

INVESTIGATIONS ON SLEEP DEPRIVATION INDUCED FATIGUE DETECTION USING MULTIMODAL FUSION

A thesis

Submitted for the award of the degree of

Doctor of Philosophy

Submitted by:

Jitender Singh Virk

Registration no: 951404007

Under the Supervision of:

Dr. M.D. Singh

Associate Professor, EIED

Thapar Institute of Engineering and Technology



Electrical and Instrumentation Engineering Department

Thapar Institute of Engineering and Technology

(Deemed to be University)

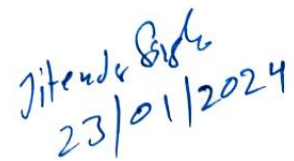
Patiala-147001, Punjab (India)

September 2023

Certificate

I hereby certify that the work presented in the thesis entitled, "Investigations on Sleep Deprivation Induced Fatigue Detection using Multimodal Fusion", for the award of the degree of Doctor of Philosophy in Electrical and Instrumentation Engineering Department (EIED). Thapar Institute of Engineering and Technology, Patiala., is an authentic record of my own work carried out under the supervision and guidance of Dr. M.D. Singh, Associate professor, EIED, Thapar Institute of Engineering and Technology, Patiala.

The matter presented in this thesis has not been submitted either in part or in full to any university or institute for the award of any degree.



(Jitender Singh Virk)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.



(Dr. M.D. Singh)

Associate Professor-EIED

Thapar Institute of Engineering and Technology

Patiala (Punjab) India

Acknowledgments

It is always challenging to acknowledge the debt of learning, which could only be paid back through gratitude. With utmost gratitude, I acknowledge the grace and blessings of the supreme Creator, Almighty for giving me the good health and strength to accomplish this research work. My heartfelt thanks and appreciation to the divine power above. Words alone cannot fully express the depth of my gratitude for all that you have done. I am thankful for every little act of kindness, every word of encouragement, and every moment of joy you have brought into my life.

I am grateful to Prof. Dr. Padmakumar Nair, Director, Thapar Institute of Engineering and Technology, Patiala (Punjab), and Dr. N. Tejo Prakash, Dean of Research and Development Cell, Thapar Institute of Engineering and Technology, Patiala (Punjab), for providing me with the facilities to complete this research work. I would like to express my heartfelt gratitude towards Prof. Dr. S.K. Singla, Head, Department of Electrical and Instrumentation Engineering, Thapar Institute of Engineering and Technology, Patiala.

I find myself filled with an overwhelming sense of gratitude and admiration, and I couldn't let the moment pass without expressing my heartfelt thanks to my supervisor Dr. Mandeep Singh (Dr. M.D. Singh; Associate Professor - EIED), for his generous and proficient support, valuable suggestions, and enthusiastic interest throughout the study. As I move forward in life, I carry with me the invaluable lessons he imparted, and I know these lessons will serve as a strong foundation for my future endeavors. I will forever be grateful for the privilege of being his student and I am so fortunate to have the opportunity to work with him.

I am deeply thankful to Prof. Dr. Mandeep Singh for developing an attitude toward research and enhancing my perspective. His role as an advisor has extended beyond academics; he has taught me valuable life lessons in perseverance, resilience, and the importance of continuous learning. He has been a role model for integrity, empathy, and compassion, traits I aspire to emulate in my own life. I wanted to express my sincere gratitude to the distinguished members of my Doctoral Committee: Dr. Saurabh Bhardwaj, Prof. Dr. Sunil Kumar Singla, and Dr. Vinay Kumar for their constructive criticism, patience, and guidance, which made learning enjoyable and enlightening.

I also express my gratitude to Dr. Usha Panjwani, Dr. Koushik Ray, and Dr. Savita Gaur, scientists at DIPAS, Defence Research, and Development Organisation India, for helping me in sponsoring the research project.

I am also thankful to my Ph.D. friends Amandeep Aulakh, Amandeep Cheema, Abhinandan Aggarwal, and Puneet Arya for their worthy help and support during data acquisition. Their voluntary participation and teamwork contributed to my success. I am grateful for the opportunity to work with a talented team that has not only made this project a triumph but also a truly enjoyable experience. Furthermore, I would like to thank our lab staff Gaurav Kumar and Jaideep Singh, and office staff Arvind Verma and Sunil Kumar for their generous support throughout the whole academic journey.

My acknowledgments would not be complete without expressing my gratitude to my family for standing by me. I am indebted to my father Dr. Narinder Singh and mother Balwinder Kaur for their consistent support, encouragement, and affection. I extend my heartfelt appreciation and thanks to my wife Ramandeep Kaur, daughter Sahaj Kaur, and son Harjp Singh for always being compassionate. Their unconditional support and presence in my life have been a persistent source of strength and motivation.

Jitender Singh
23/01/2024

Jitender Singh Virk

Dedicated
to
My Family

Abstract

Sleep is a fundamental biological need that is often overlooked in today's fast-paced and busy world. In our pursuit of productivity and success, many individuals sacrifice sleep, unaware of the profound consequences. A sleep-deprived fatigued person is prone to commit mistakes that could lead towards fatality. In today's era, sleep deprivation has become a pervasive problem that affects people of all ages and walks of life. With increasing responsibilities, technological distractions, and lifestyle choices, many individuals fail to prioritize sleep, leading to a chronic lack of rest. Youngsters spending hours on phone screens before going to bed is also another reason for prevailing sleep deprivation.

Sleep deprivation concerns the insufficient sleep needed due to which an individual does not feel fully rested and alert during the daytime. While the recommended duration varies with age, for optimal functioning, adults need 7-9 hours of sleep per night. However, surveys indicate that a significant portion of the population, especially in urban settings, regularly experiences sleep deprivation. Factors contributing to this epidemic include long working hours, excessive screen time, social obligations, and untreated sleep disorders. Therefore, it is essential to assess the fatigue. Furthermore, sleep-deprived individuals experience slower reaction times, reduced alertness, and impaired decision-making skills. These cognitive deficits can be particularly dangerous in critical situations, such as when operating heavy machinery or driving, leading to an increased risk of accidents and injuries. Sleep deprivation undermines performance and productivity across all aspects of life. In the workplace, it can lead to reduced efficiency, decreased creativity, and an increased number of errors. The employees who are Sleep-deprived are likely to struggle with decision-making and focusing, hindering their ability to perform at their best.

While reviewing the available literature we have identified that the most of existing methods to detect fatigue are uni-modal and intrusive and the major demerit of these methods is that an individual remains aware of being monitored while obtaining the samples, which results in biases. The proposed methodology is a novel approach in terms of detecting the fatigue non-intrusively and through multimodal feature fusion. In our approach, fatigue is detected by obtaining significant features from four domains: visual facial images, thermal images, voice analysis, and keystroke analysis. In the proposed approach, the samples were acquired from the volunteers; considered as subjects for feature extraction, and empirical weights were

assigned to the four different domains. Healthy and young individuals between the age group of 20 to 30 years age group participated voluntarily in the experimental study. Additionally, they abstained from consuming alcohol, tea, and coffee; or medication affecting the pattern of sleep. Initially, the model was trained with 80 samples (40 from Alert subjects and 40 from Fatigued subjects). Afterward, we tested it on the remaining 40 samples (20 from Alert subjects and 20 from fatigued subjects) and finally cross-validation was performed to evaluate the baseline performance. The results are finally compared with five different classifiers: k-nearest neighbors (kNN), random tree, support vector machines (SVM), random forest, and multilayer perceptron classifiers. The proposed multimodal feature fusion methodology has achieved an average accuracy of 93.33% in 3-fold cross-validation to detect fatigue non-intrusively.

To enhance the accuracy, we implemented the adaptive learning method and performed trials on 60 subjects in two states (Alert and Fatigued). We obtained a total of 120 samples, 60 samples for the alert state and 60 samples for the fatigued state. Initially, we experimented it only for one domain i.e., through eye cues. In this method, the 'image profile' function was used upon the segmented images of eyes to categorize the open-eye frames and closed-eye frames. Additionally, we have identified a new feature utilizing the inter-frame interval of a closed and open eye to differentiate the Alert and Fatigued state through an adaptive threshold. The proposed approach is self-learning that works in real-time, and it achieved a detection accuracy of 97.5 %. Its accuracy and sensitivity were likely to be increased further by increasing the size of the data set because it is self-learning and real-time in nature.

Subsequently, we tried the adaptive learning methodology on multimodal features and increased the sample size to 180 samples, 80 samples for the alert state and 80 samples for the fatigued state. We have proposed a self-learning framework for multimodal feature fusion from four different domains (visual spectra images, thermal images, voice, and keystrokes). We trained our model with 120 samples (60 for the Alert state and 60 for Fatigued state) and then tested it on the remaining 60 samples (30 from the Alert state and 30 for Fatigued state). The results are compared with k-nearest neighbors (kNN), support vector machines (SVM), random tree, random forest, and multilayer perceptron classifiers for cross-validation. The highest accuracy of 98.33% was achieved to detect fatigue non-intrusively. This experiment shows that the performance is improved after implementing the adaptive self-learning technique on the multimodal feature fusion.

Table of Contents

Title	Page No.
Certificate	2
Acknowledgments	3
Abstract	6
List of Figures	11
List of Tables	12
List of Abbreviations	13
Chapter 1 Introduction and Objectives	18
1.1 Introduction	18
1.2 Motivation	19
1.3 Objectives	21
1.4 Organization of the Thesis	22
Chapter 2 Literature Review	23
2.1 Literature review related to image analysis	23
2.2 Literature review related to voice analysis	28
2.3 Literature review related to reflex analysis	30
2.4 Recent Literature Based on Relevant Modern Techniques	32
2.5 Research Gaps	39

Chapter 3	Multimodal Feature Fusion for Fatigue Detection	41
3.1	Introduction	41
3.2	Material	42
3.3	Subjects	42
3.4	Methodology	43
	3.4.1 Data acquisition	45
	3.4.2 Feature extraction	45
	3.4.3. Multimodal feature fusion	52
	3.4.4. Classification	55
3.5	Results and discussion	56
3.6	Conclusion and summary	57
Chapter 4	Detecting Fatigue Through Eye Cues Using Adaptive Learning	58
4.1	Introduction	58
4.2	Novelty	59
4.3	Material	59
4.4	Subjects	60
4.5	Dataset	60
4.6	Methodology	60
	4.6.1. Pre-processing	61
	4.6.2. Morphological operations	62

	4.6.3. Features extraction	62
	4.6.4. Detection of open and closed eye pair	64
	4.6.5. Classification of Alert and Fatigue State	65
4.7	Results and discussion	69
4.8	Conclusion and summary	70
Chapter 5	Adaptive Self-Learning for Multimodal Feature Fusion to Detect Sleep Deprived Fatigue	72
5.1	Introduction	72
5.2	Subjects	72
5.3	Methodology	73
5.4	Results and discussion	75
5.5	Conclusion and summary	75
Chapter 6	Major Findings and Future Scope	77
6.1	Major findings	77
6.2	Challenges	79
6.3	Future Scope	80
Annexure-I	SCI Publications and Patent	82
Annexure-II	References	83

List of Figures

Figure	Page No.
Figure 1.1: Cause and effect of sleep deprivation	19
Figure 3.1: Representation of experimental setup	42
Figure 3.2: Exhibits methodology for fatigue detection using multimodal fusion	44
Figure 3.3: Representation of Graphical User Interface (GUI) window	45
Figure 3.4: Depiction of thermal image processing	48
Figure 3.5: Active Vs. Fatigue Keystroke pattern for the same word and same subject	50
Figure 3.6: Active Vs. Fatigue Keypress count	51
Figure 3.7: Active Vs. Fatigue Voice pattern	52
Figure 3.8: Exhibits methodology for fatigue detection using multimodal fusion	54
Figure 4.1: Experimental setup	59
Figure 4.2: Methodological representation of a comprehensive operational process	61
Figure 4.3: Represents the sequential execution of morphological operations	62
Figure 4.4: Pre-processing steps from RGB image to binary segmentation	63
Figure 4.5: Discriminating open and closed eye state	64
Figure 4.6: Illustrating discrimination in alert eye state and fatigued eye state pattern	65
Figure 4.7: Implementation methodology	68
Figure 4.8: Five new features and their ability to differentiate between two classes	70
Figure 5.1: Multimodal fusion with adaptive self-learning framework	74

