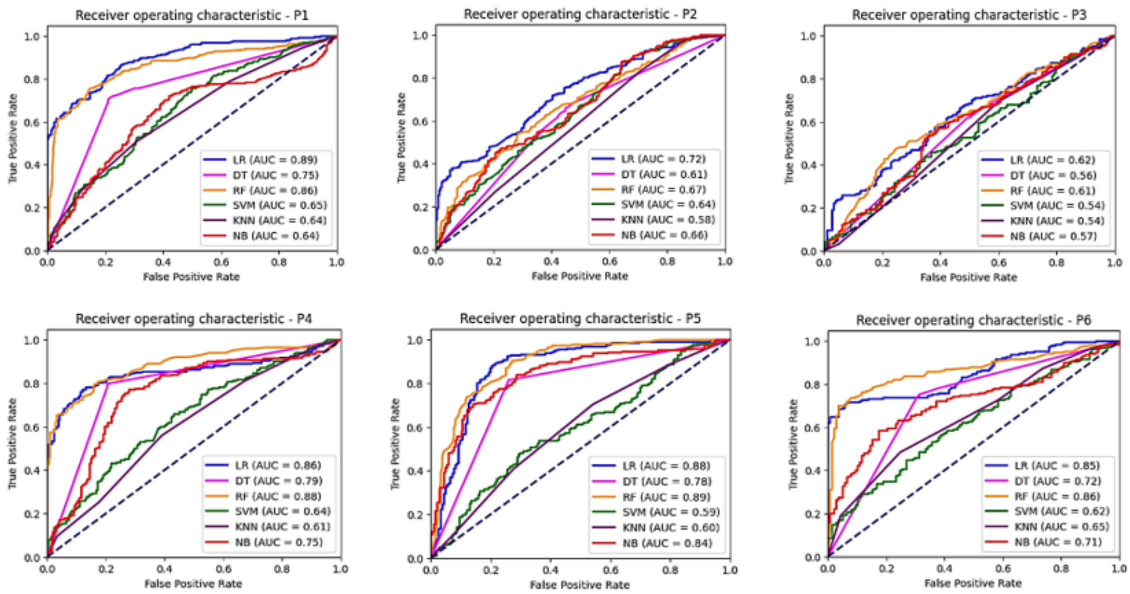


Figure 6.10: ROC curves based on classification algorithms applied after the prediction with ARIMA

Table 6.4 provides a comprehensive summary of the performance metrics extracted from diverse models compared to PredictEYE. PredictEYE, combining LSTM and Random Forest, notably achieved an accuracy rate of 86.4% and the highest F1 Score of 86.3% for Participant 4. While precision and recall did not reach similar peaks, the Random Forest model showcased superior performance across all metrics, surpassing other models.

This outcome underscores the importance of selecting the most suitable classification algorithm and rigorously evaluating model performance via multiple metrics to understand its effectiveness better. The consistently high performance of Random Forest in classifying and predicting emotional states based on LSTM time series predictions establishes it as the optimal choice for the PredictEYE model.

Figures 6.9 and 6.10 present the ROC curves alongside the corresponding Area Under the Curve (AUC) values derived from the application of classification algorithms to LSTM and ARIMA model predictions, respectively. Notably, the ROC AUC values from the classification algorithms applied to LSTM-based predictions consistently exceeded those derived from ARIMA-based predictions. This underscores the superior discriminative capability of the LSTM model's predictions across various classification thresholds.

Among the evaluated classification algorithms, Random Forest consistently exhibited higher AUC values for most participants. The elevated AUC values, when paired with the exceptional performance of the Random Forest model, signify the LSTM model's potent predictive abilities in capturing significant patterns and features for classification tasks within this context.

In the extensive analysis of classification models across all participants, the evaluation of confidence intervals within the Receiver Operating Characteristic (ROC) curves, particularly emphasized in Figures 6.9 and 6.10, revealed critical insights. These confidence intervals are valuable indicators of the precision and confidence level associated with our model estimates. Notably, the Random Forest algorithm consistently presented narrow confidence intervals for numerous participants, reflecting highly precise estimations and a robust degree of confidence that the true values fell within this range. To offer a more detailed insight, Table 6.5 furnishes the confidence interval details for participants P3 and P4.

Remarkably, Participant P4 emerged as a noteworthy standout in this analysis. The Random Forest algorithm secured the highest classification accuracy for this individual, backed by the narrow confidence intervals. These results convey a high level of confidence in the model's capacity to provide precise and accurate predictions for Participant P4. In contrast, Participant P3, while not achieving the highest accuracy, benefited from the narrow confidence intervals. This suggests that even in cases of lower accuracy, the model's predictions maintained a high level of precision and confidence.

Table 6.4: Performance evaluation of PredictEYE with other models

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
PredictEYE (LSTM+RF)	**86.4**	83.9	88.8	**86.3**
LSTM+LR	84.1	84.8	81.5	83.1
LSTM+DT	82	79.9	83.4	81.6
LSTM+SVM	58.2	59.3	40.5	48.1
LSTM+KNN	57.7	56.6	50.2	53.2
LSTM+NB	75	69.9	83.9	76.3
ARIMA+RF	80.8	84.2	74.3	78.9
ARIMA+LR	81.5	82.7	78.2	80.4
ARIMA+DT	78.9	78.2	78.2	78.2
ARIMA+SVM	60.4	59.3	58.4	58.9
ARIMA+KNN	58.5	57.3	55.9	56.6
ARIMA+NB	48.9	48.6	95	64.3

Table 6.5: Confidence Intervals associated with ROC curves

Participant	Classification model	LSTM model with 95% of CI	ARIMA model with 95% of CI
P3	LR	75.96 to 83.63	57.44 to 67.45
	DT	66.79 to 75.46	51.27 to 61.48
	RF	**78.63 to 85.92**	**56.44 to 66.44**
	SVM	55.89 to 65.23	48.57 to 58.83
	KNN	55.37 to 64.72	48.38 to 58.64
	NB	62.86 to 71.82	52.21 to 62.38
P4	LR	91.44 to 96.03	82.14 to 88.89
	DT	78.74 to 85.97	75.58 to 83.34
	RF	89.68 to 94.76	**85.25 to 91.41**
	SVM	59.31 to 68.41	59.9 to 69.08
	KNN	53.41 to 62.76	55.89 to 65.27
	NB	77.61 to 85	71.25 to 79.52

6.3.3 Validation with GSR

GSR serves as a non-invasive physiological measure that effectively reflects the activity of the sympathetic nervous system and has found widespread utility in research focusing on emotional and cognitive processes [28]. GSR is a convenient and versatile tool for assessing emotional states across a variety of contexts. In comparison to alternative physiological measures, such as heart rate or EEG, GSR is less susceptible to motion artifacts and boasts a slower response time [203]. This feature makes GSR particularly valuable for validating models that aim to predict emotional states [204].

The performance of PredictEYE was validated by collecting GSR data from participants as they watched calm and stressful videos while their eye movements were tracked. In this study, an LSTM model was trained using this data to make predictions regarding future sequences of GSR data. Simultaneously, a Random Forest model was also trained to predict participants' emotional states based on the anticipated GSR sequences. The predictions generated by the Random Forest model, utilizing GSR data, were subsequently employed to validate the emotional state predictions derived from the eye tracking data.

6.4 Discussion

Figure 6.11 illustrates the classification of emotional states into 'calm' or 'stressful' using the Random Forest algorithm. This classification is predicated on the predicted sequences of eye tracking data collected after participants viewed both calm and stressful videos. The visual representation utilizes green to signify calm and red to denote stressful states. It offers a clear visualization of the states as determined by the GSR and the emotional state predictions generated by the LSTM and ARIMA models.

An interesting pattern emerges after scrutinizing the predictions of emotional states based on eye-tracking features and comparing them with GSR data. The Random Forest algorithm excelled in accurately predicting the emotional states of all participants when using predicted data sequences from the LSTM model. However, its performance was less consistent when attempting predictions based on the ARIMA model's forecasted sequences. Notably, the algorithm exhibited inaccuracies in predicting the emotional states of participants 3 and 5 when relying on ARIMA model predictions. These observations underscore the influence of the LSTM model's predictions on the accuracy of the Random Forest algorithm in the context of emotional state prediction through eye-tracking features.

A deeper analysis focused on Participant 3, who consistently displayed calm according to the ground truth GSR during both video sessions. When using the predicted sequence from the LSTM model, the Random Forest algorithm effectively predicted the participant's emotional state at the conclusion of the stressful video. However, based on the ARIMA model's forecasted sequences, it yielded less accurate predictions. The data underwent further scrutiny through a Welch two-sample t-test, comparing data samples obtained after viewing both calm and stressful videos. The results, documented in Table 6.1, revealed that specific fea-

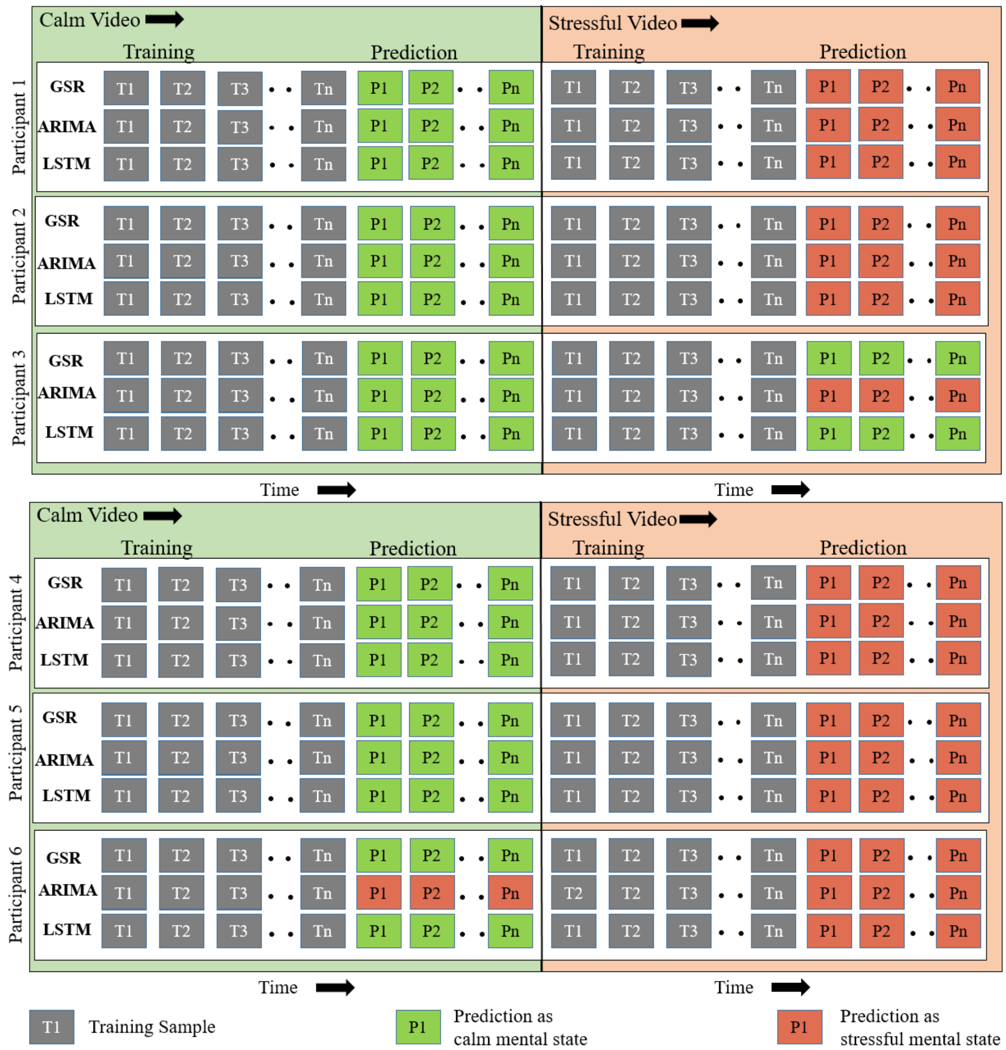


Figure 6.11: Forecasting of emotional state based on eye tracking features considering GSR as ground truth
T1, T2, ...Tn indicates the data at each time sequence used for training. P1, P2,..Pn indicates that data at each time sequence during prediction. The figure shows the emotional state prediction at the end of watching calm and stressful scenes of the video.

tures did not exhibit a significant difference at the end of the stressful video, signifying the participant's sustained calm state during the observation. These findings further emphasize the superior performance of the Random Forest algorithm in predicting emotional states when utilizing the predicted sequences generated by the LSTM model, as compared to ARIMA's predictions.

Turning to the analysis of Participant 6, Random Forest achieved precise emotional state predictions based on the forecasted sequences from the LSTM model for both videos. However, when relying on the predicted sequences from ARIMA data, the algorithm consistently classified the state as stressful for both videos, which contradicted the predictions based on GSR data. These results underscore the nuanced nature of the PredictEYE model's performance, which is dependent on the specific features used and the algorithm applied for data forecasting.

PredictEYE demonstrates the effectiveness of predicting participants' emotional states based on their emotional state while watching a video. The PredictEYE model utilizes LSTM to predict the future sequences of various eye tracking measures, such as pupil diameter, blink rate, and fixation, while participants watch specific content on a screen. These eye tracking measures provide insights into the participants' emotional states, whether they are moving towards a calm or stressful condition based on their emotional state while watching the video. For instance, when a participant is watching a stressful video, PredictEYE seeks to understand how their eye tracking metrics change over time and whether these changes lead them towards a calm or stressful state, driven by their emotional state while watching the video. The LSTM model plays a crucial role in this process by identifying patterns within each participant's eye tracking time series data and making predictions about their future eye movements and reactions. The Random Forest algorithm is then employed to collectively interpret the predicted future sequences of all eye tracking features, providing insights into the potential emotional state to which participants may be headed, considering their emotional state while watching the video. The insights gained from PredictEYE can be utilized to dynamically reorganize or skip the content being displayed to the participants, ensuring a more personalized and engaging experience based on their predicted emotional states, which are derived from their emotional states while watching the video. This approach can be applied in various domains, such as mental health and stress management, to monitor and predict individuals' emotional states in real time based on their physiological data and emotional state while watching a video.

Using time series analysis on eye tracking data with high sampling frequency can provide several benefits in predicting emotional states with personalized models. Firstly, it allows for identifying unique patterns of gaze behavior associated with different emotional states or disorders based on the emotional state while watching the video. Secondly, using time series analysis on eye-tracking data can help capture temporal dynamics and changes in emotional states over time. By capturing changes in gaze behavior and emotional states over time, personalized models can provide more accurate and timely predictions, allowing for more effective interventions and treatments.

In the PredictEYE model, eye tracking features play a significant role in predicting emotional states. The model's utilization of LSTM-based time series analysis on eye tracking data enables it to capture unique gaze patterns, fixations, and eye movements associated with different emotional states, such as calm or stressful, based on the emotional state while watching the video. By leveraging these eye tracking features, the PredictEYE model can distinguish individual behavioral patterns, making its predictions more accurate and tailored to the specific emotional states of each individual, influenced by their emotional state while watching the video.

Compared to the ground truth GSR, which predicts the emotional state, the PredictEYE model's eye tracking features offer additional information. While GSR provides valuable physiological data related to emotional state, eye tracking data goes beyond this by revealing what elements in the visual scene draw immediate attention and potentially influence the emotional state. The PredictEYE model not only detects the emotional state of an individual but also attributes the emotional state to specific scenes using the information obtained from eye tracking and the emotional state while watching the video. This feature allows for a more comprehensive understanding of the factors contributing to a person's emotional state during video viewing. By combining eye tracking and emotional state prediction, the PredictEYE model provides valuable insights into both conscious and subconscious responses, enhancing the accuracy of its predictions. PredictEYE model has been compared with existing recent personalized and not personalized models [88–91, 208, 209] in terms of stimulus, type of participants on which the study was performed, the features, algorithms used by the model, their achieved results, its analysis based on performance metrics, and type of the model as shown in Table 6.6. In eye tracking research, various stimuli, such as images, videos, tasks, and games, have been utilized to observe and analyze eye tracking measures. These measures typically include fixation, blink, saccade, and pupil diameter, employed in numerous studies to gain insights into different mental states based on the emotional state while watching the video.

The PredictEYE model focused on using video stimuli as the input and extracted features based on fixation, blink, and pupil diameter to classify emotional states as calm or stressful. Rather than comparing multiple users and attempting to understand the parameters responsible for emotional state prediction, our approach aims to observe and comprehend individual patterns, considering the idiosyncratic nature of eye tracking data.

Numerous models have adopted machine learning and deep learning algorithms to classify emotional states, and attentional states, emotional states, and identify mental disorders, as well as detect perceived workload. Among these models, PredictEYE stands out with its unique approach, utilizing LSTM-based time series data prediction and random forest algorithm to predict emotional states based on retrieved eye tracking data. The PredictEYE model falls under the personalized model category, intending to understand individual behavioral patterns based on their emotional state while watching calm and stressful videos. By learning from

Table 6.6: Comparison of PredictEYE with existing models

Model	Stimulus	Type of participant	Features	Algorithm	Result	Performance metric accuracy	Type of model
[199]	ST, AT, PT	–	PD, BR, FR,SR,ITD	CNN	Classification of attentional states	80%	Person-dependent and person-independent
[207]	SDS	Drivers	PD, BR, GD, VD,ED	CNN, LSTM	Classified the stress levels as low, medium and high	95.5%	Real time monitoring
[88]	V	ASD and TD children	FAA	DA	classification as ASD and TD	85.1%	NP
[89]	V	-	PC	IP, ML	Detection of psychomotor impairment	–	NP
[19]	AW	with cancer without cancer	VM	CNN-LSTM	Identified mental health status	Hope- 93.8% , Anxiety-94.8%, Mental Wellbeing-95%	Self-proclaimed
[90]	V	with mental disorders	PD, FR, SR, FD, FF	CNN, DLSTM	Provides objective evaluation index of patients with mental disorders	–	NP
[91]	I	Normal	SSIM, E, C, HOC, PSD for EEG, and EOG-PDE, CGF, FV, RMSF for ET	GMM, DGNN	Classified emotions under the eight event stimuli	88.10%	NP
[208]	SEOE	Excavator operators from industry	BR, BD, PD, GP	TICC, SVM	Mental fatigue detection	85%	NP
[209]	RSSS	Surgical trainees	PD, FD, GE, PERCLOS	NB	Detection of perceived workload in robotic surgical tasks.	84.7%	NP
PEYE	V	Normal	PD, FD, BD FDXY	LSTM, RF	Classified emotional state as calm and stressful	86.4%	Personalized model Indication of scene responsible for mental state

ST – switch task, AT – alignment task, PT – pairs task, SDS – stressful driving situations, V – video, AW – art works, I – images, SEOE – simulated excavator operation experiment, RSSS – robotics skill simulation session, PD – pupillometric data, BR – blink rate, FR – fixation rate, SR – saccade rate, FD – fixation duration, BD – blink duration, ITD –imaging time series data, GD – gaze dispersion, VD – vehicle data, ED – environmental data, FAA – fixation at Area of Interest(eyes,mouth, body, hands, objects, background), PC – pupil centroid, VM – visual metrics, FF – fixation frequency, SSIM – self-similarity, E – Energy, C – Complexity, HOC – High order crossing, PSD – power spectral density, EOG-PDE – electrooculography power density estimation, CGF – center gravity frequency, FV – frequency variance, RMSF – root mean square frequency, ET – eye tracking, GP – gaze position, GE – gaze entropy, PERCLOS – Percentage of eyelid closure, FDXY – fixaiton dispersion on X and Y axis, DA – discriminant analysis, IP – image processing, ML – machine learning, NB – Naive Bayes, RF – Random Forest, CNN – Convolutional Neural Network, DLSTM – Deep LSTM, GMM – Gaussian Mixed Model, DGNN – Deep Gradient Neural Network, TICC – Toeplitz Inverse Covariance-Based Clustering, TD – typically developing, NP – Not personalized, PEYE – PredictEYE these personalized patterns, the PredictEYE model predicts a person's emotional state based on their unique eye tracking responses. PredictEYE is unique in its approach as it analyzes time series eye tracking data, thoroughly understands the unique eye tracking features of that person, and predicts their emotional state and the specific scene responsible for it.

The performance of the PredictEYE model is compared to other models in terms of accuracy. The Random Forest model used in emotional state prediction has shown promising results with a maximum accuracy of 86.4%. At the same time, it could not achieve such high accuracy for all the participants, but it is unique in its approach in detecting mental states based on the emotional state while watching the video. Collecting more data over a longer period can help better understand the unique patterns of an individual's mental state, leading to more accurate predictions and improved mental health outcomes.

A stressful emotional state for Participant 1 might be attributed to scene 4, while for Participant 2, a different scene could be responsible for their emotional state, as shown in Figure 6.12. The figure illustrates the emotional state predictions of Participants 1, 2, and 3 while viewing a series of stressful scenes in a video. The depicted time span ranges from T1 to T18, with each scene labeled S1 to S18. Participant 1 experienced a state of stress, and this was attributed to scene S4, which had a noticeable impact on their emotional state. However, the same scene, S4, did not induce any changes in the emotional states of Participants 2 and 3. Participant 2, initially in a calm state, transitioned into a state of stress due to scene S8. In contrast, Participant 3 remained consistently calm throughout the entire time span, with no observed alterations in their emotional state caused by any scenes. These findings highlight the individual variability in how different participants respond to stressful stimuli and the unique triggers and reactions within their emotional states, all influenced by their emotional state while watching the video. This capability to capture and differentiate individual responses is an exceptional characteristic of PredictEYE showcasing the diverse ways in which people perceive and react to stressful situations.

PredictEYE is a tool that focuses solely on normal individuals and aims to understand changes in their emotional state by establishing a baseline period. During this baseline period, eye tracking measures are observed, and the model attempts to comprehend the trends and patterns of those measures to predict future emotional states. This approach allows for the development of personalized, data-driven interventions to support individuals' mental well-being, taking into account their emotional state while watching the video. By utilizing PredictEYE, individuals can gain insight into their emotional state and make informed decisions about their mental health care.

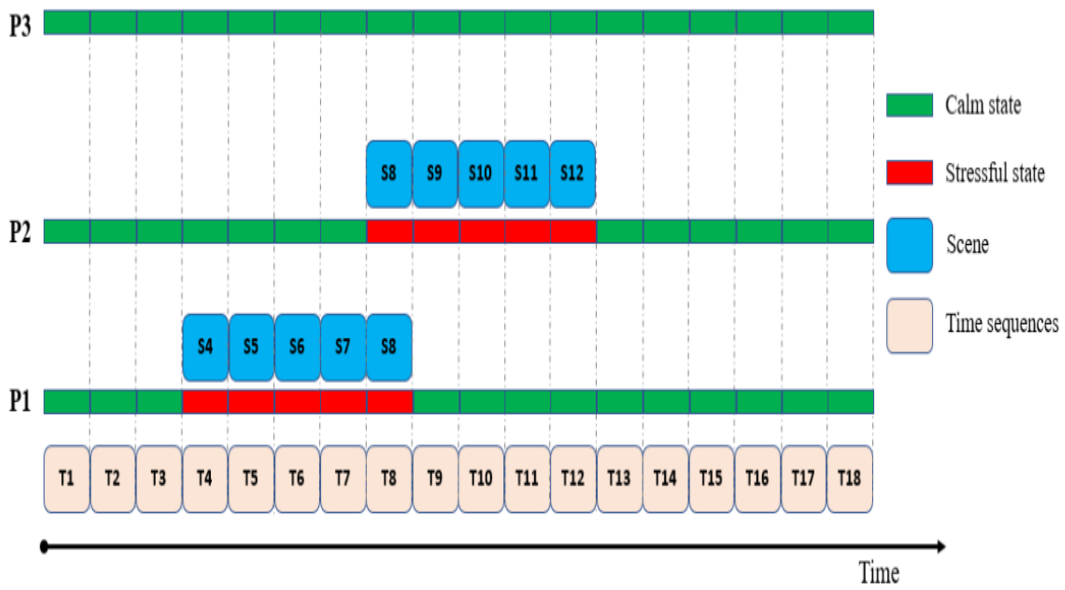


Figure 6.12: Sample scenes and gaze responsible for the emotional state of participants

P1- Participant 1, P2- Participant 2, P3- Participant 3, T1, T2..T18 - Time sequences, S4, S5, ...S12 - Scene responsible for the stressful emotional state.

One of the key advantages of using time series data analysis in the PredictEYE model was the ability to develop personalized models for each participant. By analyzing the eye tracking data during both calm and stressful video viewing, the model was able to identify the underlying patterns in each participant's data and develop a personalized model that could efficiently predict their emotional state, considering their emotional state while watching the video.

The PredictEYE model customizes its analysis of eye tracking data by utilizing LSTM-based time series models, which adapt to individual differences in a personalized manner, taking into account their emotional state while watching the video. This personalized approach involves training the model on each person's specific eye tracking data, capturing their idiosyncratic patterns and responses. Instead of comparing data across multiple participants and treating them as a homogeneous group, this personalized approach recognizes and respects the individuality of each person's cognitive and emotional processes.

The results demonstrated that the PredictEYE model could accurately predict each participant's emotional state based on their eye tracking data. Time series data analysis and personalized modeling in the PredictEYE model could be applied to larger datasets in future studies to improve the accuracy and reliability of the model for emotional state classification. PredictEYE's continuous monitoring has the potential to identify patterns in an individual's emotional state over time, revealing insights into stress and mental health conditions. These insights could facilitate early interventions or treatments before conditions worsen.

During the development of the PredictEYE system, the challenge was to build a personalized model. Extensive literature surveys led to the discovery of suitable time series analysis methods for eye tracking data, enhancing personalization, and improving emotional state prediction. Selecting the best predictive and classification model combination is crucial in developing accurate and efficient personalized models like the PredictEYE model, which could predict participants' emotional states based on their eye tracking data while taking into account their emotional state while watching the video. By selecting the LSTM and the Random Forest models as the best models for predicting the emotional state, the PredictEYE model could understand the trends in the time series data and classify participants' emotional states into calm or stressed states with better performance.

6.5 Conclusion

The PredictEYE system predicts a person's emotional state based on their emotional state while watching videos using eye tracking data. It utilizes an LSTM-based time series regression model for forecasting and a Random Forest algorithm-based classification model for predicting emotional states. Comparing the LSTM model with an ARIMA model, the LSTM outperformed. Random Forest achieved a maximum accuracy of 86.4%, precision of 83.9%, recall of 88.8%, and an F1 score of 86.3% in emotional state prediction. Eye tracking features played a significant role, similar to ground truth GSR. The model can incorporate various physiological signals for improved accuracy.

PredictEYE offers promise in predicting emotional states, providing insights into specific scenes affecting individuals' emotional states. It has applications in mental health screening and treatment monitoring. Its adaptability for webcam-based eye tracking allows continuous, non-invasive monitoring, offering insights into emotional states over time. The model's accuracy can be enhanced by parameter tuning and multivariate data analysis. Incorporating reinforcement learning can further personalize emotional state predictions, improving outcomes for mental health concerns. The adaptable and non-invasive nature of PredictEYE makes it valuable in healthcare, education, and employment settings.

This chapter explained one of the mental state parameters, emotional state, and introduced the PredictEYE time series model for the future feature prediction and its classification. The next chapter will explain the models that assess an individual's mental state based on their emotional state using eye-tracking data and other physiological measures.

Chapter 7

Multimodal Dataset Creation and the Detection of Emotional state

7.1 Representation of Emotional State on a Dual-Dimensional Space

Emotion is a complex phenomenon and it can be assessed effectively through the dual-dimensional valence and arousal model proposed by Russell [210]. This model, depicted in Figure 7.1, provides a detailed understanding of the subjective experience of emotions. Valence, which defines the pleasantness or unpleasantness of an emotional experience, includes the general emotional polarity and is divided into positive and negative. At the same time, the concept of arousal, describing the intensity level connected with feelings, is divided into low and high, which indicate the level of activation. Emotional states are represented in two different axes and that creates space for various combination of emotions. For instance, emotions with negative valence and high arousal, such as anger, are positioned in the upper left corner, while emotions with positive valence and low arousal, like calmness, are in the lower right corner.

Thus representation allows a thorough understanding of emotions by capturing the overall character and intensity of their emotions. Valence and arousal affect the cognitive processes and behaviours through different neural mechanisms in the brain. The inclusion of this concept into mental health assessments allows a more enhanced and specific approach in the assessment of an individual's emotional health.