List of Tables

Table	Page No.
Table 3.1: List of features from visual spectra facial image	47
Table 3.2: List of features from the thermal image	47
Table 3.3: List of features from Keystroke Dynamics	49
Table 3.4: List of vocal features	52
Table 3.5: Detection accuracy matrix from the individual domains	53
Table 3.6: Multimodal Feature Fusion classification accuracy	55
Table 3.7: Detection accuracy matrix from the individual and multimodal domains	56
Table 3.8: Confusion Matrix	57
Table 4.1 Significant features after feature reduction through FDR	66
Table 4.2: Alert and Fatigue state sample features of 5 subjects	69
Table 4.3: Confusion Matrix for Fatigued State	71
Table 4.4: Comparative accuracy with different classifiers	71
Table 5.1: Comparison of results with other classifiers for adaptive self-learning	75
Table 5.2: Accuracy of individual and multimodal through adaptive self-learning	76

List of abbreviations

Abbreviation	Full-Form
2D	Two Dimensional
3D	Three Dimensional
AAM	Active Appearance Model
Adaboost	Adaptive Boosting
AECS	Average of the Eye Closure Speed
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANOVA	Analysis Of Variance
ASCII	American Standard Code for Information Interchange
AU	Action Unit
AVD	ASCII Value Difference
BCNN	Bidirectional Convolutional Neural Network
BF	Blinking Frequency
BFN	Brain Function Network
BKC	Eyes Blink Count
BN	Bayesian Network
BOW	Bag of Words
CED	Closed Eye Duration
CNN	Convolutional Neural Network
COCP	Close To Open Changeover Period
CRE	Character Error
CRN	Corresponding Random Network
DAF	Differential Active Appearance Model
DAFs	Differential Active Appearance Model Features
DBN	Dynamic Bayesian Network
DFT	Discrete Fourier Transform

DHD	Directed Hausdorff Distance
DIPAS	Defence Institute of Physiology and Allied Sciences
DRDO	Defence Research, and Development Organisation
DVD	Digital Versatile Disk
ECP	Eyes Close Period
EEG	Electroencephalographic
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
EOG	Electro Oculography
ESM	Eyes Screening Mechanism
FCBF	Fast Correlation-Based Filtering
FDR	Fisher's Discriminant Ratio
FES	Fuzzy Expert System
FFP	Fundamental Frequency (Pitch)
FHFa	Angular Sum of Forehead region
FHFr	Radial Sum of Forehead Region
FOM	Frequency of Open Mouth
FPS	Fourier Power Spectrum
GA	Genetic Algorithm
GazeDis	Spatial Distribution of the Eye Gaze
GFM	Generalized Fuzzy Model
GRDB	Gradient Residual Dense Block
GSC	Galvanic Skin Conductance
GSR	Galvanic Skin Resistance
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HD	High Definition
HDD	Hard Disk Drive
HDTP	Histograms Dynamic Thermal Patterns
HMM	Hidden Markov Model

HOG	Histogram of Oriented Gradients
HR	Heart Rate
ICFC	Inter Changeover Frame Count
ID	Identity
IFICO	Inter Frame Interval Close to Open
IFIOC	Inter Frame Interval Open to Close
IKCCA	Improved Kernel Canonical Correlation Analysis
IR	Infra Red
IVIS	In-Vehicle Information Systems
JAFFE	Japanese Female Facial Expression
kNN	k Nearest Neighbors
KSR	Keystroke Rate
LAN	Local Area Network
LBP	Local Binary Patterns
LBP-TOP	Local Binary Patterns on Three Planes
LCT	Lane Change Test
LDA	Linear Discriminant Analysis,
LED	Light Emitting Diode
LTP	Local Ternary Pattern
MFCC	Mel-Frequency Cepstral Coefficient
ML	Machine Learning
MLBP	Multi-Resolution Local Binary Pattern
MMHCI	Multimodal Human-Computer Interaction
NBA	Naive Bayes Algorithm
NF	Nodding Frequency
NLD	Non-Linear Dynamics
NVIE	Natural Visible and Infrared Facial Expression
OCOP	Open to Close Changeover Period
OCR	Optical Character Recognition
OED	Open Eye Duration

P.C.	Personal Computer
PCA	Principal Component Analysis
PERCLOS	Percentage of Eyelid Closure
PerLPD	Large Pupil Dilation Percentage
PERLVO	Vertically Large Open Mouth Percentage
PerSac	Saccadic Eye Movement Percentage
PERTITL	Head Tilting Percentage
PI	Personal Identification
PNS	Non-Steer Percentage
PRFa	Angular Sum of Periorbital Region
PRFr	Radial Sum of Periorbital Region
PRV	Variation of Pupil Ratio
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
RAM	Random Access Memory
RGB	Red Green Blue
ROI	Region of Interest
RSH	Rate of Speech
SDM	Supervised Descent Method
s-FCM	Spatial fuzzy C-means Clustering
SPD	Speech Duration
SPL	Sound Pressure Level
SRC	Sparse Representation Based Classification
SSS	Stanford Sleepiness Scale
SVM	Support Vector Machine
TBC	Total Blink Count
TCD	Transcranial Doppler
TCEF	Total Closed Eye Frames
TCT	Total Changeover Time
TIFI	Total Interval Frame Interval

TOEF	Total Open Eye Frames
TS	Thermal Spectra
TST	Total String Time
TTCO	Transition Time Close to Open
TTOC	Transition Time Open to Close
TTT	Total Transition Time
USB	Universal Serial Bus
VS	Visual Spectra
WMFCC	Weighted Mel-Frequency Cepstral Coefficient

Chapter 1

Introduction and Objectives

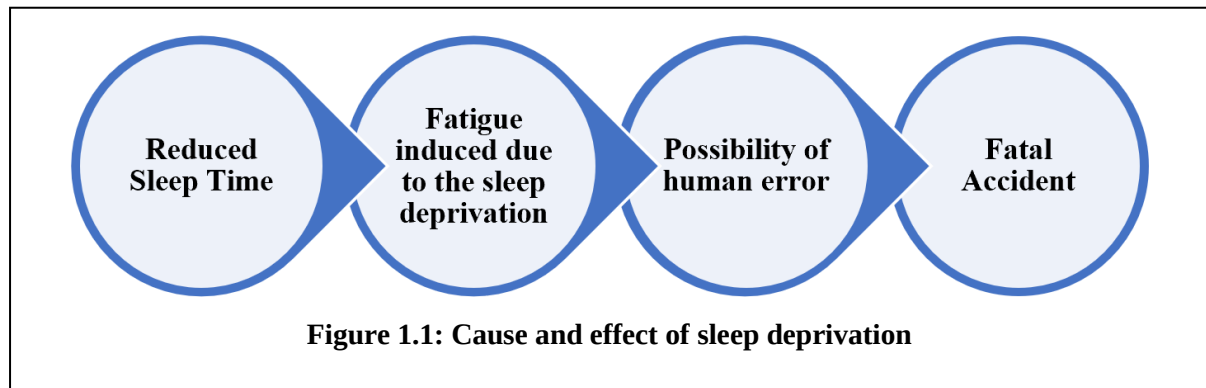
1.1 Introduction

Sleep, often regarded as a natural and effortless activity, holds a profound significance in our lives. It is a natural physiological process that permits our body and mind to have rest, repair, and rejuvenate. A well-rested mind exhibits sharper focus, improved problem-solving skills, and enhanced creativity. On the other hand, sleep deprivation impairs cognitive performance, leading to reduced attention span, slower reaction times, and compromised decision-making abilities [1]. Sleep deprivation, an increasingly prevalent issue in modern society, is a condition where individuals fail to obtain adequate and restful sleep. In the pursuit of productivity, social engagements, and technology distractions, many people prioritize their daily activities over the essential need for sleep. Sleep deprivation occurs when individuals consistently fail to obtain the recommended amount of sleep needed for optimal functioning.

The National Sleep Foundation advocates that adults should aim for 7-9 hours of sleep per night, while teenagers and children require even more [2]. However, various factors can contribute to sleep deprivation, such as work demands, social activities, family responsibilities, and the misuse of technology.

Our modern society is highly influenced by the multitasking culture in the present era. People are performing various activities simultaneously. They play multiple roles altogether, and there is a race against time. Nowadays, numerous professions, such as pilots, mariners, aviators, surgeons, and air traffic controllers, demand consistent vigilance for an extended duration. In these professions, exceptional attention is required to perform error-free tasks, and any negligence may result in a fatal accident. Sleep quality is often compromised in individuals working for an extended duration and frequently involved in shift work, which is more common in various professions. Sleep deprivation and altered circadian rhythm affect an individual's cognitive performance [3]. Social media influence and spending more time on smartphones are reasons for reduced sleeping time. Cognitive impairment leads to fatigue and a decline in attention at the workplace. In most cases, sleep deprivation is the root cause of the lack of alertness, leading to human error and may cause an accident. Figure 1.1 illustrates the cause and effect of sleep deprivation. Primary signs and symptoms of sleep deprivation include

excessive daytime sleepiness and reduced concentration. To ensure an individual's attentiveness, we need to ensure that he must have taken a sound sleep.



People with too much daytime sleepiness have difficulty staying awake even when needed and may feel drowsiness. This results in microsleeps sometimes, often in which a person dozes off for a few seconds [4]. An individual's symptoms entirely depend on the extent of the amount of sleep deprivation.

1.2 Motivation

Many people get inadequate sleep because they believe sleep deprivation is usual. An individual's symptoms are dependent on their sleep requirement. Rather than taking the necessary steps to have more sleep, they consume tea, coffee, or other energy drinks or simply try to "power through." Like other vital needs, e.g., breathing, drinking, and eating, sleeping is also a basic human need. Being well-rested is directly linked to increased physical and mental alertness. Individuals who get enough sleep exhibit improved focus, quicker reaction times, and heightened cognitive performance. On the contrary, sleep-deprived individuals often experience reduced attention spans, impaired motor skills, and diminished ability to concentrate. Numerous fatal accidents would have been prevented by recognizing the fatigue timely, and suitable preventive measures could have been taken [5]. Hence, detecting the fatigue induced due to sleep deprivation is essential.

Our research proposal is sponsored by the Defence Institute of Physiology & Allied Sciences (DIPAS)-DRDO. DIPAS-DRDO aimed to develop a non-intrusive system for detecting the fatigue state of Air Force pilots. The present research proposes to detect fatigue induced through sleep deprivation using a fusion of multiple sensor inputs in a non-intrusive manner. During a research study, DIPAS-DRDO discovered that Air Force pilots are prone to fatigue due to sleep deprivation. Fatigue due to sleep deprivation can have severe consequences for

Air Force pilots, as it impairs various cognitive and physical functions essential for safe and effective aviation operations. Sleep deprivation negatively affects cognitive abilities such as attention, memory, and decision-making. Pilots need to process vast amounts of information quickly and accurately during flight, and sleep deprivation can compromise these critical cognitive functions.

Lack of sleep leads to decreased alertness and vigilance. Pilots must maintain a high level of attention to monitor instruments, detect potential threats, and respond to changes in the environment. Sleep-deprived individuals are more prone to lapses in attention, increasing the risk of errors. Sleep deprivation slows down reaction times, which is particularly dangerous in situations where quick and precise responses are crucial. Pilots need to respond promptly to emergencies, navigate rapidly changing conditions, and avoid collisions. Sleep-deprived individuals often struggle with problem-solving and critical thinking. In aviation, where unexpected challenges can arise, impaired problem-solving skills can lead to suboptimal decision-making and difficulty managing complex situations. Moreover, microsleeps are brief episodes of sleep that occur involuntarily. They can last for just a few seconds but result in a complete loss of awareness. Pilots experiencing microsleeps, even if brief, can miss important information and jeopardize the safety of the flight. Also, Effective communication is essential in aviation, both within the cockpit and with air traffic control. Sleep-deprived pilots may struggle to communicate clearly and coherently, leading to misunderstandings and potential errors. Furthermore, inattention blindness refers to the phenomenon where individuals fail to perceive unexpected stimuli in their visual field. Sleep-deprived pilots may be more susceptible to inattention blindness, potentially overlooking critical information or hazards. In addition, precise motor skills are crucial for flying an aircraft, especially during take-offs, landing, and maneuvers. Sleep deprivation can impair fine motor coordination and manual dexterity, increasing the risk of handling errors.