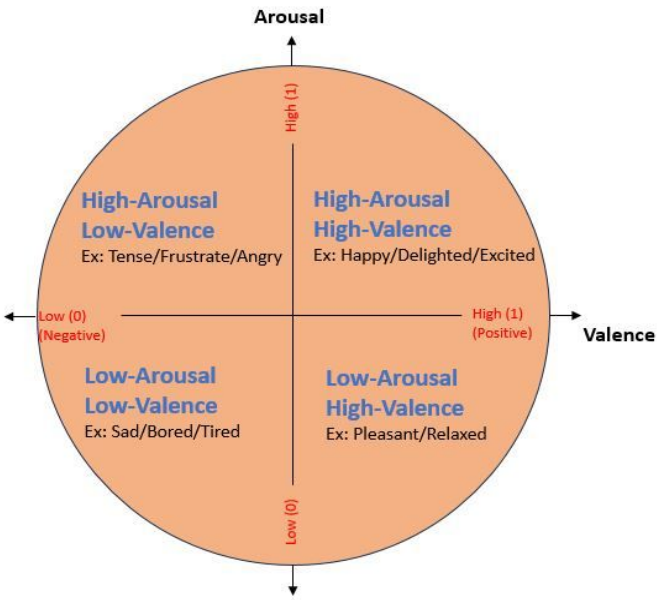


Figure 7.1: Dual-dimensional model of valence and arousal

7.2 Need of Multimodal Dataset

Emotions are complex and involve physiological processes and actions. Eye tracking alone may only provide a partial view. A multimodal approach, combining eye tracking data, physiological signals, and behavioral markers, supports and confirms results, considering individual differences and interpretations of emotional reactions. This method enhances the accuracy and reliability of systems designed to detect emotional states by considering multiple channels of data, such as face, voice tone, body language, and physiological reactions [139].

Researchers developed an emotion recognition algorithm using heart rate data from a wearable smart bracelet and 25 participants watching emotional videos [211]. Deep neural network models like EmoRL improved accuracy and reduced latency in categorizing emotions based on audio input [212]. Researchers also used respiratory signals to analyze emotional information, classifying arousal and valence into categories [142].

These studies bring out the need for a multimodal dataset in emotional state detection. The integrating data of physiological signals including the heart rate as well as the respiratory signal gives a better view of the emotion a person is experiencing. The combination of different modalities allows the algorithms to detect more types of emotion, and thus make the emotion recognition system more reliable. Hence, to enhance the efficiency of the detection of the state of emotions, it becomes essential to employ the datasets of the multimodal type.

In section 7.3, the process of emotional state detection based on the multimodal dataset EmoRPhyE is explained. The features extracted from each physiological signal are discussed in section 7.3.1, data labeling is discussed in section 7.3.2, the classification model is employed in section 7.3.3, and a detailed analysis of the results is provided in section 7.3.4. Furthermore, the discussion of the model is presented in section 7.3.5 and concludes with section 7.3.6. In section 7.4, all the models specific to cognitive load, cognitive impairment, and emotional state classification are compared with the state of the art.

7.3 EmoRPhyE Dataset and the Models to Detect the Emotional State

EmoRPhyE multimodal dataset was created by incorporating heart rate, blood flow, skin conductance, breathing, and eye tracking data while participants viewed both pleasant and unpleasant images. Extensive literature review was conducted on the identification of the various emotional states. The process of creating the detailed dataset is described in Chapter 3.

Building upon the EmoRPhyE dataset, two distinct models were developed to detect emotional states, as shown in Figure 7.2. The first model utilizes eye tracking data to observe emotional responses based on visual attention patterns. The second model integrates a broader spectrum of physiological measures, including ECG, PPG, GSR, and respiratory data. This gives a more refined possibility to study the relationship between body reactions and visual attention to emotion-related stimuli.

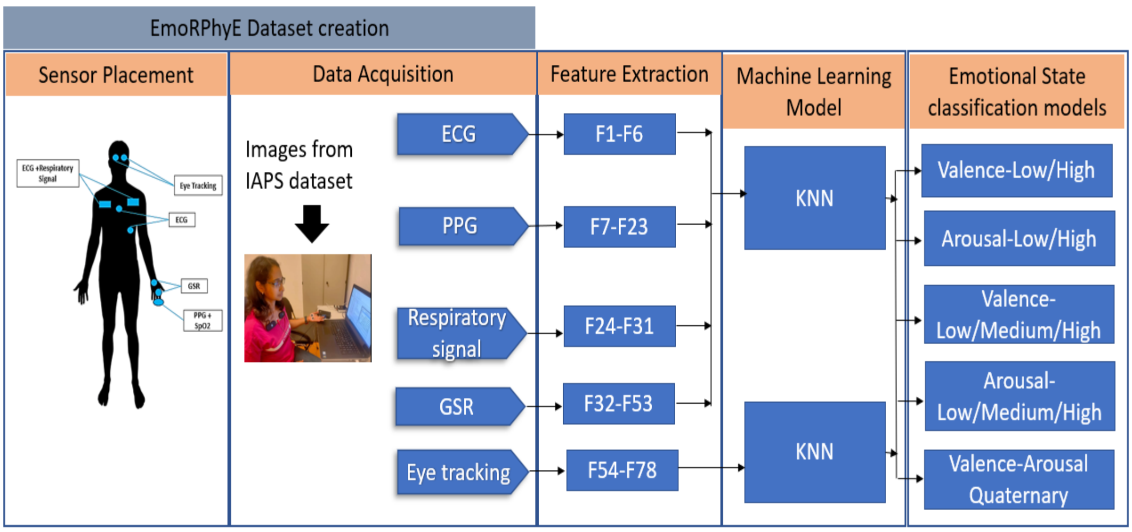


Figure 7.2: System architecture for emotional state classification
IAPS- International Affective Picture System, ECG - Electrocardiogram, PPG - Photoplethysmography, GSR - Galvanic Skin Response, KNN - K Nearest Neighbour

The EmoRPhyE dataset is a valuable resource for emotion detection, providing a variety of physiological signals to understand emotional states. It allows for comparative analysis of eye tracking data effectiveness versus multiple physiological measures. This research contributes to multimodal approaches in emotion detection, highlighting the potential synergy between visual attention and physiological responses.

7.3.1 Feature Extraction

The EmoRPhyE dataset, containing physiological data from 30 participants, was used to extract features from 192 images displaying varying levels of arousal and valence. Only time domain was considered for extracting features from the physiological signals. Since the focus was on capturing temporal patterns and variations within the signals, the frequency-specific characteristics in the process of physiological feature extraction were neglected. Each stimulus image was presented to the participant for a duration of 8 seconds, and considered as the window period for feature extraction.

Table 7.1: Features extracted from ECG signal

Feature	**Description**
Avg_HR (F1)	The mean of heart beats
Avg_RRinterval (F2)	The mean time duration between successive R-peaks in an ECG
SD_RRinterval (F3)	Dispersion of RR intervals.
RMSSD (F4)	Root mean square distance between successive RR intervals
Count_Rpeak_50 (F5)	Count of ECG R-peaks differing by more than 50 ms.
Percent_Rpeak_50 (F6)	Percentage of ECG R-peaks differing by more than 50 ms.

The R peaks were detected for the ECG signal using a modified version of the Pan-Tompkins algorithm. Subsequently, the RR interval (RRI) and the heart rate (HR) time series were derived for each window. A total of six features (F1-F6) were then extracted from the ECG signal, as detailed in Table 7.1. These features encompass metrics such as average heart rate, average RR interval, standard deviation of RR interval, root mean square distance between successive RR intervals, and the count and percentage of R peaks differing by more than 50 milliseconds. The Average Heart Rate represents the mean number of heartbeats per minute, offering a snapshot of the heart's overall pacing. The Average RR Interval provides an average measure of the time lapse between two consecutive heartbeats. The Standard Deviation of RR Interval serves as a metric for the variability or dispersion of RR intervals, indicating the consistency of the heart rate. Higher values denote increased variability. The RMSSD is a time-domain measure that quantifies the variation in consecutive RR intervals. It reflects short-term heart rate variability, with higher RMSSD values suggesting greater variability and potential parasympathetic nervous system influence. The count and percentage of R Peaks Differing by More than 50 Milliseconds are indicative of irregularities in the ECG signal. This feature quantifies the number or proportion of R peaks exhibiting a difference greater than 50 milliseconds, with a higher count and percentage pointing to increased variability in R-peak timings.

Seventeen features(F7-F23) were extracted from each PPG signal window, focusing on the detection of critical points, including valleys, systolic, diastolic, and dicrotic notches. Among these features, thirteen were timing-related, as shown in Table 7.2, with additional features such as kurtosis, skewness, the standard deviation of time between systolic peaks, and the systolic-diastolic phase ratio. The Systolic Peak is the highest pressure point during a heartbeat, reflecting the peak contraction of the heart. Diastolic Point is the lowest pressure point between heartbeats, indicating the relaxation of the heart. Dicrotic Notch is a small dip in the descending part of the pulse wave, typically following the systolic peak. Valley is the minimum point in the pulsating waveform between the diastolic point and the next systolic peak.

Regarding the respiratory signal, peaks, and valleys were extracted for each window, and the breath rate series was derived as the time differences between these peaks. The respiratory signal processing resulted in eight features (F24-F31), as detailed in Table 7.3. These features encompass parameters such as average breath rate, standard deviation of breath rate, root mean square distance of successive breath intervals, standard deviation of successive breath interval differences, average expiration time, average inspiration time, average expiration area, and average inspiration area.

GSR data yielded 22 features (F32-F53), as inspired by studies by Daniel Lopez-Martinez and Alberto Grego, and are shown in Table 7.4.

The study utilized the I-DT algorithm for the detection of fixation and saccade coordinates and identifying the blink events. Fixations were determined based on spatial and temporal proximity, with a maximum Euclidean distance of 100 pixels and a minimum time duration of 50 ms. Saccades, rapid eye movements between fixations, were identified by larger spatial distances and shorter temporal duration, with a cutoff total duration not exceeding 90 milliseconds. A total of 25 eye-tracking features (F54-F78) were extracted as shown in Table 7.5, encompassing parameters such as blink frequency, fixation duration, saccade characteristics, and pupil diameter metrics.

Table 7.2: Features extracted from PPG signal

Feature	**Description**
Pulse_propagation_time (F7)	duration between specific points in the pulsatile cycle
Systolic__systolic (F8)	duration between two successive systolic peaks
Diastolic_to_valley (F9)	duration between the diastolic point and the subsequent valley
Valley_to_diastolic (F10)	duration between subsequent valley and the diastolic point
Diastolic_to_diastolic (F11)	duration between two consecutive diastolic points
Systolic_to_dicrotic notch (F12)	duration between the systolic peak and the dicrotic notch
Dicrotic_notch_to_diastolic (F13)	duration between the dicrotic notch and the subsequent diastolic point.
Systolic_phase (F14)	Rising curve from valley to systolic peak
Diastolic_phase (F15)	descending curve from systolic peak to valley
Pulse_wave_duration (F16)	overall time duration of the pulsatile waveform
Valley_to_dicrotic_notch (F17)	duration between valley and the subsequent dicrotic notch
Dicrotic_notch_to_valley (F18)	duration between dicrotic notch and the subsequent valley
Dicrotic_notch_to_dicrotic_notch (F19)	duration between two consecutive dicrotic notches
Kurtosis (F20)	sharpness or flatness of a distribution
Skewness (F21)	asymmetry in a distribution
SD_systolic_peaks (F22)	standard deviation of duration between successive systolic peaks
systolic_diastolic_phase_ratio (F23)	proportion of time spent in heart contraction versus relaxation.

The feature “percent_interest” quantifies the proportion of interest within a designated area determined by the visual saliency of the image [213]. Computed as the percentage of fixations durations occurring within this area, derived based on the Equation 7.1. This metric measures the relative attention or interest directed toward the visually salient portion of the image, offering valuable insights into the emotional response associated with the observed image. The details of all other eye tracking features are provided in Chapter 4.

Table 7.3: Features extracted from respiratory signal

Feature	Description
Avg_breath_rate (F24)	Mean of breaths per unit of time
Std_breath_rate (F25)	standard deviation of breath rate
RMSE_breath_interval (F26)	root mean square distance of successive breath intervals
Std_breath_interval (F27)	standard deviation of successive breath interval differences
Avg_exp_time (F28)	The average expiration time
Avg_ins_time (F29)	average inspiration time
Avg_exp_area (F30)	average expiration area
Avg_ins_area (F31)	Average inspiration area

Table 7.4: Features extracted from GSR signal

Feature	Description
Max_gsr (F32)	Maximum of the GSR signal
Range_gsr (F33)	Range of the GSR signal
Std_gsr (F34)	Standard deviation of the GSR signal
IQ_gsr (F35)	Interquartile range of the GSR signal
RMS_gsr (F36)	Root mean square of the GSR signal
Avg_gsr (F37)	Average of the GSR signal
Avg_first_gsr (F38)	Average absolute value of the first differences of the GSR signal
Avg_second_gsr (F39)	Average absolute value of the second differences of the GSR signal
Avg_first_norm_gsr (F40)	Average absolute value of the first differences of the standardized GSR signal
Avg_second_norm_gsr (F41)	Average absolute value of the second differences of the standardized GSR signal
Skw_gsr (F42)	Skewness of the GSR signal
Krt_gsr (F43)	Kurtosis of the GSR signal
SumPosDiff_gsr (F44)	Sum of positive values of the first derivative of the GSR signal
SumNegDiff_gsr (F45)	Sum of negative values of the first derivative of the GSR signal
Max_scl (F46)	Maximum of the tonic curve
Max_scr (F47)	Maximum of the phasic curve
AUC_scl (F48)	Area under the tonic curve
AUC_scr (F49)	Area under the phasic curve
Avg_scl (F50)	Average of the tonic curve
Avg_scr (F51)	Average of the phasic curve
Std_scl (F52)	Standard deviation of the tonic curve
Std_scr (F53)	Standard deviation of the phasic curve

Table 7.5: Eye tracking features extracted for the classification of emotional state

Feature	Description
percent_ interest (F54)	Percentage of interest within an area of interest.
blink_freq (F55)	Frequency of blinks within a time frame
avg_ blink_dur (F56)	Average of all the duration of blinks for an image
max_blink_dur (F57)	Maximum blink duration while watching an image
min_blink_dur (F58)	Minimum blink duration while watching an image
time_to_first_fix (F59)	Time difference between starting time of first fixation in an image and starting time of an image.
first_fix_dur (F60)	First fixation duration while watching an image
avg_fix_dur (F61)	Average fixation duration while watching an image
fix_freq (F62)	Frequency of fixation within a time frame
fix_max_dur (F63)	Maximum fixation duration
fix_min_dur (F64)	Minimum fixation duration
avg_pup_dia (F65)	Average of all the pupil diameter values for an image
max_pup_dia (F66)	Maximum pupil diameter
min_pup_dia (F67)	Minimum pupil diameter
avg_sac_dur (F68)	Average saccade duration while watching an image
min_sac_dur (F69)	Minimum saccade duration
max_sac_dur (F70)	Maximum saccade duration
avg_sac_amp (F71)	Distance between starting and ending points saccade
avg_sac_vel (F72)	Velocity of the saccade
min_sac_amp (F73)	Minimum saccade amplitude while watching an image
max_sac_amp (F74)	Maximum saccade amplitude while watching an image
min_sac_vel (F75)	Minimum saccade velocity while watching an image
max_sac_vel (F76)	Maximum saccade velocity while watching an image
min_velocity (F77)	Minimum saccade velocity while watching an image
sac_freq (F78)	Frequency of saccades within a time frame

$$\text{percent\_interest} = \left(\frac{\text{Total duration of fixations within AOI}}{\text{Overall time spent for the image}}\right) \times 100 \tag{7.1}$$

The collected data were annotated using the feedback from the users. With the help of Self-Assessment Manikin method (SAM), participants estimated the valence and the arousal of each picture. At the end of each four pictures, participants had to give ratings on a scale of 1 to 9. For valence, a rating of 1 was low and 9 was high. Likewise, 1 for arousal meant low arousal and 9 meant high arousal. This structured approach let the participants to report their subjective estimates, which was aimed at defining the amount of the emotional load of the presented images.

7.3.2 Data Labelling

Considering combinations of valence and arousal as low, medium, and high, five classification models were established. The given classification offers a clear view of understanding of different emotional states, providing insights into the variations across the emotional spectrum.

1. Binary classification for low and high arousal: Data is labeled as low or high arousal.
2. Binary classification for low and high valence: Data is labeled as low or high valence.
3. Ternary classification for low, medium, and high arousal: Data is labeled into low, medium, or high arousal.
4. Ternary classification for low, medium, and high valence: Data is labeled into low, medium, or high valence.
5. Quaternary classification, where the arousal-valence plane is divided into four quadrants, each corresponding to a different combination of valence and arousal. Data is labeled into low Valence low arousal, low valence high arousal, high valence low arousal, high valence high arousal

Specific threshold levels for valence and arousal were employed to categorize these distinct groups, following the criteria detailed in Table 7.6.

Table 7.6: Labeling the data based on valence and arousal feedback

Classification	**Arousal**	**Valence**	**Threshold**
Binary-Arousal	Low	-	arousal<5
	High	-	arousal>=5
Binary-Valence	-	Low	valence<5
	-	High	valence>=5
Ternary-Arousal	Low	-	arousal<4
	Medium	-	arousal>=4 and arousal<=6
	High	-	arousal>6
Ternary-Valence	-	Low	valence<4
	-	Medium	valence>=4 and valence<=6
	-	High	valence>6
Quaternary	Low	Low	arousal<5 and valence<5
	Low	High	arousal<5 and valence>=5
	High	Low	arousal>=5 and valence<5
	High	High	arousal>=5 and valence>=5

Utilizing the same framework, physiological features and eye-tracking features underwent classification, even though the classification was conducted separately on the two datasets.

7.3.3 Classification Model

The proposed approach involves two models for emotion classification: one relies on the machine learning algorithm KNN applied to eye tracking data, and the other incorporates various physiological measures, including ECG, GSR, respiratory signals, and PPG. The primary objective is to assess the effectiveness of each model in categorizing emotions based on the features extracted from their respective datasets.

The dataset underwent division into training and testing subsets to assess the model's performance on distinct data. The training set, randomly comprising 80% of the dataset, played a crucial role in establishing relationships between physiological (or eye-tracking) features and emotional states categorized by various levels of valence and arousal. This partition exposed the algorithm to diverse emotional data, enabling it to identify meaningful correlations and enhance predictive capabilities. The remaining 20% of the data constituted the testing set, ensuring the model encountered unseen instances during evaluation. This separation guarded against overfitting and guaranteed the model's ability to generalize emotions accurately.

7.3.4 Results

The classification models of the human emotional state with the help of physiological signals and eye tracking data are developed using KNN algorithm. Five classification models that use arousal and valence dimensions have been created, and the performance of each of the created models is being compared. The arousal model and the valence model are two models that reflect different possibilities of arousal and valence; thus, each model offers a great opportunity to explore different emotional states and get a deep understanding of the human emotions.

To compare the efficiency of the models under consideration, a comparative analysis has been made. This entails comparing the predictions that are obtained from the eye tracking model with those that are obtained from the physiological measures model. The objective is to determine the accuracy of the models in identifying an individual's emotional state. The results are presented in Tables 7.7 and 7.8, providing insights into the performance of the eye tracking model and the physiological measures model across different emotional categories.

Table 7.7: Performance evaluation of emotional states based on KNN algorithm applied on eye tracking data

Emotional state classification models	**Accuracy (%)**	**Precision (%)**	**Sensitivity (%)**
Binary- Arousal	54	Low-55	Low-56
		High-55	High-53
Binary- Valence	57	Low -57	Low-64
		High-57	High-50
Ternary-Arousal	49	Low-45	Low-34
		Medium-52	Medium-55
		High-49	High- 58
Ternary-Valence	50	Low-50	Low-50
		Medium-50	Medium-32
		High-50	High-67
Quaternary	53	LALV-59	LALV-81
		LAHV-44	LAHV-26
		HALV-52	HALV-40
		HAHV-54	HAHV-67

LALV – Low Arousal-Low Valence, LAHV – Low Arousal-High Valence, HALV – High Arousal-Low Valence, HAHV – High Arousal-High Valence

Table 7.7 provides an analysis of the performance of the given models for the classification of the emotional state using the eye tracking data with the help of the KNN algorithm. Thus, when comparing binary, ternary, and quaternary models, it is seen that binary models are characterized by higher accuracy, precision, and sensitivity in comparison with ternary and quaternary models. Particularly, in the context of the binary classification framework, it can be stated that the Binary-Valence model performs better than the Binary-Arousal model, based on the higher accuracy, precision, and sensitivity values for both low and high arousal levels. Ternary and quaternary models give finer distinction to the emotions but at the same time they are less accurate and sensitive due to the complexity involved in them.

Precision signifies the model's accuracy in correctly identifying instances of a specific emotional state among all instances it classifies as that state. The model achieved a precision of 57% in classifying emotional states as low and high based on valence, indicating that 57% of the instances predicted as low were genuinely low, and similarly, 57% of the instances predicted as high were genuinely high. Precision delineates the proportion of correct identifications within the model's classifications for each emotional state.

Sensitivity, or recall, measures the model's ability to correctly identify all instances of a specific emotional state among those present in the dataset. The model could achieve a sensitivity of 64% in classifying emotional states as and low

50% in classifying as high, indicating that the model correctly identified 64% of low instances and 50% of high instances.

Table 7.8 presents the performance evaluation of emotional state classification models based on physiological measures, employing KNN algorithm. The binary classification method emerges as the superior option for emotional state classification based on physiological measures among binary, ternary, and quaternary classifications. This preference is based on its simplicity and often higher accuracy than ternary and quaternary classifications. Furthermore, within the chosen binary classification, the Binary-Arousal model achieves an accuracy of 65%. This signifies its effectiveness in accurately discerning and categorizing emotional states, particularly in distinguishing between low and high arousal levels. Therefore, the binary classification, supported by physiological signals, emerges as the optimal choice for precise emotional state detection.

The precision scores indicate that the model is more accurate in predicting instances of low arousal (69%) compared to instances of high arousal (62%). Similarly, the sensitivity scores suggest that the model is better at correctly identifying instances of low arousal (60%) than instances of high arousal (54%). While the model performs relatively well in detecting low arousal, its accuracy in identifying high arousal instances may require improvement.

Table 7.8: Performance evaluation of emotional states based on KNN algorithm applied on various physiological data

Emotional state classification models	**Accuracy (%)**	**Precision (%)**	**Sensitivity (%)**
Binary- Arousal	65	Low-69	Low-60
		High-62	High-54
Binary- Valence	56	Low -56	Low-64
		High-56	High- 48
Ternary-Arousal	48	Low-47	Low-60
		Medium-46	Medium-40
		High-52	High- 53
Ternary-Valence	45	Low-42	Low-50
		Medium-48	Medium-35
		High-45	High-50
Quaternary	37	LALV-31	LALV-48
		LAHV-43	LAHV-31
		HALV-45	HALV-28
		HAHV-37	HAHV-45

LALV – Low Arousal-Low Valence, LAHV – Low Arousal-High Valence, HALV – High Arousal-Low Valence, HAHV – High Arousal-High Valence

When comparing the performance of both models for emotional state detection based on eye tracking and physiological measures, it was observed that the

Binary-Arousal model exhibited superior classification accuracy when utilizing physiological data. Conversely, all other models displayed enhanced performance when utilizing eye tracking data for emotional state detection. However, the models' overall performance could be improved by integrating both eye tracking and physiological measures, leveraging the complementary information provided by each modality.

7.3.5 Discussion

The study introduces EmoRPhyE, a novel multimodal dataset includes synchronized physiological data and eye gaze data collected while viewing emotion-evoking images. This dataset, uses objectively scored feedback, attempts to improve consistency in emotional research across different domains. With an emphasis on crucial elements including stimulus selection, duration, and trial design, the study highlights the difficulty of identifying emotional states in response to stimuli with low emotional effect.

The stimuli were divided into 48 groups, each with 4 images that have same valence and arousal levels and are displayed randomly to ensure a dynamic experience. The study utilized an emotion assessment approach to explore the usability of a dataset, analyzing physiological signals and eye-tracking data separately, and extracting relevant information.

The EmoRPhyE dataset is a unique combination of subject count, stimulus variety, and physiological signal diversity as shown in Table 7.9. It includes data from 30 subjects, a competitive number compared to other datasets like DEAP [214] and AMIGOS [215]. The dataset uses 192 images from the IAPS dataset, a controlled emotional response, and various physiological signals, including ECG, PPG, GSR, respiratory signals, and eye tracking data. It also captures valence and arousal feedback from participants.

The emotional state detection models, utilizing the KNN algorithm and incorporating both eye tracking data and physiological measures such as ECG, respiratory signal, GSR, and PPG, demonstrated comparable performance levels. Both models showed competence in binary classification based on arousal and valence, outperforming ternary and quaternary models. Notably, the Binary-Valence model achieved marginally better accuracy with eye tracking data, while other physiological measures exhibited stronger classification performance for the Binary-Arousal model.

Classification results indicated promise, particularly in binary and ternary classifications, but highlighted sensitivity challenges, especially for high arousal and high valence states in physiological signals. The approach using both physiologi-

Table 7.9: Overview of multimodal datasets

Dataset	**Subject**	**Stimulus**	**Physiological signals**	**Feedback from participant**
Deap [214]	32	120 videos	GSR, skin temperature BVP, Respiration pattern, EMG, EOG, EEG, frontal video, audio signals	Valence, arousal, dominance, liking, familiarity
MAHNOB_HCI [215]	27	20 emotional videos	GSR, EEG, ECG respiration amplitude, skin temperature, eye gaze data	valence, arousal on ternary scale
SEED [216]	15	Videos of 4 min duration	EEG signals	negative, neutral, positive
AMIGOS [217]	40	16 short emotional videos and 4 long videos	EEG, ECG, GSR,	valence, arousal, control, familiarity, liking
EmoRPhyE	30	192 images from IAPS datset	ECG, PPG, GSR, Respiratory signals, Eye Tracking Data	Valence, Arousal

cal and eye-tracking data showcased balanced performance across different emotional states. However, there is room for improvement, particularly in refining and optimizing models for better accuracy. This study presents a significant contribution by demonstrating the potential benefits of combining multiple modalities to understand emotional responses comprehensively. Limitations include focusing on time-domain features in physiological signals, suggesting avenues for future research involving more specific features and advanced classification algorithms, and potentially utilizing deep learning techniques for enhanced model performance.

The model is a versatile framework, employing a foundational machine-learning algorithm for classification tasks. Improvement opportunities lie in replacing the existing classification algorithm with more advanced alternatives. Additionally, the model's flexibility allows for tailored training according to the specifics of available data across various applications, enabling classification into multiple classes as per the specific requirements of each scenario.

7.3.6 Conclusion

This study carries significant implications for advancing emotion assessment, spanning bioengineering and cognitive sciences. The curated database is a valuable resource, providing a robust foundation for algorithm development in emotion assessment through machine learning or deep learning techniques. Incorporating multimodal and synchronized physiological signals, along with subjective and objective data collection, makes this database particularly promising for investigations in cognitive science, offering insights into psychological aspects.

The database and algorithm's potential in enhancing emotion assessment tasks is evident through the use of physiological signals. Further exploration of the database and AI techniques could improve objective assessment of emotions from diverse signals. The emotional state detection model is innovative and holds promise for understanding and evaluating human emotions objectively. Future research could involve integrating eye tracking with other physiological measures to further enhance accuracy and depth of emotion assessment tasks.

7.4 Comparison of the Proposed Work with the Existing Models

The presented research addresses significant gaps in mental health assessment using eye gaze tracking. It fills the void of publicly available datasets and proposes innovative models for predicting cognitive load, cognitive impairment, and emotional state. The inclusion of multimodal data and physiological signals improves the understanding of mental health states. This work provides valuable resources for benchmarking, replicability, and collaboration among researchers, emphasizing the potential of eye tracking technology in mental health assessment.

The presented research addresses significant gaps in mental health assessment. It addresses the lack of publicly available datasets and presented new paradigms for estimating cognitive load, cognitive decline, and mood state. The interpretation of mental health states could be enhanced with the inclusion of multimodal physiological data. This study highlights the importance of eye tracking technology in mental health assessment, promoting collaboration among researchers and sharing benchmarks for replication. Table 7.10 shows the comparison with existing datasets.