The detection of fatigue in Air Force pilots is advantageous for several reasons, as it directly impacts the safety and effectiveness of flight operations. Detecting fatigue helps prevent pilots from operating aircraft when their mental and physical states are compromised, reducing the risk of accidents. Fatigue is a known factor contributing to aviation accidents. By identifying signs of fatigue in pilots, measures can be taken to prevent potential mishaps and ensure the overall safety of flight operations. Detecting fatigue allows for proactive measures to be taken, such as adjusting flight schedules or providing additional support, to ensure pilots maintain optimal situational awareness during missions. Identifying signs of fatigue enables

commanders and aviation medical personnel to make informed decisions regarding pilot assignments, ensuring that pilots in peak conditions are assigned critical tasks.

Pilots who are well-rested and alert are more likely to execute missions with precision and effectiveness. Fatigue detection contributes to maintaining a high level of operational readiness, ensuring that air-force missions are carried out successfully. Detection of fatigue allows for the adjustment of work schedules and duty hours. Moreover, it is essential for maintaining peak performance and mission readiness. Implementing optimized schedules helps prevent overexertion and ensures that pilots have sufficient time for rest and recovery between missions. Accidents and errors resulting from fatigue can be costly in terms of both human lives and equipment. Detecting and managing fatigue helps to prevent incidents, leading to potential cost savings for the Air Force.

1.3 Objectives

Research objectives are crucial components of any research study as they provide a clear direction and purpose for the investigation. These objectives enabled to structure of the study effectively and served as a roadmap, helping in designing the methodology, collecting relevant data, and analyzing it in a systematic and organized manner. It ensures relevancy by specifying the purpose of the research work. At the end of our investigation, we compared the achieved results against the initial objectives to determine the extent to which the research has met its intended goals.

In the present work, an effort is made to develop a computer-aided system to assess fatigue due to sleep deprivation. The work has been carried out with an orientation toward the investigation of fatigue induced due to inadequate sleep. The four objectives of the research work are as follows:

1. To identify the significant physiological parameters from different domains for detecting sleep deprived induced fatigue
2. To study the relation between alertness and fatigue due to sleep deprivation
3. To develop a computer-aided system for sleep deprived induced fatigue detection using multimodal features
4. To test and validate the developed classifier for fatigue detection

1.5 Organization of the Thesis

The organization of the thesis follows a logical flow of ideas, making it convenient to navigate through the structure with well-defined sections and headings allowing us to locate specific information quickly and efficiently. This provides a general outline of the chapters. Chapter 1 explains the background, motivation, and objectives, which is an introduction to the present work. This chapter also includes the challenges faced during the present research work. Chapter 2 contains the summary of the prior art and contains a review of available literature from all four domains i.e., visual facial images, thermal images, voice analysis, and body reflexes. Moreover, research gaps till the present date are identified and documented in this chapter. Chapter 3 is about the multimodal feature fusion technique. This chapter contains information about the methodology, experimental setup, and results of the feature fusion technique for fatigue detection. Chapter 4 contains the methodology for detecting fatigue through eye cues by implementing an adaptive learning method on the features from one domain. Chapter 5 describes the implementation of adaptive learning on multimodal feature fusion to enhance the results. Chapter 6 contains the discussions on major findings, the outcome of the research, and the future prospects of this research work.

Chapter 2

Literature Review

The available literature has been reviewed, it involves analyzing existing scholarly works and research relevant to our research work. Various modern techniques and concepts related to fatigue detection and sleep deprivation are also considered. In previous research work, various intrusive and nonintrusive approaches were applied for detecting fatigue. This research work is aimed at investigating multimodal systems to detect fatigue induced due to sleep deprivation. Hence, we have considered intrusive and nonintrusive techniques and recent research considering all four domains for the study of available. This literature review is further subdivided into four sub-parts for fatigue detection through facial expressions, voice analysis, reflexes, and other relevant literature indirectly associated with the research work. The resume of the relevant literature in chronological order is presented in this section.

2.1 Literature review related to image analysis

Fatigue detection through visual facial features is an active research area, and numerous systems are developed for fatigue detection through visual facial features, incorporating eye-tracking for eye movements, such as longer eye blinks. Fatigue can also be detected by analyzing several other visual facial features, such as drooping eyelids, redness in the eyes, and changes in the expressions of the face. In the recent past, the non-intrusive fatigue detection method was reported in 2004 by Ji Q. et al. A non-intrusive prototype computer vision system for monitoring driver's vigilance employing the Bayesian Network model was developed. They used some facial features like eyelid movement, head movement, etc. They reported that their system is reasonably accurate in characterizing human fatigue [5]. However, they determined the fatigue was due to continuous driving only.

A non-invasive method that infers user stress in real-time from various modalities is presented by Liao W. et al. They provided a Dynamic Bayesian Network (DBN) structure to represent fatigue [6]. The optical character recognition (OCR) algorithm was applied by Dinges D.F. et al. to detect the changes in facial expressions when individuals feel fatigued during the performance. A computer vision application to distinguish the difference in human facial

expressions during a stretched spaceflight duration may be a practical approach to recognizing fatigue during critical operations [7].

A machine learning scheme to determine actual human behavior during drowsiness is employed by Vural E. et al. Automatic classifiers for 30 facial actions from the Facial Action Coding system were developed using machine learning on a separate database of spontaneous expressions. The system could predict sleep with approximately 90% accuracy across subjects [8]. Similarly, an approach to detect the drowsiness that works on greyscale images was proposed by Flores M.J. et al. in 2008. They considered eye blinking rate as an indicator to detect driver's fatigue [9]. A facial expression detection technique is proposed by Cheon Y. et al. that detects sequential facial expression images through the differential active appearance model (AAM). The difference in the AAM parameters between a reference (neutral expression) face image and an input face image is used to compute the differential AAM features (DAFs). The experiment exhibits that the DAFs enhance facial expression detection over conventional AAM features by 20%, and the sequence-based k-NN (k Nearest Neighbors) classifier demonstrates 95% facial expression recognition [10].

Fan X. et al. conceptualized monitoring human fatigue from facial image sequences by deploying the Gabor-based dynamic representation. The fused image sequence is divided into rectangular sub-image sequences as dynamic units to get the dynamic features. Then a histogram of each dynamic unit is computed and combined. An Ada-Boost algorithm is applied to select a subset of the most discriminative dynamic features and build a robust classifier [11]. Likewise, a vision system for monitoring driver's vigilance is presented by Yao K.P. et al. The level of alertness is determined by integrating several facial features in the input video sequence. Estimated parametric values are collected and analyzed at fixed intervals to provide real-time vigilance [12]. Brookhuis K.A. et al. discussed traffic accidents caused by inadequate mental workload, i.e., either too low vigilance or too high stress. The modern driving simulator environment controls environmental conditions, scenarios, and stimuli and enables physiological measurement of heart rate, brain activity, etc. Research in this field is often ethically unacceptable. Therefore, advanced driving simulators are profitable in two ways; they enable the study of realistic conditions without any objective risks [13].

Over and above the effects of in-vehicle information systems (IVIS), usage of eye blinks in a simulated Lane Change Test (LCT) are examined by Benedetto S. et al. a driving task designed by the International Organization for Standardization [14]. Moore S. et al. attempted to produce pose-invariant facial expression classifiers. A sequential 2-stage approach is taken for facial

expression classification to investigate the effects of yaw variations from frontal to profile views. Local binary patterns (LBPs) and variations of LBPs as texture descriptors are investigated. Feature vectors contain concatenated feature histograms built from each sub-block. Multi-class support vector machines are adopted for facial expression classifiers [15]. Whereas recognizing facial expressions on multiple manifolds, a comprehensive structure is applied by Xiao R. et al. Initially, each expression's inherent features are learned independently, and then a genetic algorithm (GA) is employed [16]. While Ji Y. et al. worked on Facial expression machine analysis in the Human-Computer Interaction (HCI) context. They proposed two new features on the Spatio-temporal plane to describe the micro transformation of a human face and facial expressions. The SVM (Support Vector Machine) system efficiently recognizes a single image's presentation in a sequence. Then, the probabilities of all the frames are used to predict the class [17]. Sandbach G. et. al. proposed a facial expression identification technique that exploits 3D motion-based features. Results demonstrated that 3D information improves the identification accuracy compared to 2D data in an automatic manner [18]. As well as the characteristics of facial expressions and gestures for automatic affect identification are discussed by Chen S. et al., They employed an Image Histogram of Oriented Gradients (HOG) features through temporal normalization and a bag of words (BOW) to acquire the desired information [19].

Analysis and recognition of facial expressions in continuous videos are addressed by Dornaika F. et al. They studied classifiers' performance that exploits head pose-independent temporal facial action parameters. The results show that using dimension reduction techniques on the extracted time series can improve classification performance. Moreover, these experiments show that the best recognition rate can be above 90% [20]. Forsman P.M. et al. sought to develop a methodology to detect driver fatigue before the imminent accident risk. They evaluated eighty-seven different driver drowsiness detection metrics on twenty-nine participants subjected to a night shift situation for ten simulated workdays for other driving sessions. Afterward, they correlated with an independent measure of fatigue, i.e., on a psychomotor vigilance test administered before each drive. Also, they simulated their findings across eight curved road segments for validation [21]. A method for facial expression recognition, combined curvelet transform, is proposed by Tang M. et al. with an improved support vector machine (SVM) based on particle swarm optimization (PSO). Curvelet transform is applied to extract facial expression features [22]. A method based on natural 3D visual features for human emotion recognition is introduced by Yun T. et al. The 3D geometric

information and color density information of the facial expressions are extracted by the 3D Gabor library to construct visual feature vectors. An improved kernel canonical correlation analysis (IKCCA) algorithm is adopted for the final decision [23].

In 2013, Sharma N. et al. used thermal imaging of facial regions. Their work involves facial regions captured in videos in thermal (TS) and visible (VS) spectra and introduces their stress database. Results show that the stress recognition rate was 72%, obtained from HDTP (Histograms Dynamic Thermal Patterns) features fused with Local Binary Patterns on three planes (LBP-TOP) features [24]. A stress detection system based on the analysis of facial expressions was developed by Gao H. et al. A near-infrared camera is used to capture the face's frontal view, and then a face tracker is applied to track facial landmarks. A pose normalization step compensates for pose-mismatch due to head movement and camera view. Their results demonstrated that the system has a more than 90% detection rate for indoor tests [25]. Eskil M.T. et al. proposed an approach to facial expression recognition that capitalizes on the anatomical structure of the human face. Facial muscle influence regions are estimated through a semi-automatic customization process and projected to the image plane to determine feature points. The relative displacement of each feature point between two image frames is treated as evidence of muscular activity. Feature point displacements are projected back to the 3D space to estimate the new coordinates of the wireframe vertices. They demonstrated muscle force's representative power based on three classifiers, NBA, SVM, and Adaboost [26].

A statistical pattern recognition approach to locating landmarks in an image is presented by Benitez-Quiroz C.F. et al. explored the fact that several object classes exhibit a homogeneous structure such that each landmark position provides information about the other landmarks' status [27]. Similarly, an approach to the face recognition problem using directional and texture information from face images is proposed by Mehta R. et al. For texture feature extraction, a low-dimensional and computationally effective local descriptor was utilized in 2014 [28]. Methodology for detecting face and facial features is introduced by Naruniec J. et al. The experiment focuses on detecting the eye centers, which are crucial in many computer vision systems such as face recognition, eye movement detection, or iris recognition [29]. Several techniques to improve the recognition rate and execution time of facial expression recognition systems are employed by Owusu E. et al. Face detection was carried out by applying the Viola-Jones descriptor. An algorithm was formulated to select Gabor wavelets from the extracted features. A multilayer feed-forward neural network is trained with sample datasets from JAFFE

and Yale databases. The average recognition rate in JAFFE dataset is 96.83% and in the Yale dataset, it is 92.22% [30].

A framework for emotion detection from participants' fixation patterns is proposed by Vaidya A.R. et al. They stated that detecting the most subtle emotions is best through fixations to the eyes. The results suggest another viewpoint on several puzzling dissociations in the analytic and neuropsychological approach to the study of eye gaze [31]. In addition, Wang S. et al. proposed an implicit multiple-emotional video tagging system. The audience's expressions are inferred through a multi-expression recognition model consisting of an image-driven expression measurement recognition and a Bayesian network. The video tags are obtained from the recognized expressions by another Bayesian Network, capturing the relations between expressions and emotions. The experiments conducted on the JAFFE and NVIE databases demonstrate that expression recognition performance is improved by considering the relations among multiple expressions [32]. Vitale J. et al. emphasized the skills of intelligent agents and robots to operate within social environments. This work used Simulation Theory and neuroscience findings on Mirror Neuron Systems as the basis for a novel computational model to handle affective facial expressions for neuropsychological research. The model is based on a probabilistic mapping of observations from multiple identities [33].