Based on the research gap specified earlier, Table 7.11 presents novel features like error rate, scanpath comparison score, and inattentional blindness score, which differ from traditional models and significantly enhance understanding of cognitive functions, especially visual search speed and focused attention, providing a more

Table 7.10: Comparison with existing datasets

Dataset		Mental state parameter	Data
COLET [218]		Cognitive Load	Eye tracking data
CL-Drive [219]		Cognitive Load	EEG, ECG, EDA, eye tracking data
Deap [214]		Emotional State	GSR, BVP, Respiration pattern, skin temperature, EMG, EOG, EEG, frontal video, audio signals
MAHNOB_HCI [215]		Emotional State	GSR, respiration amplitude, skin temperature, EEG, ECG, eye gaze data
SEED [216]		Emotional State	EEG
AMIGOS [217]		Emotional State	EEG, ECG, GSR
Proposed Thesis Work	ET_MT_CL [49]	Cognitive Load	Pupil diameter, Blink count
	ET_TMT_CL	Cognitive Load	35 features based on fixation, saccade, blink, pupil diameter and time
	ET_TMT_CI [14, 15]	Cognitive Impairment	Low-level fixation based features, Middle -level AOI based features, High-level features- Error rate, Scanpath comparison score, Inattentional Blindness Score
	ET_Video_ES [36, 38]	Emotional State	Fixation, Blink and Pupil diameter based 5 features
	EmoRPhyE (Multimodal)	Emotional State	ECG, PPG, GSR, Respiratory signals, Eye Tracking Data

insightful perspective on cognitive impairment.

The research proposes the ETMT model to address the lack of assistive tools for healthcare professionals in mental health assessments. Compared to traditional TMT, the ETMT model is more accessible and efficient, making it particularly useful in clinical and residential settings with a shortage of mental health professionals. Comparisons are provided in Table 7.12. The ETMT model provides a comprehensive view of cognitive functioning by extracting 13 distinct features,

Table 7.11: Comparison with the features extracted

Model	Stimuli	Features
[120]	Reading short paragraphs and answering questions based on it	Saccade and fixation in each AOI, fixation duration, saccade amplitude, first fixation duration in an AOI
[119]	Word Memory Test	Saccade frequency, pupil size, dwell rate in an AOI
[106]	Memory task, attention task and calculation task	Fixation duration within the AOI, saccade frequency, smooth pursuit
[35]	Memory and attention tasks	Fixation duration
[220]	Digit span task, Spatial span task	Fixation, saccade frequency, saccade latency
ETMT [15, 36]	Eye tracking based TMT	Fixation based features AOI based features High level features- **Error rate**, **Scanpath comparison score**, **Inattentional blindness score**

unlike the traditional TMT which focuses on task completion time. This helps healthcare professionals conduct accurate assessments, bridging the research gap and enabling more accurate cognitive health assessments.

Table 7.12: Comparison of ETMT with Traditional TMT

Tests	Low cost	No specialized training to administer	Easy to detect in early stages	shorter time duration for the tests	Aable to address lower motor neuron atrophy.	Does not stress the participant
ETMT [15, 36]	✓	✓	✓	✓	✓	✓
Traditional TMT [109]	✓	X	✓	✓	✓	✓

The integration of eye tracking with mental health indicators and the development of holistic assessment models is an unexplored area that has the potential to provide a comprehensive understanding of an individual's mental state. This research gap can be addressed by focusing on mental state parameters such as cognitive load, impairment, and emotional state simultaneously. While other models in the field often concentrate on one parameter, this thesis takes a more comprehensive approach by targeting all three. This comprehensive approach can provide a complete understanding of an individual's mental state, thus addressing

Table 7.13: Comparison with existing models

Model		**CL**	**CI**	**ES**
[40–43, 99, 168]		Y		
[35, 106, 115, 119, 120, 220]			Y	
[47, 48, 59, 61, 62, 143]				Y
Proposed Thesis Work	ECL-1 [49] and ECL-2	Y		
	ETMT [14, 15]		Y	
	PredictEYE [36, 38]			Y
	Emotional state detection model			Y

the research gap identified. Table 7.13 shows a comparison with existing models, focusing on one particular assessment, whereas the thesis aims to understand cognitive load, cognitive impairment, and emotional state.

Chapter 8

Conclusion and Scope for further Research

The thesis has significantly contributed to developing both unimodal and multimodal datasets for identifying and classifying mental state parameters. Novel features like error rate, scanpath comparison score, and inattentional blindness have been extracted to enhance the precision of mental state assessments. The ECL-2 model achieves a notable 94% accuracy in cognitive load assessment, employing TMT stimuli, meticulous feature extraction, and the Random Forest algorithm for accurate and interpretable cognitive load assessments. The Eye-Tracking Trail Making Test (ETMT), model introduces a valuable tool for evaluating cognitive abilities, specifically in visual perception and attentional processes. Comprehensive scores, such as focused attention and visual search speed, are provided by the ETMT model, which offers insightful information on cognitive impairment based on task performance. This eye tracking version of the Trail Making Test serves as an assistive tool, providing healthcare experts with valuable indicators of cognitive impairment based on task performance.

In emotional state prediction, PredictEYE outperforms ARIMA with an accuracy of 86.4%. It uses LSTM for forecasting and Random Forest for classification, focusing on eye tracking features during video viewing. The system integrates personalized univariate time series models for more precise mental health assessments. A multimodal dataset, EmoRPhyE, is developed to research emotional states with diverse stimuli of pleasant and unpleasant images. This thesis establishes a robust framework for identifying and classifying mental state parameters.

In future the ETMT model can be integrated with eye-tracking software as a plugin tool in the automated screening of cognitive impairment becomes useful for the improvement of healthcare professionals. Applying the PredictEYE model for the eye tracking by webcam guarantees constant and nonintrusive surveillance of the mental state. Applying reinforcement learning in PredictEYE can improve the accuracy and individuality of the results in the future. The EmoRPhyE dataset demonstrates the potential of combining eye tracking with physiological measures to enhance understanding of mental and emotional well-being, paving the way for advanced AI techniques.

References

[1] L. Clements, "The role of science in mental health: Insights from the wellcome global monitor.." https://wellcome.org/reports/wellcome-global-monitor-mental-health/2020, 2021 (accessed 2023).

[2] IPSOS, "Frequency of visiting or consulting with other type of medical specialists or healthcare professionals in india." https://www.statista.com/statistics/916688/india-frequency-of-visits-to-other-type-of-medical-specialists/, 2019 (accessed 2023).

[3] IPSOS, "Global health service monitor 2021." https://www.ipsos.com/en-uk/global-health-service-monitor-2021, 2021 (accessed 2023).

[4] V. Hermann, N. Durbeej, A.-C. Karlsson, and A. Sarkadi, "'feeling down one evening doesn't count as having mental health problems'—swedish adolescents' conceptual views of mental health," *Journal of Advanced Nursing*, vol. 79, no. 8, pp. 2886–2899, 2023.

[5] J. Bergamin, J. Luigjes, J. Kiverstein, C. L. Bockting, and D. Denys, "Defining autonomy in psychiatry," *Frontiers in psychiatry*, vol. 13, p. 801415, 2022.

[6] P. Fusar-Poli, G. S. de Pablo, A. De Micheli, D. H. Nieman, C. U. Correll, L. V. Kessing, A. Pfennig, A. Bechdolf, S. Borgwardt, C. Arango, *et al.*, "What is good mental health? a scoping review," *European neuropsychopharmacology*, vol. 31, pp. 33–46, 2020.

[7] C. R. Cloninger, I. M. Salloum, and J. E. Mezzich, "The dynamic origins of positive health and wellbeing," *International Journal of Person Centered Medicine*, vol. 2, no. 2, p. 179, 2012.

[8] R. Z. Zheng, *Cognitive load measurement and application: a theoretical framework for meaningful research and practice*. Routledge, 2017.

[9] A. Grisenti, G. Nollo, M. Dalprá, F. D. Natale, M. D. Cecco, A. Francesconi, A. Fornaser, P. Tomasin, N. Garau, L. Guandalini, *et al.*, “Technological infrastructure supports new paradigm of care for healthy aging: The living lab ausilia,” in *Ambient Assisted Living: Italian Forum 2019 10*, pp. 85–99, Springer International Publishing, 2021.

[10] R. S. Murthy, “National mental health survey of india 2015–2016,” *Indian journal of psychiatry*, vol. 59, no. 1, p. 21, 2017.

[11] M. Kivimäki, G. D. Batty, J. Pentti, M. J. Shipley, P. N. Sipilä, S. T. Nyberg, S. B. Suominen, T. Oksanen, S. Stenholm, M. Virtanen, *et al.*, “Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study,” *The Lancet Public Health*, vol. 5, no. 3, pp. e140–e149, 2020.

[12] J. Gruber, M. J. Prinstein, L. A. Clark, J. Rottenberg, J. S. Abramowitz, A. M. Albano, A. Aldao, J. L. Borelli, T. Chung, J. Davila, *et al.*, “Mental health and clinical psychological science in the time of covid-19: Challenges, opportunities, and a call to action.,” *American Psychologist*, vol. 76, no. 3, p. 409, 2021.

[13] N. Okkels, C. B. Kristiansen, and P. Munk-Jørgensen, *Mental health and illness in the city.* Springer, 2017.

[14] C. Jyotsna, J. Amudha, R. Rao, and R. Nayar, “Intelligent gaze tracking approach for trail making test,” *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 6299–6310, 2020.

[15] J. Chandrasekharan, A. Joseph, A. Ram, and G. Nollo, “Etmt: A tool for eye-tracking-based trail-making test to detect cognitive impairment,” *Sensors*, vol. 23, no. 15, p. 6848, 2023.

[16] F. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, “Towards multilevel mental stress assessment using svm with ecoc: an eeg approach,” *Medical & biological engineering & computing*, vol. 56, no. 1, pp. 125–136, 2018.

[17] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke, “The trier social stress test: principles and practice,” *Neurobiology of stress*, vol. 6, pp. 113–126, 2017.

[18] G. Acerbi, E. Rovini, S. Betti, A. Tirri, J. F. Rónai, A. Sirianni, J. Agrimi, L. Eusebi, and F. Cavallo, “A wearable system for stress detection through

physiological data analysis," in *Italian forum of ambient assisted living*, pp. 31–50, Springer, 2016.

[19] S. B. Shafiei, Z. Lone, A. S. Elsayed, A. A. Hussein, and K. A. Guru, "Identifying mental health status using deep neural network trained by visual metrics," *Translational psychiatry*, vol. 10, no. 1, pp. 1–8, 2020.

[20] S. Krishnan, J. Amudha, and S. Tejwani, "Gaze exploration index (ge i)-explainable detection model for glaucoma," *IEEE Access*, vol. 10, pp. 74334–74350, 2022.

[21] R. Martinez-Maldonado, "" i spent more time with that team" making spatial pedagogy visible using positioning sensors," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pp. 21–25, 2019.

[22] C. Tsiourti, A. Weiss, K. Wac, and M. Vincze, "Multimodal integration of emotional signals from voice, body, and context: Effects of (in) congruence on emotion recognition and attitudes towards robots," *International Journal of Social Robotics*, vol. 11, no. 4, pp. 555–573, 2019.

[23] M. J. Junokas, R. Lindgren, J. Kang, and J. W. Morphew, "Enhancing multimodal learning through personalized gesture recognition," *Journal of Computer Assisted Learning*, vol. 34, no. 4, pp. 350–357, 2018.

[24] C. R. Henrie, R. Bodily, R. Larsen, and C. R. Graham, "Exploring the potential of lms log data as a proxy measure of student engagement," *Journal of Computing in Higher Education*, vol. 30, no. 2, pp. 344–362, 2018.

[25] D. Di Mitri, M. Scheffel, H. Drachsler, D. Börner, S. Ternier, and M. Specht, "Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data," in *Proceedings of the seventh international learning analytics & knowledge conference*, pp. 188–197, 2017.

[26] M. N. Giannakos, K. Sharma, I. O. Pappas, V. Kostakos, and E. Velloso, "Multimodal data as a means to understand the learning experience," *International Journal of Information Management*, vol. 48, pp. 108–119, 2019.

[27] R. Pernice, Y. Antonacci, M. Zanetti, A. Busacca, D. Marinazzo, L. Faes, and G. Nollo, "Multivariate correlation measures reveal structure and strength of brain-body physiological networks at rest and during mental stress," *Frontiers in neuroscience*, vol. 14, p. 1427, 2020.

[28] R. Markiewicz, A. Markiewicz-Gospodarek, and B. Dobrowolska, "Galvanic skin response features in psychiatry and mental disorders: A narrative review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 20, p. 13428, 2022.

[29] D. Carneiro, J. C. Castillo, P. Novais, A. Fernández-Caballero, and J. Neves, "Multimodal behavioral analysis for non-invasive stress detection," *Expert Systems with Applications*, vol. 39, no. 18, pp. 13376–13389, 2012.

[30] S. Klinghammer, T. Voitsekhivska, N. Licciardello, K. Kim, C.-K. Baek, H. Cho, K.-J. Wolter, C. Kirschbaum, L. Baraban, and G. Cuniberti, "Nanosensor-based real-time monitoring of stress biomarkers in human saliva using a portable measurement system," *ACS sensors*, vol. 5, no. 12, pp. 4081–4091, 2020.

[31] K. Rayner, "Eye movements in reading and information processing: 20 years of research.," *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.

[32] P. A. Punde, M. E. Jadhav, and R. R. Manza, "A study of eye tracking technology and its applications," in *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, pp. 86–90, IEEE, 2017.

[33] Y. Wang, H.-L. Lyu, X.-H. Tian, B. Lang, X.-Y. Wang, D. St Clair, R. Wu, and J. Zhao, "The similar eye movement dysfunction between major depressive disorder, bipolar depression and bipolar mania," *The World Journal of Biological Psychiatry*, vol. 23, no. 9, pp. 689–702, 2022.

[34] M. Rantanen, J. Hautala, O. Loberg, J. Nuorva, J. K. Hietanen, L. Nummenmaa, and P. Astikainen, "Attentional bias towards interpersonal aggression in depression–an eye movement study," *Scandinavian journal of psychology*, vol. 62, no. 5, pp. 639–647, 2021.

[35] K. Tadokoro, T. Yamashita, Y. Fukui, E. Nomura, Y. Ohta, S. Ueno, S. Nishina, K. Tsunoda, Y. Wakutani, Y. Takao, *et al.*, "Early detection of cognitive decline in mild cognitive impairment and alzheimer's disease with a novel eye tracking test," *Journal of the neurological sciences*, vol. 427, p. 117529, 2021.

[36] C. Jyotsna, J. Amudha, A. Ram, and G. Nollo, "Inteleye: An intelligent tool for the detection of stressful state based on eye gaze data while watching video," *Procedia Computer Science*, vol. 218, pp. 1270–1279, 2023.

[37] J. L. Kröger, O. H.-M. Lutz, and F. Müller, "What does your gaze reveal about you? on the privacy implications of eye tracking," *Privacy and Identity Management. Data for Better Living: AI and Privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2. 2 International Summer School, Windisch, Switzerland, August 19–23, 2019, Revised Selected Papers 14*, pp. 226–241, 2020.

[38] C. Jyotsna, J. Amudha, A. Ram, D. Fruet, and G. Nollo, "Predicteye: Personalized time series model for mental state prediction using eye tracking," *IEEE Access*, vol. 11, pp. 128383–128409, 2023.

[39] M. Burch, R. Haymoz, and S. Lindau, "The benefits and drawbacks of eye tracking for improving educational systems," in *2022 Symposium on Eye Tracking Research and Applications*, pp. 1–5, 2022.

[40] M. D. Babu, D. JeevithaShree, G. Prabhakar, K. P. S. Saluja, A. Pashilkar, and P. Biswas, "Estimating pilots' cognitive load from ocular parameters through simulation and in-flight studies," *Journal of Eye Movement Research*, vol. 12, no. 3, 2019.

[41] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, "Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze," *PloS one*, vol. 13, no. 9, p. e0203629, 2018.

[42] M. Behroozi and C. Parnin, "Can we predict stressful technical interview settings through eye-tracking?," in *Proceedings of the Workshop on Eye Movements in Programming*, pp. 1–5, 2018.

[43] J. Zagermann, U. Pfeil, and H. Reiterer, "Measuring cognitive load using eye tracking technology in visual computing," in *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization*, pp. 78–85, 2016.

[44] N. Nourbakhsh, F. Chen, Y. Wang, and R. A. Calvo, "Detecting users' cognitive load by galvanic skin response with affective interference," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 3, pp. 1–20, 2017.

[45] L. Pu, D. Pan, H. Wang, X. He, X. Zhang, Z. Yu, N. Hu, Y. Du, S. He, X. Liu, *et al.*, "A predictive model for the risk of cognitive impairment in community middle-aged and older adults," *Asian Journal of Psychiatry*, vol. 79, p. 103380, 2023.

[46] T. D. Wilcockson, D. Mardanbegi, B. Xia, S. Taylor, P. Sawyer, H. W. Gellersen, I. Leroi, R. Killick, and T. J. Crawford, "Abnormalities of saccadic eye movements in dementia due to alzheimer's disease and mild cognitive impairment," *Aging (Albany NY)*, vol. 11, no. 15, p. 5389, 2019.

[47] W. Tang, S. Wu, T. Vigier, and M. P. Da Silva, "Influence of emotions on eye behavior in omnidirectional content," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, 2020.

[48] N. Baharom, N. Jayabalan, M. Amin, and S. Wibirama, "Positive emotion recognition through eye tracking technology," *Journal of Advanced Manufacturing Technology (JAMT)*, vol. 13, no. 2 (1), 2019.

[49] C. Jyotsna and J. Amudha, "Eye gaze as an indicator for stress level analysis in students," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1588–1593, IEEE, 2018.

[50] IPSOS, "Ipsos global health service monitor 2023." `https://www.ipsos.com/sites/default/files/ct/news/documents/2023-09/Ipsos-Global-Health-Service-Monitor-2023-WEB.pdf`, 2023 (accessed 2023).

[51] M. S. Gautham, G. Gururaj, M. Varghese, V. Benegal, G. N. Rao, A. Kokane, B. S. Chavan, P. K. Dalal, D. Ram, K. Pathak, *et al.*, "The national mental health survey of india (2016): Prevalence, socio-demographic correlates and treatment gap of mental morbidity," *International Journal of Social Psychiatry*, vol. 66, no. 4, pp. 361–372, 2020.

[52] "Mental health care analysis." https://timesofindia.indiatimes.com/blogs/voices/mental-health-care-analysis/.

[53] R. Z. Marandi and P. Gazerani, "Aging and eye tracking: in the quest for objective biomarkers," *Future Neurology*, vol. 14, no. 4, p. FNL33, 2019.

[54] M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Wearable eye tracking for mental health monitoring," *Computer Communications*, vol. 35, no. 11, pp. 1306–1311, 2012.

[55] K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Computerized Medical Imaging and Graphics*, vol. 65, pp. 176–190, 2018.

[56] S. Akshay and K. Vincent, "Identification of parkinson disease patients classification using feed forward technique based on speech signals," *Int. J. Eng. Adv. Technol*, vol. 8, no. 5, pp. 1769–1778, 2019.

[57] L. Tao, Q. Wang, D. Liu, J. Wang, Z. Zhu, and L. Feng, "Eye tracking metrics to screen and assess cognitive impairment in patients with neurological disorders," *Neurological Sciences*, vol. 41, no. 7, pp. 1697–1704, 2020.

[58] M. Polden, T. D. Wilcockson, and T. J. Crawford, "The disengagement of visual attention: An eye-tracking study of cognitive impairment, ethnicity and age," *Brain Sciences*, vol. 10, no. 7, p. 461, 2020.

[59] M. Oliva and A. Anikin, "Pupil dilation reflects the time course of emotion recognition in human vocalizations," *Scientific reports*, vol. 8, no. 1, p. 4871, 2018.

[60] U. Ernsberger and H. Rohrer, "Sympathetic tales: subdivisons of the autonomic nervous system and the impact of developmental studies," *Neural Development*, vol. 13, no. 1, pp. 1–21, 2018.

[61] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Eye-tracking analysis for emotion recognition," *Computational intelligence and neuroscience*, vol. 2020, 2020.

[62] S. Alghowinem, M. AlShehri, R. Goecke, and M. Wagner, "Exploring eye activity as an indication of emotional states using an eye-tracking sensor," in *Intelligent Systems for Science and Information: Extended and Selected Results from the Science and Information Conference 2013*, pp. 261–276, Springer, 2014.

[63] C. Mestre, J. Gautier, and J. Pujol, "Robust eye tracking based on multiple corneal reflections for clinical applications," *Journal of biomedical optics*, vol. 23, no. 3, pp. 035001–035001, 2018.

[64] Y. Ebisawa and A. Nakashima, "Increasing precision of pupil position detection using the corneal reflection," *The Institute of Image Information and Television Engineers*, vol. 62, no. 7, pp. 1122–1126, 2008.

[65] C. Zhang, J.-N. Chi, Z.-H. Zhang, and Z.-L. Wang, "A novel eye gaze tracking technique based on pupil center cornea reflection technique," *Jisuanji Xuebao(Chinese Journal of Computers)*, vol. 33, no. 7, pp. 1272–1285, 2010.

[66] Z. Li, P. Guo, and C. Song, "A review of main eye movement tracking methods," in *Journal of Physics: Conference Series*, vol. 1802, p. 042066, IOP Publishing, 2021.

[67] "How do tobii eye trackers work?." https://connect.tobii.com/s/article/How-do-Tobii-eye-trackers-work?language=en_US.

[68] M. Pomplun and S. Sunkara, "Pupil dilation as an indicator of cognitive workload in human-computer interaction," in *Human-Centered Computing*, pp. 542–546, CRC Press, 2019.

[69] J. Wang, P. Antonenko, M. Celepkolu, Y. Jimenez, E. Fieldman, and A. Fieldman, "Exploring relationships between eye tracking and traditional usability testing data," *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 483–494, 2019.

[70] S. Białowas and A. Szyszka, "Eye-tracking in marketing research," *Managing Economic Innovations–Methods and Instruments*, vol. 1, no. 69, pp. 91–104, 2019.

[71] V. Clay, P. König, and S. Koenig, "Eye tracking in virtual reality," *Journal of eye movement research*, vol. 12, no. 1, 2019.

[72] V. Sundstedt, *Gazing at games: An introduction to eye tracking control*. Springer Nature, 2022.

[73] J. Mucke, M. Schwarzkopf, and J. Siegmund, "Reyeker: Remote eye tracker," in *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–5, 2021.

[74] C.-h. Chen, C. Monroy, D. M. Houston, and C. Yu, "Using head-mounted eye-trackers to study sensory-motor dynamics of coordinated attention," *Progress in Brain Research*, vol. 254, pp. 71–88, 2020.

[75] I. T. Hooge, D. C. Niehorster, R. S. Hessels, J. S. Benjamins, and M. Nyström, "How robust are wearable eye trackers to slow and fast head and body movements?," *Behavior Research Methods*, pp. 1–15, 2022.

[76] S. Dowiasch, P. Wolf, and F. Bremmer, "Quantitative comparison of a mobile and a stationary video-based eye-tracker," *Behavior research methods*, vol. 52, pp. 667–680, 2020.

[77] H.-Y. Chang, H.-F. Wu, Y.-C. Chang, Y.-S. Tseng, and Y.-C. Wang, "The effects of a virtual simulation-based, mobile technology application on nursing students' learning achievement and cognitive load: Randomized controlled trial," *International Journal of Nursing Studies*, vol. 120, p. 103948, 2021.

[78] A. P. Field and R. R. Wilcox, "Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers," *Behaviour research and therapy*, vol. 98, pp. 19–38, 2017.

[79] P. Cohen, S. G. West, and L. S. Aiken, *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology press, 2014.

[80] S. Tuarob, C. S. Tucker, S. Kumara, C. L. Giles, A. L. Pincus, D. E. Conroy, and N. Ram, "How are you feeling?: A personalized methodology for predicting mental states from temporally observable physical and behavioral information," *Journal of biomedical informatics*, vol. 68, pp. 1–19, 2017.

[81] Y. Liu, J. M. Ritchie, T. Lim, Z. Kosmadoudi, A. Sivanathan, and R. C. Sung, "A fuzzy psycho-physiological approach to enable the understanding of an engineer's affect status during cad activities," *Computer-Aided Design*, vol. 54, pp. 19–38, 2014.

[82] T. Azim, M. A. Jaffar, and A. M. Mirza, "Fully automated real time fatigue detection of drivers through fuzzy expert systems," *Applied Soft Computing*, vol. 18, pp. 25–38, 2014.

[83] A. A. Obinikpo and B. Kantarci, "Big sensed data meets deep learning for smarter health care in smart cities," *Journal of Sensor and Actuator Networks*, vol. 6, no. 4, p. 26, 2017.

[84] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard, and J. Tørresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," *Pervasive and Mobile Computing*, vol. 51, pp. 1–26, 2018.

[85] E. Serrano and J. Bajo, "Discovering hidden mental states in open multi-agent systems by leveraging multi-protocol regularities with machine learning," *Sensors*, vol. 20, no. 18, p. 5198, 2020.

[86] F. Li, G. Xu, and S. Feng, "Eye tracking analytics for mental states assessment–a review," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2266–2271, IEEE, 2021.

[87] Q. Wei, H. Cao, Y. Shi, X. Xu, and T. Li, "Machine learning based on eye-tracking data to identify autism spectrum disorder: A systematic review and meta-analysis," *Journal of Biomedical Informatics*, p. 104254, 2022.

[88] G. Wan, X. Kong, B. Sun, S. Yu, Y. Tu, J. Park, C. Lang, M. Koh, Z. Wei, Z. Feng, *et al.*, "Applying eye tracking to identify autism spectrum disorder in children," *Journal of autism and developmental disorders*, vol. 49, pp. 209–215, 2019.

[89] S. Adhikari and D. E. Stark, "Video-based eye tracking for neuropsychiatric assessment," *Annals of the New York Academy of Sciences*, vol. 1387, no. 1, pp. 145–152, 2017.

[90] Z. Xiang, W. Gao, T. Tao, L. Wang, C. Fan, and L. Xu, "Psychological test of patients with mental disorders based on eye movement data fusion algorithm," in *2021 4th International Conference on Information Systems and Computer Aided Education*, pp. 2440–2445, 2021.

[91] Q. Wu, N. Dey, F. Shi, R. G. Crespo, and R. S. Sherratt, "Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks," *Applied Soft Computing*, vol. 110, p. 107752, 2021.

[92] A. Yadav, C. Jha, and A. Sharan, "Optimizing lstm for time series prediction in indian stock market," *Procedia Computer Science*, vol. 167, pp. 2091–2100, 2020.

[93] P. T. Yamak, L. Yujian, and P. K. Gadosey, "A comparison between arima, lstm, and gru for time series forecasting," in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 49–55, 2019.

[94] T. Hulsen, S. S. Jamuar, A. R. Moody, J. H. Karnes, O. Varga, S. Hedensted, R. Spreafico, D. A. Hafler, and E. F. McKinney, "From big data to precision medicine," *Frontiers in medicine*, p. 34, 2019.

[95] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of covid-19 using deep learning models: India-usa comparative case study," *Chaos, Solitons & Fractals*, vol. 140, p. 110227, 2020.

[96] M. Bartels and S. P. Marshall, "Measuring cognitive workload across different eye tracking hardware platforms," in *Proceedings of the symposium on eye tracking research and applications*, pp. 161–164, 2012.

[97] A. Febiyani, A. Febriani, J. Ma'Sum, *et al.*, "Calculation of mental load from e-learning student with nasa tlx and sofi method," *Jurnal Sistem Dan Manajemen Industri*, vol. 5, no. 1, pp. 35–42, 2021.

[98] P. L. A. Pratiwi, A. Pawennari, and I. N. Afiah, "Analysis of the influence of mental workload on employee performance with the subjective workload assessment technique (swat) method at pt. bumi jasa utama (kalla group)," *Journal of Sustainability Industrial Engineering and Management System*, vol. 1, no. 1, pp. 23–31, 2022.

[99] Y. Yamada and M. Kobayashi, "Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults," *Artificial intelligence in medicine*, vol. 91, pp. 39–48, 2018.

[100] S. Hutt, A. E. Stewart, J. Gregg, S. Mattingly, and S. K. D'Mello, "Feasibility of longitudinal eye-gaze tracking in the workplace," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. ETRA, pp. 1–21, 2022.

[101] H. Rahman, M. U. Ahmed, S. Barua, P. Funk, and S. Begum, "Vision-based driver's cognitive load classification considering eye movement using machine learning and deep learning," *Sensors*, vol. 21, no. 23, p. 8019, 2021.

[102] A.-L. Le Cunff, E. Dommett, and V. Giampietro, "Neurophysiological measures and correlates of cognitive load in attention-deficit/hyperactivity disorder (adhd), autism spectrum disorder (asd) and dyslexia: A scoping review and research recommendations," *European Journal of Neuroscience*, vol. 59, no. 2, pp. 256–282, 2024.