Moreover, problems related to the face detectors developed in the last two decades for images are discussed by Yan J. et al. They proposed a hierarchical part-based structural context model to explicitly capture facial images. Also, they presented a phrase-based representation for body detection and jointly encoded the outputs of face and body detectors [34]. A non-intrusive fatigue detection system based on video analysis of drivers is demonstrated by Azim T. et al. The parameters used for detecting fatigue are eye closure duration and yawning. The information about the eyes and mouth is passed to the Fuzzy Expert System (FES) which classifies the driver's state. The system has been tested with different sequences recorded in day and night driving conditions and with users belonging to other races and genders [35].

Besides this, McIntire L.K. et al. showed that sustained attention or vigilance declines due to the extended duration of task performance. As performance is reduced, blink rate and duration increase. Results suggest that eye blink frequency and duration may be considered an indicator of arousal level [36]. The empirical mode decomposition (EMD) technique for facial emotion recognition is proposed by Ali H. et al. In this method, facial features are obtained from successive projections of the Radon transform of a 2D image [37]. In 2015, a real-time 3D

facial Action Unit (AU) intensity estimation and emotion recognition system, presented by Zhang Y. et al. was evaluated with the Bosphorus 3D database [38].

Recently, a smile detection approach based on an Extreme Learning Machine (ELM) was proposed by An L. et al. It is tested using the SVM and LDA classifiers [39]. Smith R.S. et al. used a multi-class classifier to detect a face. Feature extraction is performed by generating many multi-resolution local binary pattern (MLBP) features and then selecting from these using fast correlation-based filtering (FCBF). A weighted decoding scheme is proposed with the weights computed using first-order Walsh coefficients. The bias and variance properties of the classifier are measured, and sources of error can be reduced through bootstrapping and weighted decoding [40]. Ouyang Y. et al. proposed a method that is composed of three components. First, Histograms of Oriented Gradients (HOG) and Local Binary Patterns (LBP) were employed to extract the features. Second, Sparse Representation-based Classification (SRC) is used. Finally, the Cohn–Kanade database experiments show that the proposed method performs better than many existing methods [41]. Pawar R. et al. propose a Drowsiness Detection System for preventing accidents caused due to drowsy driving. Their system was based on a Convolutional Neural Network (CNN) model interfaced with a Raspberry Pi microcontroller and a webcam to capture visual facial images of the driver [42]. An algorithm to detect drowsiness by analyzing human face structure from video samples was proposed by Inthanon P. et al. Their study exhibits the significance of facial landmarks from eyes and mouth components that can help to detect drowsiness effectively by using the Nvidia Nano Jetson Tool. Nvidia Jetson Nano is a tool that evaluates images by tracking eye motions above 35 FPS and yawning above 50 FPS to predict drowsiness [43].

2.2 Literature review related to voice analysis

Human voice features can be considered for detecting fatigue due to sleep deprivation. Vocal features include speech analysis to detect the changes in pitch, tone, and rate of speech, which are significant indicators of fatigue. The voice of a fatigued person tends to become more monotone and less expressive. The human voice is also related to a person's drowsiness and alertness; therefore, it is considered for the study. A range of issues associated with the analysis, modeling, and recognition of speech under stress is considered by Hansen J.H.L. et al. They explored how individuals differ in their perception of anxiety, how it affects the speech production system, and the cues associated with perceiving stress [44]. While, Tawari A. et al. used contextual information, which is evident that affect recognition by a human is primarily

influenced by the context information. They introduced features based on cepstrum analysis of pitch and intensity contours and reported an overall accuracy of around 84% [45]. A classifier to recognize affective categories from prosody in natural and acted speech strength is demonstrated by Fernandez R. et al. The work is evaluated on a database of acted emotions compared to human perceptual recognition of five affective categories. Then the model is evaluated on two different corpora of fully spontaneous, naturally occurring speech between people. They obtained 70% detection rate with 30% false alarms [46]. Schuller B. et al. attempt to provide a short overview of automatic recognition of emotion from speech. They described automated processing, including features, classification, robustness, evaluation, implementation, and system integration. They organized the first comparative challenge on emotion recognition from speech, i.e., the INTERSPEECH 2009 Emotion Challenge. They include the Challenges of the database, Sub-Challenges, participants and their approaches, and the fusion of results. Finally, they address the everlasting problems and future promising attempts [47]. Furthermore, a hierarchical computational structure to recognize emotions was introduced by Lee C.C. et al. They evaluated the classification framework using vocal features [48].

Spectrally weighted Mel-frequency cepstral coefficient (WMFCC) features for emotion recognition from speech are proposed by Bozkurt E. et al. The formant locations carry emotion-related information, therefore, critical spectral bands around formant locations can be emphasized when calculating MFCC features. They evaluated the proposed WMFCC features and the standard spectral and prosody features using HMM-based classifiers on the spontaneous FAU Aibo emotional speech corpus. The results show that classifiers with WMFCC features perform significantly better than those with standard spectral features [49]. A variety of classifiers and vocal features for fatigue detection through speech analysis were evaluated by Krajewski J. et al. Within the subject, a partial sleep deprivation model test was conducted on seventy-seven participants for eight consecutive hours, and 372 voice samples were recorded. Bagging procedures achieve the best performance for male and female speakers on phonetics and NLD feature sets. The models for the phonetic feature set achieved 78.3% for males and 68.5% for female speaker classification accuracy. The framework for the NLD feature set achieved 77.2% for males and 76.8% for female speakers [50]. Three novel methods for speaker identification are presented by Bhardwaj S. et al. Two ways utilize both the continuous density hidden Markov model (HMM) and the generalized fuzzy model (GFM), which has the advantages of both the Mamdani and Takagi–Sugeno models. Their method is

inspired by how humans cash in on mutual acquaintances while identifying a speaker. A similar multimodal or hybrid classifier could be used on fatigue-related features [51]. In various experiments, Caraty M.J. et al. used prolonged oral reading corpora to analyze and detect vocal fatigue. Using the audio feature set, a two-class SVM classifier gave an unweighted accuracy of 94.1% for the training set and 68.2% for the test set [52]. Body reflex is an unplanned involuntary action in response to a stimulus. Sleep deprivation also causes muscle fatigue and affects body reflexes, like keyboard typing and hand grips. Jaimes A. et al. studied the methods focused on the body, gesture, and eye gaze, including emotion analysis from audio [53].

2.3 Literature review related to reflex analysis

Jaimes A. et al. reviewed the significant approaches to Human-Computer Interaction, giving an overview of the field from a computer vision perspective. They focused on the body, gesture, gaze, and affective interaction. They discussed multimodal fusion, highlighting challenges, open issues, and emerging applications for Human-Computer Interaction (HCI) research. Also, they emphasized the computer vision approach to recognize gestures [53]. Schleicher R. et. al. examined changes in various oculomotor variables to increase sleepiness in 129 participants. Regarding oculomotor parameters, like blink rate, delay in eyelid reopening, and eyelid closure speed were acknowledged as the finest indicators of drowsiness. Saccadic parameters also exhibit significant changes with an increase in sleepiness. Finally, implications for the design and development of drowsiness detection devices were discussed [54].

The possibility of detecting cognitive and physical stress by monitoring keyboard interactions is explored by Vizer L.M. et al. with the eventual goal of detecting cognitive and physical function changes. Results show that it is possible to classify cognitive and physical stress conditions relative to non-stress conditions based on keystroke and linguistic features [55]. A survey on sensors used to investigate techniques for modeling stress is carried out by Sharma N. et al. using unobtrusive sensors to detect fatigue of those individuals who would need to monitor stress levels regularly (e.g., vehicle drivers and patients with illnesses linked to stress). Further, they discussed various computational techniques to establish optimal sensor fusion for fatigue recognition and classification [56].

Lim Y.M. et al. investigated how keyboard and mouse features can be combined to detect human stress, particularly cognitive stress induced by time pressure and mental arithmetic problems. The results show that the complexity of the mental arithmetic problem and time pressure affect user behavior, mouse behavior, and keyboard behavior significantly. Mouse

movement directions could be an essential variable that stress can affect and must be examined [57]. In addition to this, Nahin A.F.M. et al. discussed an approach to detect emotions by analyzing the body reflexes like keyboard typing patterns of the user and the type of text (words, sentences) typed by them. Several machine learning algorithms analyzed keystroke timing attributes and text patterns. The combined result of keystroke dynamics and text pattern analysis showed improved accuracy over different results, as this combined approach shows above 80% accuracy in identifying emotions [58].

Cui Z. et al. designed a lightweight neural network model for fatigue detection. It can quickly identify the opening and closing states of the driver's eyes and mouth in a video. Afterward, it calculates the driver's PERCLOS (Percentage of Eyelid Closure over the Pupil) and FOM (Frequency of Open Mouth) according to the coding sequence. Finally, the multi-feature fusion judgment algorithm is used to realize the judgment of the driver's fatigue state through multiple reflexes. The experimental results show that their method has an accuracy rate of 98.30% for drowsiness and yawning behaviors in a real vehicle environment [59].

Xu T. et al. proposed a unified framework to simultaneously perform personal identification (PI) and driving fatigue detection using a convolutional attention neural network (CNN-Attention). The performance was assessed using EEG data collected through a wearable dry-sensor system from 31 healthy subjects undergoing a 90-minute simulated driving task. The proposed scheme achieved the best ($p < 0.01$) performance in both PI (98.5%) and fatigue detection (97.8%). Besides, the spatial-temporal structure of the proposed framework exhibits an optimal balance between classification performance and computational efficiency [60].

Ngai W.K. et al. proposed a multimodal approach that uses 2-channel electroencephalography (EEG) signals for emotion recognition. Also, they studied the use of facial images versus facial depth as the face modality. They adopted the common arousal–valence model of emotions and the convolutional neural network to process the spatiotemporal information for emotion recognition. Experiments were conducted on the modality and emotion data, and the results showed accuracies of 67.8% and 77.0% in valence recognition and arousal recognition, respectively [61]. Zhang J. et al. reviewed the emotion recognition methods based on multi-channel EEG signals and multimodal physiological signals. According to the standard pipeline for emotion recognition, they reviewed different feature extraction (e.g., wavelet transform and non-linear dynamics), feature reduction, and ML classifier design methods (e.g., k-nearest neighbor (KNN), naive Bayesian (NB), support vector machine (SVM) and random forest (RF)).

Furthermore, the EEG rhythms that are highly correlated with emotions are analyzed, and the correlation between different brain areas and emotions is discussed. Also, they evaluated various ML and deep learning algorithms for emotion recognition and suggested several open problems and future research directions in the fast-growing area of AI [62]. Li R. et al. proposed a method to measure a driver's fatigue state by detecting the driver's grip force on the steering wheel while driving. They tested the grip force of 36 drivers on the steering wheel in conscious states (Alert) and fatigue states under actual road driving conditions. Using the Stanford Sleepiness Scale (SSS). They divided drivers into Alert and fatigue Groups. During 20-minute real-road driving trials, they measured the steering wheel grip force, electroencephalogram index ($R = (\alpha + \theta)/\beta$), and blink frequency of each driver synchronously. They observed that ΔF , the difference between the maximum/minimum grip force and the standard deviation of the grip force, σF , for each driver, highly correlated with the driver's fatigue state. In the fatigue state, both ΔF and σF increased significantly. They examined these force indices using analysis of variance (ANOVA) and validated them against the R-value, blink frequency, and the driver's self-reported fatigue state. Using the grip force in fatigue detection, they achieved an overall recognition rate of 86.6% and an individual recognition rate of 88.3% [63]. Hooda R. et al. aimed to understand fatigue, its repercussions, and techniques to detect fatigue using machine learning (ML) approaches. They categorized the methods that can be used to detect fatigue into four diverse groups, i.e., Mathematical Models, Rule-Based Implementation, ML, and Deep Learning. Their study compares various algorithms to find the most promising approach that can be used to detect fatigue [64].

Frost C. et al. reviewed the effects of sleep on firefighter occupational performance and health. They studied quantitative analysis methods involving sleep and occupational performance in firefighters. Relevant data were extracted, and study quality was evaluated with Newcastle-Ottawa Scale. They concluded that firefighters with poor sleep measures had worse cognitive performance and overall health, but acute sleep restriction did not impact physical capabilities. However, chronic insufficient sleep would likely negatively impact physical performance due to a decline in numerous health measures [65].