[103] H. C. Pearson, D. E. MacKenzie, D. T. Oystreck, and D. A. Westwood, "Using eye-tracking technology to measure cognitive function in mild traumatic brain injury: A scoping review," *Journal of concussion*, vol. 7, p. 20597002231194151, 2023.

[104] R. G. Ross, A. Olincy, S. K. Mikulich, A. D. Radant, J. G. Harris, M. Waldo, N. Compagnon, S. Heinlein, S. Leonard, G. O. Zerbe, *et al.*, "Admixture analysis of smooth pursuit eye movements in probands with schizophrenia and their relatives suggests gain and leading saccades are potential endophenotypes," *Psychophysiology*, vol. 39, no. 6, pp. 809–819, 2002.

[105] K. Conner, C. Y. Sweeney, T. Brown, L. Childs, S. Rogers, and T. Gregory, "Practical applications of physical activity for successful cognitive aging," *JAAPA*, vol. 30, no. 8, pp. 30–35, 2017.

[106] A. Oyama, S. Takeda, Y. Ito, T. Nakajima, Y. Takami, Y. Takeya, K. Yamamoto, K. Sugimoto, H. Shimizu, M. Shimamura, *et al.*, "Novel method for rapid assessment of cognitive impairment using high-performance eye-tracking technology," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[107] R. Morimoto, H. Kawanaka, Y. Hicks, and R. Setchi, "Development of recreation game for measurement of eye movement using tangram," *Procedia Computer Science*, vol. 192, pp. 4924–4932, 2021.

[108] H. P. Surendran, M. Narmadha, S. Kalavagunta, A. Sasidharan, and D. Dutta, "Preservation of cognitive function after brain irradiation," *Journal of Oncology Pharmacy Practice*, p. 10781552221077037, 2022.

[109] Y. Guo, "A selective review of the ability for variants of the trail making test to assess cognitive impairment," *Applied Neuropsychology: Adult*, pp. 1–12, 2021.

[110] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""mini-mental state": a practical method for grading the cognitive state of patients for the clinician," *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–198, 1975.

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

[112] C. d. A. Faria, H. V. D. Alves, and H. Charchat-Fichman, "The most frequently used tests for assessing executive functions in aging," *Dementia & Neuropsychologia*, vol. 9, pp. 149–155, 2015.

[113] C. Zucchella, A. Federico, A. Martini, M. Tinazzi, M. Bartolo, and S. Tamburin, "Neuropsychological testing," 2018.

[114] R. T. Lange, G. L. Iverson, M. J. Zakrzewski, P. E. Ethel-King, and M. D. Franzen, "Interpreting the trail making test following traumatic brain injury: comparison of traditional time scores and derived indices," *Journal of clinical and experimental neuropsychology*, vol. 27, no. 7, pp. 897–906, 2005.

[115] R. P. Fellows, J. Dahmen, D. Cook, and M. Schmitter-Edgecombe, "Multicomponent analysis of a digital trail making test," *The Clinical Neuropsychologist*, vol. 31, no. 1, pp. 154–167, 2017.

[116] S. L. Hicks, R. Sharma, A. N. Khan, C. M. Berna, A. Waldecker, K. Talbot, C. Kennard, and M. R. Turner, "An eye-tracking version of the trail-making test," *PLoS One*, vol. 8, no. 12, p. e84061, 2013.

[117] D. Martinez-Marquez, S. Pingali, K. Panuwatwanich, R. A. Stewart, and S. Mohamed, "Application of eye tracking technology in aviation, maritime, and construction industries: a systematic review," *Sensors*, vol. 21, no. 13, p. 4289, 2021.

[118] C. K. Ramachandra and A. Joseph, "Ieyegase: An intelligent eye gaze-based assessment system for deeper insights into learner performance," *Sensors*, vol. 21, no. 20, p. 6783, 2021.

[119] E. Tomer, T. Lupu, L. Golan, M. Wagner, and Y. Braw, "Eye tracking as a mean to detect feigned cognitive impairment in the word memory test," *Applied Neuropsychology: Adult*, 2018.

[120] K. C. Fraser, K. L. Fors, D. Kokkinakis, and A. Nordlund, "An analysis of eye-movements during reading for the detection of mild cognitive impairment," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1016–1026, 2017.

[121] J. A. Saxon, J. C. Thompson, J. M. Harris, J. Ealing, H. Hamdalla, A. Chaouch, C. Young, D. Blackburn, T. Majeed, C. Gall, *et al.*, "The edinburgh cognitive and behavioral als screen (ecas) in frontotemporal dementia," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 21, no. 7-8, pp. 606–613, 2020.

[122] T. Taule, A. S. Morland, M. A. Renså, J. Aßmus, O.-B. Tysnes, and T. Rekand, "Edinburgh cognitive and behavioral amyotrophic lateral sclerosis screen (ecas) in norway: protocol for validation and a prospective cohort study," *Contemporary Clinical Trials Communications*, vol. 14, p. 100347, 2019.

[123] J. Keller, A. Krimly, L. Bauer, S. Schulenburg, S. Böhm, H. E. Aho-Özhan, I. Uttner, M. Gorges, J. Kassubek, E. H. Pinkhardt, *et al.*, "A first approach to a neuropsychological screening tool using eye-tracking for bedside cognitive testing based on the edinburgh cognitive and behavioural als screen," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 18, no. 5-6, pp. 443–450, 2017.

[124] M. Proudfoot, R. A. Menke, R. Sharma, C. M. Berna, S. L. Hicks, C. Kennard, K. Talbot, and M. R. Turner, "Eye-tracking in amyotrophic lateral

sclerosis: a longitudinal study of saccadic and cognitive tasks," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 17, no. 1-2, pp. 101–111, 2016.

[125] X. Li, S. Jia, Z. Zhou, Y. Jin, X. Zhang, C. Hou, W. Zheng, P. Rong, and J. Jiao, "The role of the montreal cognitive assessment (moca) and its memory tasks for detecting mild cognitive impairment," *Neurological Sciences*, vol. 39, no. 6, pp. 1029–1034, 2018.

[126] F. J. Fraga, G. Q. Mamani, E. Johns, G. Tavares, T. H. Falk, and N. A. Phillips, "Early diagnosis of mild cognitive impairment and alzheimer's with event-related potentials and event-related desynchronization in n-back working memory tasks," *Computer methods and programs in biomedicine*, vol. 164, pp. 1–13, 2018.

[127] M. Russo, M. Amboni, P. Barone, M. T. Pellecchia, M. Romano, C. Ricciardi, and F. Amato, "Identification of a gait pattern for detecting mild cognitive impairment in parkinson's disease," *Sensors*, vol. 23, no. 4, p. 1985, 2023.

[128] J. Kulisevsky and J. Pagonabarraga, "Cognitive impairment in parkinson's disease: tools for diagnosis and assessment," *Movement disorders*, vol. 24, no. 8, pp. 1103–1110, 2009.

[129] D. Aarsland, L. Batzu, G. M. Halliday, G. J. Geurtsen, C. Ballard, K. Ray Chaudhuri, and D. Weintraub, "Parkinson disease-associated cognitive impairment," *Nature Reviews Disease Primers*, vol. 7, no. 1, pp. 1–21, 2021.

[130] D. Lagun, C. Manzanares, S. M. Zola, E. A. Buffalo, and E. Agichtein, "Detecting cognitive impairment by eye movement analysis using automatic classification algorithms," *Journal of neuroscience methods*, vol. 201, no. 1, pp. 196–203, 2011.

[131] M. D. Crutcher, R. Calhoun-Haney, C. M. Manzanares, J. J. Lah, A. I. Levey, and S. M. Zola, "Eye tracking during a visual paired comparison task as a predictor of early dementia," *American Journal of Alzheimer's Disease & Other Dementias®*, vol. 24, no. 3, pp. 258–266, 2009.

[132] Z. Liu, Z. Yang, Y. Gu, H. Liu, and P. Wang, "The effectiveness of eye tracking in the diagnosis of cognitive disorders: A systematic review and meta-analysis," *PloS one*, vol. 16, no. 7, p. e0254059, 2021.

[133] D. Venugopal, J. Amudha, and C. Jyotsna, "Developing an application using eye tracker," in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTE-ICT)*, pp. 1518–1522, IEEE, 2016.

[134] S.-J. Moon, C.-H. Park, S. I. Jung, J.-W. Yu, E.-C. Son, H. N. Lee, H. Jeong, S. Jang, E. Park, and T.-D. Jung, "Effects of an eye-tracking linkage attention training system on cognitive function compared to conventional computerized cognitive training system in patients with stroke," in *Healthcare*, vol. 10, p. 456, MDPI, 2022.

[135] F. Júlio, G. Caetano, C. Januário, and M. Castelo-Branco, "The effect of impulsivity and inhibitory control deficits in the saccadic behavior of premanifest huntington's disease individuals," *Orphanet Journal of Rare Diseases*, vol. 14, pp. 1–13, 2019.

[136] U. Leon-Dominguez, J. F. Martín-Rodríguez, and J. León-Carrión, "Executive n-back tasks for the neuropsychological assessment of working memory," *Behavioural brain research*, vol. 292, pp. 167–173, 2015.

[137] I. Vigo, L. Coelho, and S. Reis, "Speech-and language-based classification of alzheimer's disease: A systematic review," *Bioengineering*, vol. 9, no. 1, p. 27, 2022.

[138] J. Nie, Q. Qiu, M. Phillips, L. Sun, F. Yan, X. Lin, S. Xiao, and X. Li, "Early diagnosis of mild cognitive impairment based on eye movement parameters in an aging chinese population," *Frontiers in aging neuroscience*, vol. 12, p. 221, 2020.

[139] R. J. Larsen and Z. Prizmic-Larsen, "Measuring emotions: Implications of a multimethod perspective.," 2006.

[140] L. Costantini, C. Pasquarella, A. Odone, M. E. Colucci, A. Costanza, G. Serafini, A. Aguglia, M. B. Murri, V. Brakoulias, M. Amore, *et al.*, "Screening for depression in primary care with patient health questionnaire-9 (phq-9): A systematic review," *Journal of affective disorders*, vol. 279, pp. 473–483, 2021.

[141] A. Behera, P. Matthew, A. Keidel, P. Vangorp, H. Fang, and S. Canning, "Associating facial expressions and upper-body gestures with learning tasks for enhancing intelligent tutoring systems," *International Journal of Artificial Intelligence in Education*, vol. 30, pp. 236–270, 2020.

[142] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, "Respiration-based emotion recognition with deep learning," *Computers in Industry*, vol. 92, pp. 84–90, 2017.

[143] U. Ballenghein, O. Megalakaki, and T. Baccino, "Cognitive engagement in emotional text reading: concurrent recordings of eye movements and head motion," *Cognition and Emotion*, 2019.

[144] M. E. Król and M. Król, "A novel machine learning analysis of eye-tracking data reveals suboptimal visual information extraction from facial stimuli in individuals with autism," *Neuropsychologia*, vol. 129, pp. 397–406, 2019.

[145] R. E. Nielsen, J. Banner, and S. E. Jensen, "Cardiovascular disease in patients with severe mental illness," *Nature Reviews Cardiology*, vol. 18, no. 2, pp. 136–145, 2021.

[146] M. Lohani, B. R. Payne, and D. L. Strayer, "A review of psychophysiological measures to assess cognitive states in real-world driving," *Frontiers in human neuroscience*, vol. 13, p. 57, 2019.

[147] J. Xu, Y. Wang, F. Chen, H. Choi, G. Li, S. Chen, and S. Hussain, "Pupillary response based cognitive workload index under luminance and emotional changes," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 1627–1632, 2011.

[148] M. Stikic, C. Berka, D. J. Levendowski, R. F. Rubio, V. Tan, S. Korszen, D. Barba, and D. Wurzer, "Modeling temporal sequences of cognitive state changes based on a combination of eeg-engagement, eeg-workload, and heart rate metrics," *Frontiers in neuroscience*, vol. 8, p. 342, 2014.

[149] Z. Zhang, Y. Gao, and J. Wang, "Effects of vision and cognitive load on anticipatory and compensatory postural control," *Human Movement Science*, vol. 64, pp. 398–408, 2019.

[150] J. Zhou, K. Yu, F. Chen, Y. Wang, and S. Z. Arshad, "Multimodal behavioral and physiological signals as indicators of cognitive load," in *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2*, pp. 287–329, 2018.

[151] D. Rosen, A. Palatnik, and D. Abrahamson, "A better story: An embodied-design argument for generic manipulatives," in *Using mobile technologies in the teaching and learning of mathematics*, pp. 189–211, Springer, 2018.

[152] K. Sharma, P. Dillenbourg, and M. Giannakos, "Stimuli-based gaze analytics to enhance motivation and learning in moocs," in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, vol. 2161, pp. 199–203, IEEE, 2019.

[153] B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. Pea, "Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups," *International Journal of Computer-Supported Collaborative Learning*, vol. 13, no. 3, pp. 241–261, 2018.

[154] S. Mu, M. Cui, and X. Huang, "Multimodal data fusion in learning analytics: A systematic review," *Sensors*, vol. 20, no. 23, p. 6856, 2020.

[155] M. C. Sáiz-Manzanares, I. R. Pérez, A. A. Rodríguez, S. R. Arribas, L. Almeida, and C. F. Martin, "Analysis of the learning process through eye tracking technology and feature selection techniques," *Applied Sciences*, vol. 11, no. 13, p. 6157, 2021.

[156] D. Kraft, K. van Laerhofen, and G. Bieber, "Carecam: Concept of a new tool for corporate health management," in *The 14th PErvasive Technologies Related to Assistive Environments Conference*, pp. 585–593, 2021.

[157] K. H. Abdulkareem, M. A. Mohammed, A. Salim, M. Arif, O. Geman, D. Gupta, and A. Khanna, "Realizing an effective covid-19 diagnosis system based on machine learning and iot in smart hospital environment," *IEEE Internet of Things Journal*, 2021.

[158] D. JOSEPH, "The world's most relaxing film." `https://www.youtube.com/watch?v=dkFNdABPhC0`, 2015 (accessed 2023).

[159] "The tiger movie." `https://https://www.youtube.com/watch?v=W0SUBdaMjKs`, 2015 (accessed 2023).

[160] S. Uday, C. Jyotsna, and J. Amudha, "Detection of stress using wearable sensors in iot platform," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 492–498, IEEE, 2018.

[161] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 71–78, 2000.

[162] P. J. Lang, M. M. Bradley, B. N. Cuthbert, *et al.*, *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* NIMH, Center for the Study of Emotion & Attention Gainesville, FL, 2005.