2.4 Recent Literature Based on Relevant Modern Techniques

Overall, detecting fatigue is a complex task, and it often requires a combination of techniques to achieve accurate results. However, with the improvement in modern technologies and machine learning algorithms, detecting fatigue through multimodal features is becoming easier.

Tuncer T. et al. developed an EEG-based intelligent system for driver fatigue detection. They proposed a framework that includes texture descriptors for fatigue detection. The proposed scheme contains pre-processing, feature generation, informative feature selection, and classification with shallow classifier phases. In the pre-processing, discrete cosine transform, and fast Fourier transform are used together. To improve the detection performance, they utilized discrete wavelet transforms as a pooling method, in which the functional brain network-based feature describes the relationship between fatigue and brain network organization. The proposed framework achieved about 98% classification accuracy for fatigue detection using EEG signals in the experiments [66]. Kartsch V.J. et al. studied the various research efforts made to compute the degree of drowsiness by eye-tracking and analyzing physiological features through EEG signals [67]. Similarly, Ye C. et al. assessed mental fatigue through EEG to prevent the risk of performance degradation. The accuracy of their language understanding experimental approach is 87.9% [68].

Wu N. et al. experimented on the speech signal by extracting the Mel frequency cepstrum coefficient (MFCC) and implementing machine learning techniques for estimating fatigue. They compared various algorithms, including the self-adaption quantum genetic algorithm (SQGA), back propagation neural network (BPNN), k-nearest neighbor (kNN), and support vector machine (SVM). 94.0% detection accuracy was obtained only by SVM [69]. Similarly, Siam A.I. et al. experimented with emotion recognition through facial images and speech. Their real-time deep learning approach included machine vision and speech modalities. Furthermore, they used principal component analysis (PCA) for feature decomposition. They claimed 97% detection accuracy for simulation only [70].

Alnuaim A.A. proposed a real-time approach to emotion detection through speech to develop modules for speech emotion recognition (SER) to support applications related to human-computer interfaces (HCI). In their research, they discussed the results of the various classification approaches and feature fusion to enhance emotion detection accuracy. Unlike most previous studies, their analysis is based on multi-lingual data sets [71]. Tuncer T. et al. proposed a framework based on EEG to detect driver fatigue. In their approach, they presented a hybrid three-layered feature selection method. Their non-intrusive method obtained 97.29% classification accuracy through EEG signals for driver fatigue detection [72].

Fatima B. et al. investigated micro-sleep patterns and developed a cost-effective solution for detecting driver fatigue. They distinguished open and closed eyes through Viola Jones algorithm. Micro-sleep patterns are determined to detect fatigue state and eventually trigger a

warning alarm. They achieved an average accuracy of 96.5% and 95.4% for SVM and AdaBoost, respectively [73].

In recent research, O’Keeffe K. et al. investigated the effect of fatigue on human performance and investigated the efficacy of the existing methods for inducing mental fatigue. In their experiment, they conducted six sessions with 12 participants and performed two cognitive tests; the first was an AX-continuous performance test (AX-CPT), and the second was the TloadDback test [74]. Heaton K.J. et al. worked on the prediction of the changes the performance through electrodermal activity and speech-motor coordination. They interpreted the decline in cognitive performance and alertness with the increase in workload [75]. Sikander G. et al. studied the various techniques for detecting driver fatigue. Nowadays, automakers are installing driver assistance technologies (ADAS) in automobiles which include fatigue detection devices. They reviewed and compared numerous approaches and methodologies for the detection of fatigue through physiological features and vehicular features [76]. Doudou M. et al. proposed an embedded methodology to monitor drivers’ alertness and detect real-time drowsiness levels. Their research is based on vehicle and driver behavior through video and physiological signals to determine the applicability and accuracy of intrusiveness and practical usage viewpoint [77].

In another recent research, Zhao C. et al. emphasized driving safety in the modern era of smart devices and the proliferation of the Internet of Vehicles (IoV). They implemented a Bayesian convolutional neural network (BCNN) data selection strategy and introduced the uncertainty-weighted asynchronous aggregation (UWAA) algorithm [78]. Moreover, Zhang Y. et al. experimented with detecting mental fatigue through EEG signals to reduce the probability of traffic accidents. Their research work is based upon logistic regression, one-way analysis of variance, and recursive feature elimination (logistic-ARFE), and they implemented Gaussian SVM for classification purposes and achieved an accuracy of 73.33% [79].

In one more research, Chen J. et al. developed a BP neural network model for detecting fatigue through eye closure, yawning, and percentage of eye closure time (PERCLOS) from the recorded video. The accuracy of the proposed model with facial expressions increased by 8.4%. The proposed model can filter out artifacts caused due to facial expressions while detecting driver fatigue [80]. Furthermore, Li Y. et al. proposed a lightweight wearable device based upon a convolution neural network for detecting driver fatigue through eye images [81].

Nowadays thermal imaging is contributory for detecting fatigue. Singh J. et al. presented an approach for sinusitis detection using thermography. They assessed the effectiveness of active dynamic thermography (ADT) and passive static thermography (PST) by performing a comparative analysis. Their study exhibits encouraging results for the application of ADT for sinusitis diagnosis. [82]. Rodríguez A.O.R. et al. analyzed the emotional data for brain-computer interfaces through learning environments by collecting vital signs. Brain-computer interfaces provide a means to capture emotional metrics from brain activity [83]. Tan W.R. et al. proposed an approach for human skin detection in color images through fusion strategy. Their approach, an eye detector is used to refine the skin model for a specific person. Their approach reduces computational cost and improves accuracy [84].

Singh S. et al. presented an approach for the Multimodal fusion of medical images to integrate multisensory medical data into a single image. They used a hybrid layer decomposition technique scheme to maximize the structural details and suppress noise and artifacts. Their methodology is based on a convolutional neural network (CNN) [85]. Schirra C.O. et al. worked on cardiac imaging using CT to decrease motion blur and to improve the signal-to-noise ratio. They detected local motion vector fields from high-contrast objects and used motion-compensated filtered back projection reconstruction. This results in enhanced image quality compared to gated reconstruction and SNR is improved due to the reduced level of motion artifacts [86]. Singh S. et al. proposed a methodology for Multimodal medical image sensor fusion (MMISF) for the visualization of vital information acquired through multiple imaging sensors. The proposed MMISF approach is comparatively superior to several other approaches and provides enhanced visually fused images with improved computational measures [87]. Dhanawansa, V. et al. proposed an approach to analyzing the temporal relationship between the 3D head pose angles and object displacement and demonstrated its validity via an application on the EYEDIAP video dataset. They also performed an attention assessment of an individual to understand one's behavioral patterns in cognitive disorders [88].

Rodriguez-Marek E. et. al. proposed a two-dimensional empirical mode decomposition (2D-EMD) by utilizing an inpainting technique. The most natural use for 2D-EMD is in image processing. They introduced an algorithm for EMD using inpainting for three surfaces i.e., upper, lower and mean surface. Upper and lower surfaces are related to local maxima and minima, respectively whereas the mean surface is calculated through the average of the upper and lower surfaces. Intrinsic mode functions (IMFs) are obtained by using these three surfaces, through an iterative process, the lower IMFs obtained with 2D-EMD show image edges [89].

Bhardwaj S. et al. proposed an approach for the prediction of chaotic time series through the combination of the Hidden Markov Model (HMM) and Neural Network (NN). They proposed an algorithm based upon HMM for Pattern identification and NN associated with each cluster for prediction of the system's output. Simulation results exhibit that the performance is much better as compared to the existing methods [90].

Singh D. et al. conducted a study for the identification of nonlinear variations of heart activities during operative attention through multifractal analysis. They observed that the heart rhythm shows the existence of a long-range correlation and Gaussian spectrum shape. Additionally, Fractal variations of heart dynamics were observed to quantify internal and external operative attention for baseline and recovery sessions, it may be helpful for developing brain-computer interfaces [91]. Farina, D. et al. worked on the transfer matrix describing the relation between distributions of transmembrane voltage (TMV) within the cardiac tissue and patient's body surface potential maps (BSPMs). It can be used to reconstruct cardiac activity non-invasively, which can be an important diagnostic tool in clinical practice. Moreover, they reduced the computation time by 50% with no loss in accuracy [92].

De Palma, P. et al. explored the application of modern computer hardware and software to the collection and analysis of behavioral data which involves work on child-directed speech [93]. Bhatt U. et al. paper addressed the issue of generating a high-resolution (HR) image from a single low-quality or low-resolution (LR) image. They worked on Discrete wavelet transform (DWT) with the Stationary wavelet transform (SWT) to generate or increase the resolution of the image. They considered Canny edge extraction to get a sharper image and comparatively better results based on image quality parameters in the super-resolution (SR) process [94]. Gupta J. et al. investigated human posture analysis for various implications in the field of sports and medical science including patient monitoring, lifestyle analysis, elderly care, etc. In their findings, most of the work is based on computer vision techniques. Nowadays, real-time solutions based on the Internet of Things are being planned and used for human posture recognition and detection [95]. Malhotra S. et al. developed a model for automatic age-invariant face recognition model using facial features. They extracted the features from the eye, mouth, and nose in the face image and calculated the eigenvalue and eigenvectors of a covariance matrix. Then evaluated the difference between all features to get a difference vector. Their experimental results show that this method outperforms the usage of the face aging data sets however, a fusion of age-invariant face recognition model further improves the accuracy [96]. Fairhurst M. C., et al. identified characteristics of the signature that affect the resilience

to fraudulent penetration for both automated and non-automated identity authentication [97]. A framework for EMG pattern identification using a neural network is proposed by Pancholi S., et al. They used TensorFlow to build the learning models and improved the classification accuracy and time complexity [98].

Singh D. et al. proposed an architecture of sensor fusion framework for environment perception. They used a fuzzy logic algorithm for autonomous mobile robot mapping by taking information from various sensors e.g., ultrasonic, and infrared sensors are fused to obtain better perception. The qualitative comparisons show the improvement in the results where the resultant map is extremely near to the reference map [99]. Singla S.K. et al. proposed a biometric methodology for the speaker verification system through an individual's voice characteristics using LabVIEW. They used a Mel-scale filter bank for feature extraction, and for hypothesis testing, they used a minimum distance classifier. They attained an accuracy of nearly 94% [100]. Fairhurst M. C., et al. studied Biometric-based access control to protect personal information or personal data. They proposed a machine-based perceptual mechanism for handwritten signature analysis to authenticate decisions that are conventionally made by visual inspection [101]. Joshi D. et al. worked on speaker identification through image processing techniques. They proposed a novel approach for feature extraction through Radon transform (RT) after obtaining the spectrogram from the speech sample. Also, PCA was used for feature reduction and to reduce computational complexity [102].

Koh M.-S. et al. emphasized the facilitation of Graphical user interfaces (GUIs) for engineering work using software packages. Simulation of several engineering models seems simpler when changing a variable in the GUI rather than coding to run a program. Typically, GUIs have been developed by using programming languages such as C, C++, C#, or JAVA. On the other hand, MatLab is also preferred for developing interactive GUIs. Nowadays MatLab is not only used for mathematical modeling but also be used for the designing and simulation of various dynamic systems [103]. Singh S. et al. investigated speech enhancement techniques through 12 different practical noises on five different language databases. were used. They used MatLab for algorithm implementation [104].

De Palma Paul investigated the work in automatic speech recognition. He emphasized that advancement in mobile computing devices is shifting the consumer paradigm away from desktop computing, and perhaps away from the point-and-click. Often times a computer user's eyes or hands are not available, in such situations the point-and-click model has clear limitations [105]. Koh M.-S. et al. presented an algorithm for speech enhancement. Their

algorithm is applicable to color noises, e.g., babble noises, without applying the whitening process. The results of the experiment exhibit the effectiveness of the algorithm to combat color noises [106].

An investigation was done by Sharma A. et al. on 10 healthy subjects which included both males and females of the 20-30 years age group. They assessed the impact of graded HUT and HRT on the sympathetic and parasympathetic nervous systems. Experimental study Results show that graded HUT leads to an increase in sympathetic activity but a decrease in parasympathetic activity, which is observed through higher tilt angles [107]. Gope J.C. et al. proposed a novel technique to analyze sentiments through opinion mining using machine learning. Hence, they categorized the positive and negative sentiments expressed in a piece of text. We acquired an accuracy of 91.90% with the Random Forest classifier and 97.52% of accuracy with the deep learning approach [108].

Wijesinghe A. developed an automated system for speech analysis to detect changes in communication patterns and differentiate autistic traits. They obtained a dataset from 40 participants and considered a three-stage processing system; the first stage is for thresholding for vocal activity and silence detection. The second stage is based on adopting machine learning techniques to develop a classifier for the isolated vocals and the third stage is for recognizing autistic traits in the speech patterns [109]. Xu Z. et al. performed an experimental study on Federated Learning (FL) which is a variant of ML to protect the data and sensitive information over the Internet of Medical Things (IoMT) devices to ensure privacy [110]. Gupta N. et al. proposed a technique for smart supervising of off-road vehicles through IoT sensor technologies. They presented a hybridized computational intelligence methodology for health monitoring and diagnosis (HM&D) of off-road vehicles, considering microphones as sensors, which are economical [111]. Ding F. et al. proposed a methodology to prevent biometric data, specifically facial information. They proposed an innovative approach for countermeasure against the attacks of DeepFake antforensics. It serves as a defense tool against anti-forensics for shielding facial bio information amid Industry 4.0 [112]. As fatigue detection technologies continue to advance, they have the potential to improve safety and performance.

This literature review is a critical and integral component of our research endeavor. This literature review provided the necessary background and context for our research. By reviewing existing literature, we have identified gaps that helped in building a framework for our study. It provided insights into various research methodologies employed in previous studies to analyze the strengths and limitations of various methods and interpretations of previous studies.

Also, it assisted in determining what aspects of the topic have already been explored and what areas remain uncharted. The literature review supports the significance of the research by demonstrating research gaps. Moreover, it helped in defining the research scope and boundaries of the research.

2.5 Research Gaps

Research is a continuous process, evolving novel methods and concepts for the betterment of humanity. Researchers have made numerous efforts in sleep-deprived fatigue detection, but there is still a need to do a lot for refinement.

Identifying and addressing research gaps is a crucial aspect of developing a robust and meaningful research proposal. Research gaps highlight areas where current knowledge is incomplete or insufficient. Filling these gaps contributed to the expansion and advancement of existing knowledge. Identifying research gaps helps justify the need for our study and demonstrates that research addresses an important issue of detection of fatigue due to sleep deprivation; that has not been adequately explored. Also, these research gaps provided an opportunity to introduce originality and innovation into our work.

Addressing gaps in existing literature helped in formulating research questions and guided the direction of our study. A thorough understanding of existing literature prevented duplication of efforts. Knowing what has already been studied allowed us to tailor our research to areas that have not yet been explored and refine our approach. Furthermore, these research gaps served as the foundation for a purposeful research proposal within the broader scholarly conversation.

Currently, there is no single, definitive test or device that can objectively determine the level of sleep deprivation or fatigue. While certain technologies like actigraphy or polysomnography can provide some insights, they may not be readily accessible or practical for everyday use.

After reviewing the available literature, the following research gaps have been identified.

1. Most of the present systems for detecting sleep-deprived induced fatigue are intrusive, i.e., various biomedical sensors have been employed on the body of ambulatory subjects. The major drawback of intrusive methods is that the subject under the test is aware that he or she is being observed, so the test result may be biased or distorted. Moreover, the subject's movement is also minimal during the measurement. Therefore, there is a need to develop a non-intrusive method for better fatigue assessment.

2. As per available literature, features from only one domain are implemented, i.e., uni-modal systems have been used. However, few researchers so far have implemented bimodal systems for this purpose. It may be possible that the outcome will be significantly more accurate by using a multimodal approach, hence increasing the number of features/parameters for observation. Moreover, cost and accessibility are other factors, implementing fatigue detection systems, such as advanced technology or biofeedback devices, can be costly and may not be feasible for all settings or populations.
3. Another research gap is that the quantitative value for sleep-deprived induced fatigue has not been used. A quantitative value to measure the level of fatigue due to sleep deprivation could be more helpful. This will enable the scientists to grade the fatigue numerically. Identifying fatigue induced due to sleep deprivation may require consideration of the broader context.

In our proposed methodology, we have attempted to fill the first and second research gaps by developing a non-intrusive and multimodal framework to detect sleep-deprived induced fatigue. However, the third research gap is considered as a future scope due to a wider research prospect and time constraints.

Moreover, quantification of fatigue is also a significant area of research. There is no standard or marker to detect sleep-deprived fatigue. Detecting fatigue is subjective and multifactorial because individuals may perceive and respond to fatigue differently. Proposed multimodal feature fusion is subject-specific which can be generalized. AI techniques can be implemented to eliminate the requirement of a controlled environment and to overcome the limitations.

Chapter 3

Multimodal Feature Fusion for Fatigue Detection

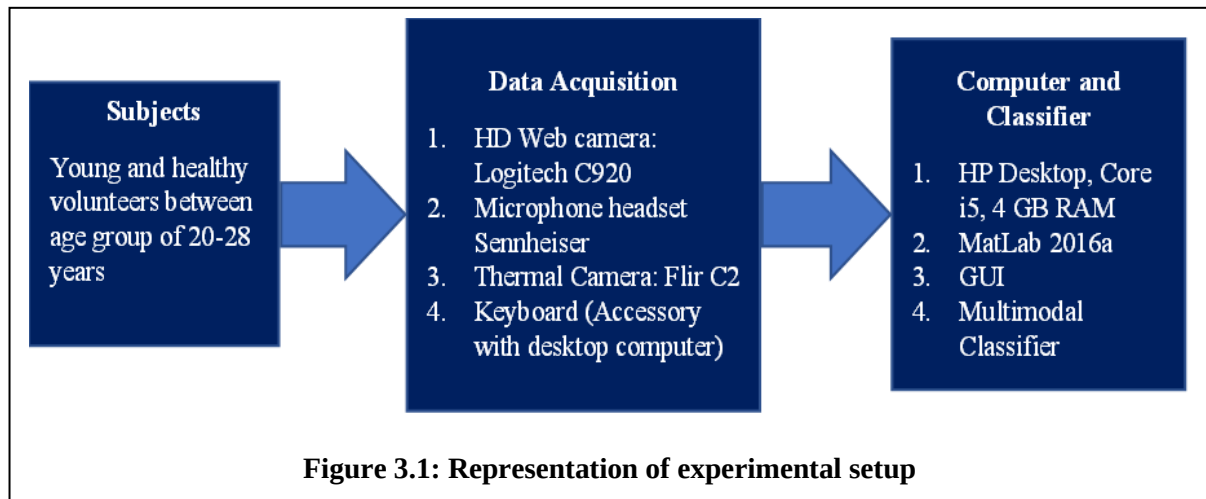
3.1 Introduction

A person who is feeling fatigued due to sleep deprivation is likely to make errors that may even prove to be fatal. Therefore, it is essential to recognize the fatigue. Multimodal feature fusion is significant for fatigue detection due to its ability to leverage information from various sources or modalities, enhancing the overall accuracy, reliability, and robustness of fatigue detection systems. Fatigue is a complex and multifaceted phenomenon that can manifest in various ways, both behaviorally and physiologically. Multimodal feature fusion allows for the integration of information from diverse sources, such as facial expressions, eye movements, thermal images, and cognitive performance (reflex actions), providing a more comprehensive view of the individual's fatigue state. Multimodal feature fusion enhances the applicability of fatigue detection systems in real-world, complex scenarios such as driving, aviation, healthcare, and industrial operations. The combination of various modalities makes the system more robust and suitable for deployment in practical situations.

The novelty of the proposed methodology to detect fatigue is that it is based on multimodal feature fusion and is non-intrusive in nature. In the proposed research work fatigue is detected by acquiring significant features from four domains, i.e., visual images, thermal images, keystroke dynamics, and voice features. Young healthy volunteers (n=60) of the 20 to 30 years age group participated in the experimental study. Also, during the study participants abstained from the consumption of tea, coffee, alcohol, and other drugs that could affect their sleep patterns. During the experimental study samples of participants are acquired from all four domains for feature extraction, and empirical weights are assigned to the respective domains. The obtained results are compared with the five classifiers; k- Nearest Neighbours (kNN), Multilayer Perceptron, Random Forest, Random Tree, and Support Vector Machines (SVM) classifiers. The proposed multimodal nonintrusive technique has attained an average detection accuracy of 93.33% in 3-fold cross-validation.

3.2 Material

The multimodal fatigue detection system's apparatus consists of a desktop computer of Make: HP, Model: R-15 (including wireless keyboard and mouse) used as an interface for signal processing and computation task. A high definition (HD) webcam of Make: Logitech, Model: C920, captures visual spectrum images. Flir's C2 thermal camera captures thermal images, as shown in Figure 3.1. A software tool, MatLab version: 2016a, incorporated with the image processing toolbox, is used for multimodal information processing and categorization.



3.3 Subjects

We considered individuals for voluntary participation as subjects who possess certain qualities that will contribute to the success of the study and ensure the validity of the results. Volunteers provided informed consent to understand the purpose of the study, its procedures, potential risks, and benefits, and willingly agreed to participate without coercion.

Volunteers were committed to the study requirements, following protocols, and completing necessary tasks. All volunteers were in good physical and mental health to minimize confounding factors that could influence the study's results. We confirmed the availability and accessibility as per project requirements. Also, we confirmed the participants should not have any pre-existing conditions, medications, or lifestyle factors that could significantly impact the study's outcomes. Participants were aware of their right to withdraw from the study at any time without consequences.

A group of 120 healthy subjects was constituted between 20 years to 30 years. They also abstained from caffeine, alcohol consumption, psychoactive and psychotropic drugs, influencing sleep patterns. All subjects were well informed about the study's purpose, and we

instructed them to fill out a subjective sleepiness scoring questionnaire. Subjects were kept awake for 18 hours and called for data acquisition in a controlled environment. When they had taken sound sleep for more than 8 hours, the same subjects called again for the baseline (Alert state) data acquisition. For consistent circadian rhythm, data was acquired in the morning between 7:00 AM to 9:00 AM. Data acquisition comprises four domains, 1. Visual facial image capturing. 2. Thermal image capturing. 3. Voice acquisition. 4. Keystroke data acquisition. Data acquired from the mentioned four domains are required for cumulative dataset preparation. This dataset is used for developing, testing, and validating the classifier. The entire procedural process is discussed in section 3.4.

3.4 Methodology

A multi-modal feature fusion methodology involves combining information from different modalities (such as text, images, audio, etc.) to enhance the overall representation and performance of a system, typically in the context of machine learning or artificial intelligence. The goal is to leverage the complementary information provided by different modalities to improve the accuracy and robustness of a model. Attention mechanisms allow the model to focus on specific parts of the input data, giving more weight to relevant modalities.

Multi-modal attention mechanisms can dynamically adjust the contribution of different modalities based on the context of the input.

This experimental study aims to identify the significant physiological parameters from different domains for fatigue detection and develop a computer-aided system using multimodal features. Due to sleep deprivation, there will be a considerable change in the various physiological parameters.

The significant parameters for detecting human fatigue are eye cues, facial expressions, body reflexes, voice patterns, and the facial region's thermal spectrum. As shown in Figure 3.2, an experimental apparatus has been prepared to detect these physiological parameters and behavioral changes. Fatigue induced due to sleep deprivation can be distinguished through facial expressions, gestures, speech patterns, and physiological indicators [113]. Tang L. et al. proposed an image fusion framework for infrared and visible images in real time [114].