[163] N. Du, F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang, "Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving," *Transportation research part C: emerging technologies*, vol. 112, pp. 78–87, 2020.

[164] V. Skaramagkas, G. Giannakakis, E. Ktistakis, D. Manousos, I. Karatzanis, N. S. Tachos, E. Tripoliti, K. Marias, D. I. Fotiadis, and M. Tsiknakis, "Review of eye tracking metrics involved in emotional and cognitive processes," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 260–277, 2021.

[165] K. Mundnich, B. M. Booth, M. l'Hommedieu, T. Feng, B. Girault, J. L'hommedieu, M. Wildman, S. Skaaden, A. Nadarajan, J. L. Villatte, *et al.*, "Tiles-2018, a longitudinal physiologic and behavioral data set of hospital workers," *Scientific Data*, vol. 7, no. 1, p. 354, 2020.

[166] R. Martins and J. Carvalho, "Eye blinking as an indicator of fatigue and mental load—a systematic review," *Occupational safety and hygiene III*, vol. 10, pp. 231–235, 2015.

[167] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley, and S. A. Bunge, "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?," *Developmental cognitive neuroscience*, vol. 25, pp. 69–91, 2017.

[168] T. Čegovnik, K. Stojmenova, G. Jakus, and J. Sodnik, "An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers," *Applied ergonomics*, vol. 68, pp. 1–11, 2018.

[169] S. Wasef, I. Laksono, P. Kapoor, D. Tang-Wei, D. Gold, A. Saripella, S. Riazi, S. Islam, M. Englesakis, J. Wong, *et al.*, "Screening for subjective cognitive decline in the elderly via subjective cognitive complaints and informant-reported questionnaires: a systematic review," *BMC anesthesiology*, vol. 21, no. 1, pp. 1–9, 2021.

[170] H. E. Whitson, R. Malhotra, A. Chan, D. B. Matchar, and T. Østbye, "Comorbid visual and cognitive impairment: relationship with disability status and self-rated health among older singaporeans," *Asia Pacific Journal of Public Health*, vol. 26, no. 3, pp. 310–319, 2014.

[171] R. Jorde, E. B. Mathiesen, S. Rogne, T. Wilsgaard, M. Kjærgaard, G. Grimnes, and H. Schirmer, "Vitamin d and cognitive function: the tromsø study," *Journal of the Neurological Sciences*, vol. 355, no. 1-2, pp. 155–161, 2015.

[172] R. C. Petersen and S. Negash, "Mild cognitive impairment: an overview," *CNS spectrums*, vol. 13, no. 1, pp. 45–53, 2008.

[173] C. M. Carvalho, F. L. Seixas, A. Conci, D. C. Muchaluat-Saade, J. Laks, and Y. Boechat, "A dynamic decision model for diagnosis of dementia, alzheimer's disease and mild cognitive impairment," *Computers in Biology and Medicine*, vol. 126, p. 104010, 2020.

[174] R. U. Haque, A. L. Pongos, C. M. Manzanares, J. J. Lah, A. I. Levey, and G. D. Clifford, "Deep convolutional neural networks and transfer learning for measuring cognitive impairment using eye-tracking in a distributed tablet-based environment," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 11–18, 2020.

[175] C. Belkhiria and V. Peysakhovich, "Eog metrics for cognitive workload detection," *Procedia Computer Science*, vol. 192, pp. 1875–1884, 2021.

[176] W. Albert and T. Tullis, *Measuring the user experience*. Elsevier, 2010.

[177] S. Popelka and M. Beitlova, "Scanpath comparison using scangraph for education and learning purposes: Summary of previous educational studies performed with the use of scangraph," in *2022 Symposium on Eye Tracking Research and Applications*, pp. 1–6, 2022.

[178] M. Ekelund, H. Fernsund, S. Karlsson, and E. Mac Giolla, "Does expertise reduce rates of inattentional blindness? a meta-analysis," *Perception*, vol. 51, no. 2, pp. 131–147, 2022.

[179] C. Kreitz, P. Furley, D. Memmert, and D. J. Simons, "Inattentional blindness and individual differences in cognitive abilities," *PLOS one*, vol. 10, no. 8, p. e0134675, 2015.

[180] J. K. Seegmiller, J. M. Watson, and D. L. Strayer, "Individual differences in susceptibility to inattentional blindness.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 37, no. 3, p. 785, 2011.

[181] I. . Sánchez-Cubillo, J. A. Periáñez, D. Adrover-Roig, J. M. Rodríguez-Sánchez, M. Ríos-Lago, J. Tirapu, and F. Barceló, "Construct validity

of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities," *Journal of the International Neuropsychological Society*, vol. 15, no. 3, pp. 438–450, 2009.

[182] D. L. Mann, J. Causer, H. Nakamoto, and O. R. Runswick, "Visual search behaviors in expert perceptual judgements," *Anticipation and decision making in sport*, pp. 59–78, 2019.

[183] C. Kosel, D. Holzberger, and T. Seidel, "Identifying expert and novice visual scanpath patterns and their relationship to assessing learning-relevant student characteristics," in *Frontiers in Education*, vol. 5, p. 612175, Frontiers Media SA, 2021.

[184] Y. Zhang, T. Wilcockson, K. I. Kim, T. Crawford, H. Gellersen, and P. Sawyer, "Monitoring dementia with automatic eye movements analysis," in *Intelligent Decision Technologies 2016*, pp. 299–309, Springer, 2016.

[185] C. Palliya Guruge, S. Oviatt, P. Delir Haghighi, and E. Pritchard, "Advances in multimodal behavioral analytics for early dementia diagnosis: A review," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 328–340, 2021.

[186] K. Chandrika and J. Amudha, "A fuzzy inference system to recommend skills for source code review using eye movement data," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1743–1754, 2018.

[187] R. Fahimi and N. D. Bruce, "On metrics for measuring scanpath similarity," *Behavior Research Methods*, vol. 53, no. 2, pp. 609–628, 2021.

[188] M. Borys and M. Plechawska-Wójcik, "Eye-tracking metrics in perception and visual attention research," *EJMT*, vol. 3, pp. 11–23, 2017.

[189] K. Arbuthnott and J. Frank, "Trail making test, part b as a measure of executive control: validation using a set-switching paradigm," *Journal of clinical and experimental neuropsychology*, vol. 22, no. 4, pp. 518–528, 2000.

[190] H. Chauhan, S. Satapathy, and A. K. Sahoo, "Farmer workplace discomfort levels leading to adverse mental health," *International Journal of Social Ecology and Sustainable Development (IJSESD)*, vol. 13, no. 1, pp. 1–14, 2022.

[191] M. Seifallahi, A. H. Mehraban, F. Khamseh, M. Salehi, G. Hajati, M. Khansari, M. Ahmadzadeh, and M. Babaie, "Detection of mild cog-

nitive impairment from gait using adaptive neuro-fuzzy inference system," *Biomedical Signal Processing and Control*, vol. 71, p. 103195, 2022.

[192] S. Mahadas, C. Semkewyc, S. Suresh, and G. K. Hung, "Scan path during change-detection visual search," *Computers in Biology and Medicine*, vol. 131, p. 104233, 2021.

[193] H. Cogo-Moreira, S. H. Krance, C.-Y. Wu, K. L. Lanctôt, N. Herrmann, S. E. Black, B. J. MacIntosh, J. S. Rabin, M. Eid, W. Swardfager, *et al.*, "State, trait, and accumulated features of the alzheimer's disease assessment scale cognitive subscale (adas-cog) in mild alzheimer's disease," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 9, no. 1, p. e12376, 2023.

[194] A. N. Kurti, D. Davis, R. Redner, B. Jarvis, I. Zvorsky, D. R. Keith, H. Bolivar, T. J. White, P. Rippberger, C. Markeish, *et al.*, "A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change.," *Translational issues in psychological science*, vol. 2, no. 2, p. 128, 2016.

[195] O. Postolache, D. J. Hemanth, R. Alexandre, D. Gupta, O. Geman, and A. Khanna, "Remote monitoring of physical rehabilitation of stroke patients using iot and virtual reality," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 562–573, 2020.

[196] P. Rathnayaka, N. Mills, D. Burnett, D. De Silva, D. Alahakoon, and R. Gray, "A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring," *Sensors*, vol. 22, no. 10, p. 3653, 2022.

[197] Y. Yang, H. Wang, R. Jiang, X. Guo, J. Cheng, and Y. Chen, "A review of iot-enabled mobile healthcare: technologies, challenges, and future trends," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9478–9502, 2022.

[198] T. Sharon, "Self-tracking for health and the quantified self: Re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare," *Philosophy & Technology*, vol. 30, no. 1, pp. 93–121, 2017.

[199] L.-M. Vortmann, J. Knychalla, S. Annerer-Walcher, M. Benedek, and F. Putze, "Imaging time series of eye tracking data to classify attentional states," *Frontiers in Neuroscience*, vol. 15, p. 625, 2021.

[200] K. G. Ranjan, D. S. Tripathy, B. R. Prusty, and D. Jena, "An improved sliding window prediction-based outlier detection and correction for volatile time-series," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 34, no. 1, p. e2816, 2021.

[201] H. Yu, E. B. Klerman, R. W. Picard, and A. Sano, "Personalized wellbeing prediction using behavioral, physiological and weather data," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4, IEEE, 2019.

[202] S. Kaushik, A. Choudhury, P. K. Sheron, N. Dasgupta, S. Natarajan, L. A. Pickett, and V. Dutt, "Ai in healthcare: time-series forecasting using statistical, neural, and ensemble architectures," *Frontiers in big data*, vol. 3, p. 4, 2020.

[203] Tobii, "Tobii connect." `https://connect.tobii.com/s/article/galvanic-skin-response-gsr?language=en_US`, 2023 (accessed 2023).

[204] IMotions, "Galvanic skin response (gsr): The complete pocket guide." `https://imotions.com/blog/learning/research-fundamentals/galvanic-skin-response/`, 2020 (accessed 2023).

[205] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the arima model on the covid-2019 epidemic dataset," *Data in brief*, vol. 29, p. 105340, 2020.

[206] Y.-T. Tsan, D.-Y. Chen, P.-Y. Liu, E. Kristiani, K. L. P. Nguyen, and C.-T. Yang, "The prediction of influenza-like illness and respiratory disease using lstm and arima," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1858, 2022.

[207] L. Mou, C. Zhou, P. Zhao, B. Nakisa, M. N. Rastgoo, R. Jain, and W. Gao, "Driver stress detection via multimodal fusion using attention-based cnn-lstm," *Expert Systems with Applications*, vol. 173, p. 114693, 2021.

[208] J. Li, H. Li, W. Umer, H. Wang, X. Xing, S. Zhao, and J. Hou, "Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology," *Automation in Construction*, vol. 109, p. 103000, 2020.

[209] C. Wu, J. Cha, J. Sulek, T. Zhou, C. P. Sundaram, J. Wachs, and D. Yu, "Eye-tracking metrics predict perceived workload in robotic surgical skills training," *Human factors*, vol. 62, no. 8, pp. 1365–1386, 2020.

[210] J. A. Russell, "A circumplex model of affect.," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[211] L. Shu, Y. Yu, W. Chen, H. Hua, Q. Li, J. Jin, and X. Xu, "Wearable emotion recognition using heart rate data from a smart bracelet," *Sensors*, vol. 20, no. 3, p. 718, 2020.

[212] E. Lakomkin, M. A. Zamani, C. Weber, S. Magg, and S. Wermter, "Emorl: continuous acoustic emotion classification using deep reinforcement learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4445–4450, IEEE, 2018.

[213] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Transactions on circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 2941–2959, 2018.

[214] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[215] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, 2011.

[216] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, vol. 83, pp. 19–52, 2022.

[217] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2018.

[218] E. Ktistakis, V. Skaramagkas, D. Manousos, N. S. Tachos, E. Tripoliti, D. I. Fotiadis, and M. Tsiknakis, "Colet: A dataset for cognitive workload estimation based on eye-tracking," *Computer Methods and Programs in Biomedicine*, vol. 224, p. 106989, 2022.

[219] P. Angkan, B. Behinaein, Z. Mahmud, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "Multimodal brain–computer interface for in-vehicle driver

cognitive load measurement: Dataset and baselines," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[220] M. Polden, T. D. Wilcockson, and T. J. Crawford, "The disengagement of visual attention: An eye-tracking study of cognitive impairment, ethnicity and age," *Brain Sciences*, vol. 10, no. 7, p. 461, 2020.

List of Publications

(Based on the Research Work)

Papers Published

1. Jyotsna C. and Joseph Amudha. “Eye gaze as an indicator for stress level analysis in students.” In 2018 International conference on advances in computing, communications and informatics (ICACCI), pp. 1588-1593. IEEE, 2018.

2. Jyotsna, C., J. Amudha, Raghavendra Rao, and Ravi Nayar. “Intelligent gaze tracking approach for trail making test.” Journal of Intelligent & Fuzzy Systems 38, no. 5 (2020): 6299-6310. (Impact Factor 2)

3. Jyotsna, C., J. Amudha, and Sreedevi Uday. “A Personalized Healthcare Platform for Monitoring Mental Health of a Person During COVID.” In Proceedings of International Conference on Computing and Communication Networks: ICCCN 2021, pp. 309-317. Singapore: Springer Nature Singapore, 2022.

4. Jyotsna, C., J. Amudha, Amritanshu Ram, and Giandomenico Nollo. “IntelEye: An Intelligent Tool for the Detection of Stressful State based on Eye Gaze Data While Watching Video.” Procedia Computer Science 218 (2023): 1270-1279.

5. Chandrasekharan Jyotsna, Amudha Joseph, Amritanshu Ram, and Giandomenico Nollo. “ETMT: A Tool for Eye-Tracking-Based Trail-Making Test to Detect Cognitive Impairment.” Sensors 23, no. 15 (2023): 6848.(Impact Factor 3.9)

6. Jyotsna C., Amudha J, Amritanshu Ram, Damiano Fruet, Giandomenico Nollo, “PredictEYE: Personalised Mental State Prediction from Eye Tracking Data”, IEEE Access (Impact Factor 3.9).

Papers Communicated

1. D. Fruet, D. Ferrante, C. Jyotsna, B. Treccani, C. Mulatti, J. Amudha, G. Nollo, “EmoRPhyE: a database integrating Physiological measurements and Eye tracking data for Emotional state Recognition”, IEEE Transactions on Affective Computing.(Impact Factor 11.2)