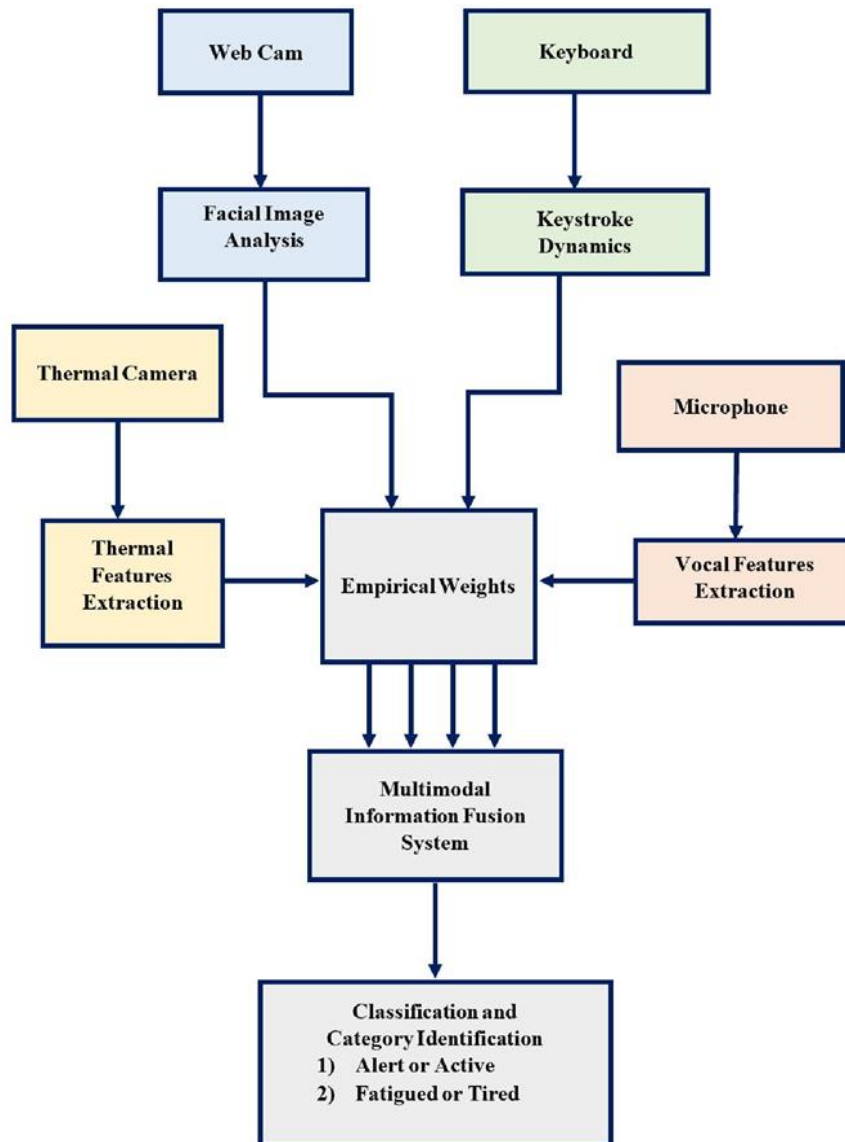


Figure 3.2: Exhibits methodology for fatigue detection using multimodal fusion

The overall procedure is carried out in five stages, as depicted in Figure 3.2. This apparatus designed by us is a multimodal feature fusion from 4 domains: 1. Visual spectra facial image analysis, 2. Keystroke analysis, 3. Voice analysis, and 4. Thermal spectra image analysis. The whole methodology consists of five stages: 1. Data acquisition, 2. Feature extraction, 3. Multimodal feature fusion, 4. Processing and analysis, and 5. Classification and detection. A graphical user interface has been developed for interface purposes and to facilitate the whole process as represented in Figure 3.3.

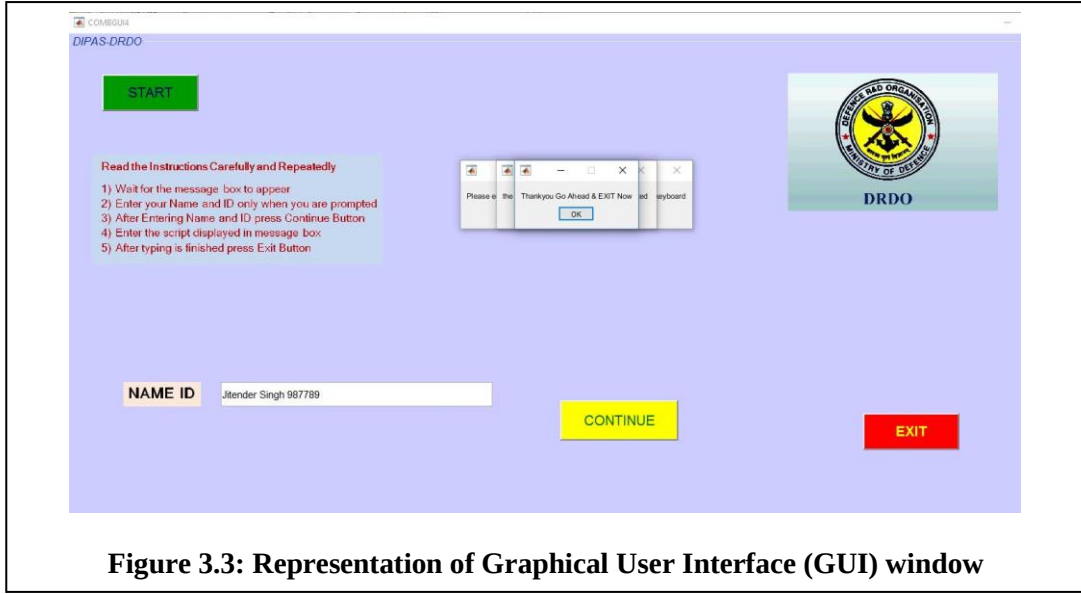


Figure 3.3: Representation of Graphical User Interface (GUI) window

3.4.1 Data Acquisition

The foremost step is to acquire data in a noise-free and constrained environment. A protocol has been made to keep the dataset uniform and get reliable results, and the apparatus has been installed in an isolated chamber with diffused lights for consistency in ambient light. A direct light source was not there in the chamber, reflecting the subject's face. Also, we maintained consistency in ambient temperature, which was within the range of 22°C to 28°C. External sound interference was negligible. The subject's seating distance was 50 to 60 centimeters away from the apparatus.

3.4.2 Feature Extraction

Significant features have been obtained from the four domains to detect fatigue, i.e., a. Visual image features, b. Thermal image features, c. Keystroke dynamics, and d. Vocal features.

a) Visual spectra Facial image features: We obtain discrete pictures from the webcam, and the facial region is confined automatically from the discrete images using the Viola-Jones algorithm. Afterward, the Viola-Jones algorithm detects the eye-pair region automatically in the particular facial region [73]. The eye-pair region is cropped and converted into a grayscale image for processing time minimization. Here, cropping and converting RGB images to grayscale images have improved overall efficiency. In the subsequent step, histogram equalization is performed, followed by morphological operations.

$$S_N = T(r_N) = \sum_{j=0}^N P_r(r_j) = \sum_{j=0}^N \left(\frac{n_j}{n} \right) \quad (3.1)$$

Here in equation 3.1 [115], S_N represents the intensity value of the resulting image, corresponding to the intensity amount of an input image r_N ; N represents total grayscale levels and P represents probability. Morphological operations incorporate four steps, the first step is the disk structure element per retina's shape. The second and third stages comprise the opening of the image function, trailed by the closing of the image closing function to form a filter. The filtered image is processed through a dilation operation to get notable edges from the dilated image obtained in the fourth step.

After morphological operations, the image is converted into binary, later utilized for the decision-making process. We used the image-profile function on a binary image segment to differentiate between closed and open eye pairs. The image profile function provided us with notable information through the pixel value plot of a line. We decide a threshold limit based on pixel values. Hence, a binary classifier is formed to classify open and closed eyes.

Open and closed eyes are differentiated by a binary algorithm developed for classification. Eight perpendicular lines are formed when an eye pair is open across the image profile line. These perpendicular lines characterize the number of white pixels across the lines. The opened eyes, nose, and pupil segments make more than eight perpendicular lines. Whereas in closed eyes, only nose boundaries are observed, resulting in only two perpendicular lines in a function of the image profile. This binary algorithm can automatically distinguish open or closed eye pairs across the perpendicular line count.

Afterward, significant features have been extracted from the images, followed by feature reduction. Feature reduction is performed through Fisher's discrimination ratio (FDR). Elevated FDR denotes the feature's capability to discriminate between two classes is higher [116]. The FDR value for the n th feature is calculated as described in equation 3.2.

$$FDR_n = \left| \frac{\mu_{an} - \mu_{bn}}{\sqrt{(\sigma_{an}^2 + \sigma_{bn}^2)}} \right| \forall n \in [1, 10] \quad (3.2)$$

Here, μ_a = mean of class 'a'; μ_b = mean of class 'b'; σ_a = standard deviation of class 'a' and σ_b = standard deviation of class 'b'. Based on the highest FDR and minimum correlation coefficient, the Six most significant features have been identified out of 10 facial image features, as shown in Table 3.1.

Table 3.1: List of features from visual spectra facial image

S. No.	Features	Description
1	Eyes Close Period (ECP)	The time period for closed eyes in a 3-minute video sample
2	Open to Close Changeover Period (OCOP)	The time for the eye state to change from an open state to the closed state
3	Close to Open Changeover Period (COCP)	The time for the eye state to change from the closed state to an open state
4	Total Changeover Time (TCT)	Time period for total changeover states (OCOP+COCP)
5	Inter Changeover Frame Count (ICFC)	Changeover frames count
6	Eyes Blink Count (BKC)	Total eye blinks number in a 3-minute video sample

b) Thermal image features: Traditional methods to assess fatigue include self-reporting and physiological measurements like heart rate and cortisol levels. However, advancements in technology have led to the emergence of innovative techniques, such as using thermal imaging, to detect fatigue non-invasively. Thermal imaging is a non-contact technique that captures and measures the infrared radiation emitted by the subject. By detecting these thermal patterns, thermal imaging cameras create visual representations of the temperature variations in the human body. Overall, five features (listed in Table 3.2) are considered significant from the captured thermal images. The facial region is initially cropped and extracted from the image acquired from the thermal camera to extract the pixel count value. Then, the extracted thermal image of the facial region is converted into grayscale.

Table 3.2: List of features from the thermal image

S. No.	Features	Description
1	Pixel count	Total number of white pixels in binary image
2	FHFa	Angular Sum of Forehead region
3	FHFr	Radial sum of forehead region
4	PRFa	Angular sum of periorbital region
5	PRFr	Radial sum of periorbital region

Cropping the image and converting it to grayscale reduces the processing time significantly. Afterward, the grayscale image is transformed into a binary image, as shown in Figure 3.4. To extract the other four thermal features (two from the periorbital region and two from the forehead region), the ROIs from the thermal image have been selected.

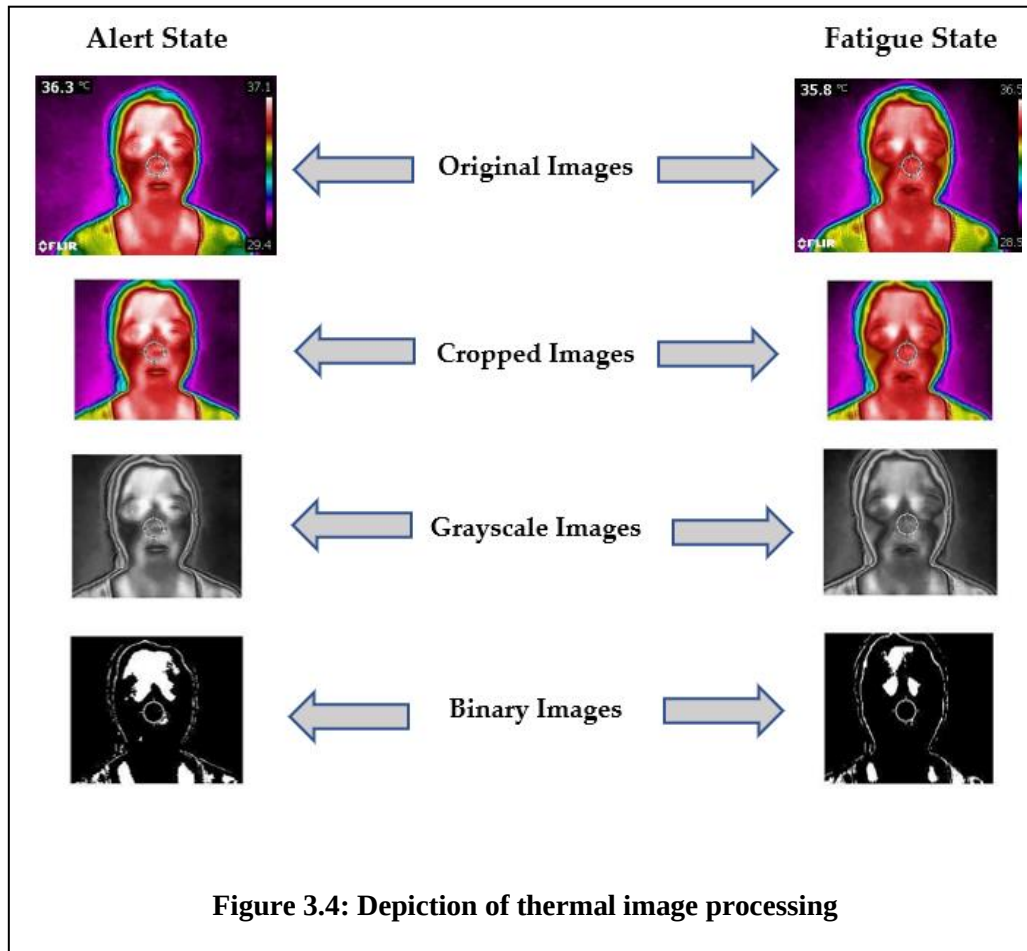


Figure 3.4: Depiction of thermal image processing

During experiments, it has been observed that the periorbital region and forehead region are significant ROIs. Thermal and visual spectrum images are captured by the same camera, i.e., Flir C2. The visible spectra image is resized to match the thermal image size, i.e., 320 x 240 Pixels. Afterward, the eye pair region is extracted from the visual spectrum image using the Viola-Jones algorithm [73], and we get a bounded box over the eye pair. The coordinates of the bounded box to get the ROIs for the periorbital region from the thermal image. The exact size bounded box is put adjoining above this to obtain forehead features. After getting ROIs, the next step is to obtain FPS (Fourier Power Spectrum) features from selected ROIs. This texture model provides texture features related information, like contrast, grain size, and orientation. The Discrete Fourier Transform (DFT) approach has been used to quantify the

texture features. Afterward, the Angular sum and the Radial sum of the DFT have been calculated to obtain the texture features. FPS features are computed through the frequency domain and power spectrum.

$$|F(u, v)|^2 = F(u, v)F^*(u, v) \quad (3.3)$$

Here, the DFT of an image is $F(u, v)$, and $F^*(u, v)$ is the complex conjugate of the DFT of an image. To obtain function $S_r(\theta)$ spectral features are represented as polar coordinates. Also, frequency is defined as $S_r(\theta)$, and direction is expressed as $S_\theta(r)$. Wedge analysis is performed by evaluating $S_\theta(r)$ for a fixed value of θ . It is the spectrum behavior along the radial direction from an origin. Whereas Ring analysis is performed by evaluating $S_r(\theta)$ for a fixed value of r . It is the behavior of the entire range beside a concentric circle on the origin. The summation of these discrete values gives the global interpretation:

$$S_\theta = \sum_{\theta=0}^{\pi} S_\theta(r) \quad (3.4)$$

$$S_r = \sum_{\theta=0}^{R_0} S_r(\theta) \quad (3.5)$$

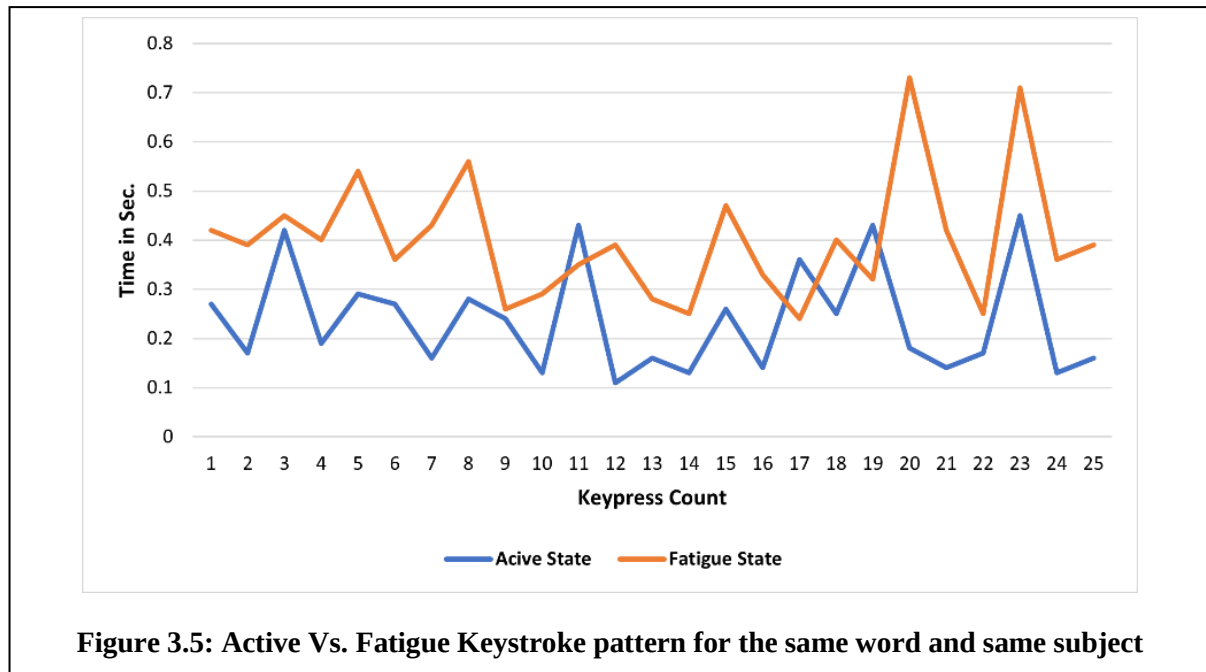
Where R_0 is the radius of the concentric circle on the source. In the current texture model, two features: S_r and S_θ are computed by determining the texture's orientation.

c) Keystroke dynamics: It is a behavioral biometric; used to observe the keyboard actions when the subject types a sentence through a keyboard. Through experiments, the variation in the keystroke attributes for fatigued subjects can be distinguished from the alert subjects. Keystroke dynamics measures the user's typing pattern when an individual subject types on the keyboard. The features from keystroke dynamics are listed in Table 3.3.

Table 3.3: List of features from Keystroke Dynamics

S. No.	Features	Description
1	KSR	Keystroke Rate
2	CRE	Character Error
3	AVD	ASCII Value Difference
4	TST	Total String Time

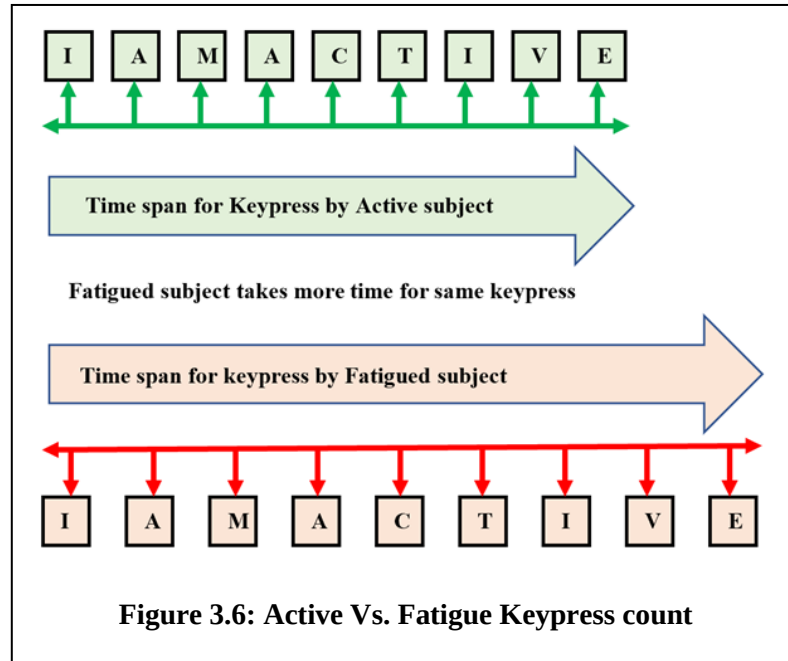
A total of 4 features have been identified to measure keystroke dynamics. The first feature is the Keystroke rate. It is the total number of keypresses by an individual to complete one string, as shown in Figure 3.5.



Different sentences of similar character length have been used in the experiment, i.e., around 48 to 54 characters, considering space keypress. The purpose of distinct sentences is that the individual does not become habitual to typing one sentence every time.

The second feature is Character Error (CRE) which is the difference in characters; this feature is measured by calculating the total number of mismatches between the string displayed on the screen and the string entered by the subject using a keyboard. Each character is compared one by one with the preliminary assumption that the subject is prone to make mistakes if the subject is exhausted. In contrast, the errors attempted by an active subject will be negligible.

The third feature is the ASCII Value Difference (AVD); here, the preliminary assumption is that sometimes the subject is alert, but change may occur in characters' place while typing a sentence in a hurry i.e. prone to typo-error. There will be no difference in ASCII values of that particular sentence in such cases. The fourth feature is the total time taken by an individual to enter a sentence or string, i.e., whole string time, as shown in Figure 3.6.



Let us take a string; "With the new day comes new strength and new thoughts" as an example. This particular string consists of 52 characters, including the space keypress. If the subject is fatigued, he will take more time within the keypress. Also, there are chances he will make some spelling mistakes. Suppose the subject is alert and makes spelling mistakes, which can be detected by the characters' differences in the ASCII value. We hypothesize that if the subject is alert, there will be no ASCII difference. Also, the fatigued subject takes more time to complete all the keypresses than the alert subjects.

d) Vocal features: 5 significant features are extracted from the subjects' voice samples to detect fatigue, as listed in Table 3.4. The first feature is the Fundamental Frequency or pitch (FFP) of auditory sensation produced by the speech.

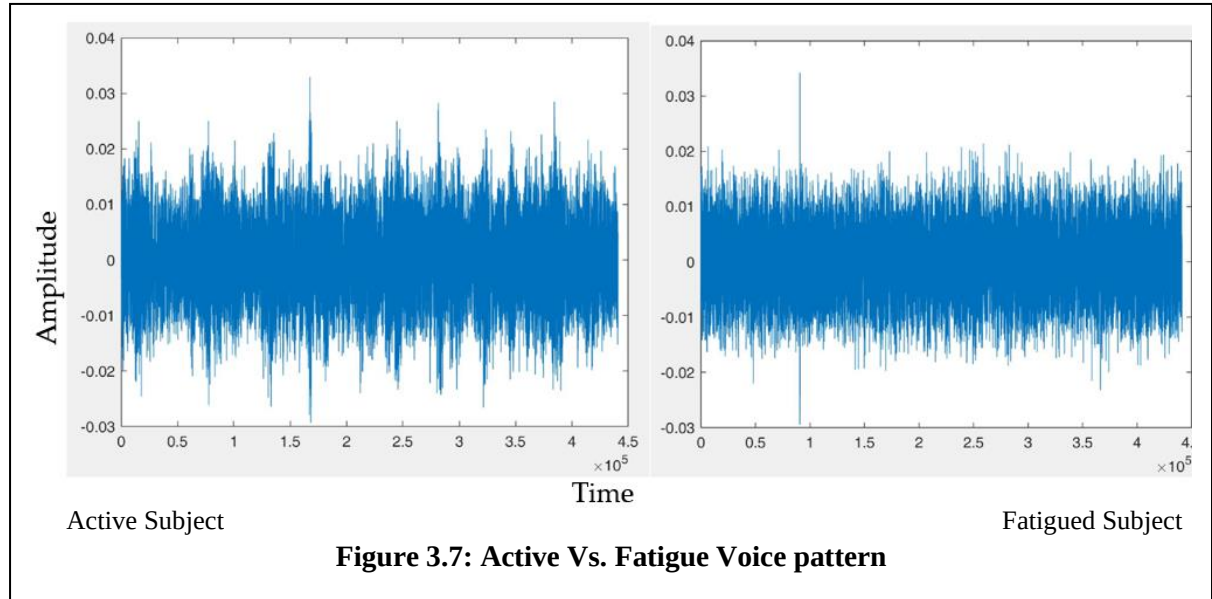
The second feature is the Rate of Speech (RSH), also known as voiced, and unvoiced duration represents temporal speech rhythm characteristics such as pause patterns. The third feature is the Sound Pressure Level (SPL), which is the difference between the pressure of the sound wave produced and the ambient pressure. The fourth feature is Power Spectral Density (PSD); it is a limited average power of a speech signal described by the average power spectral density. Power spectral density is the area under its Fourier transform magnitude. The average power of a signal $s(t)$ is P_{avg} ; computed as in equation 3.6 [117].

Table 3.4: List of vocal features

S. No.	Features	Description
1	FFP	Fundamental Frequency (Pitch)
2	RSH	Rate of Speech
3	SPL	Sound Pressure Level
4	PSD	Power Spectral Density
5	SPD	Speech Duration

$$P_{avg} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(t)^2 dt \quad (3.6)$$

Speech Duration (SPD) is the fifth feature, representing the time taken by a subject to complete a sentence fragment. A sample voice spectrum for 10 second audio of the Active subject and the Fatigued subject is shown in Figure 3.7.

**Figure 3.7: Active Vs. Fatigue Voice pattern**

3.4.3. Multimodal feature fusion

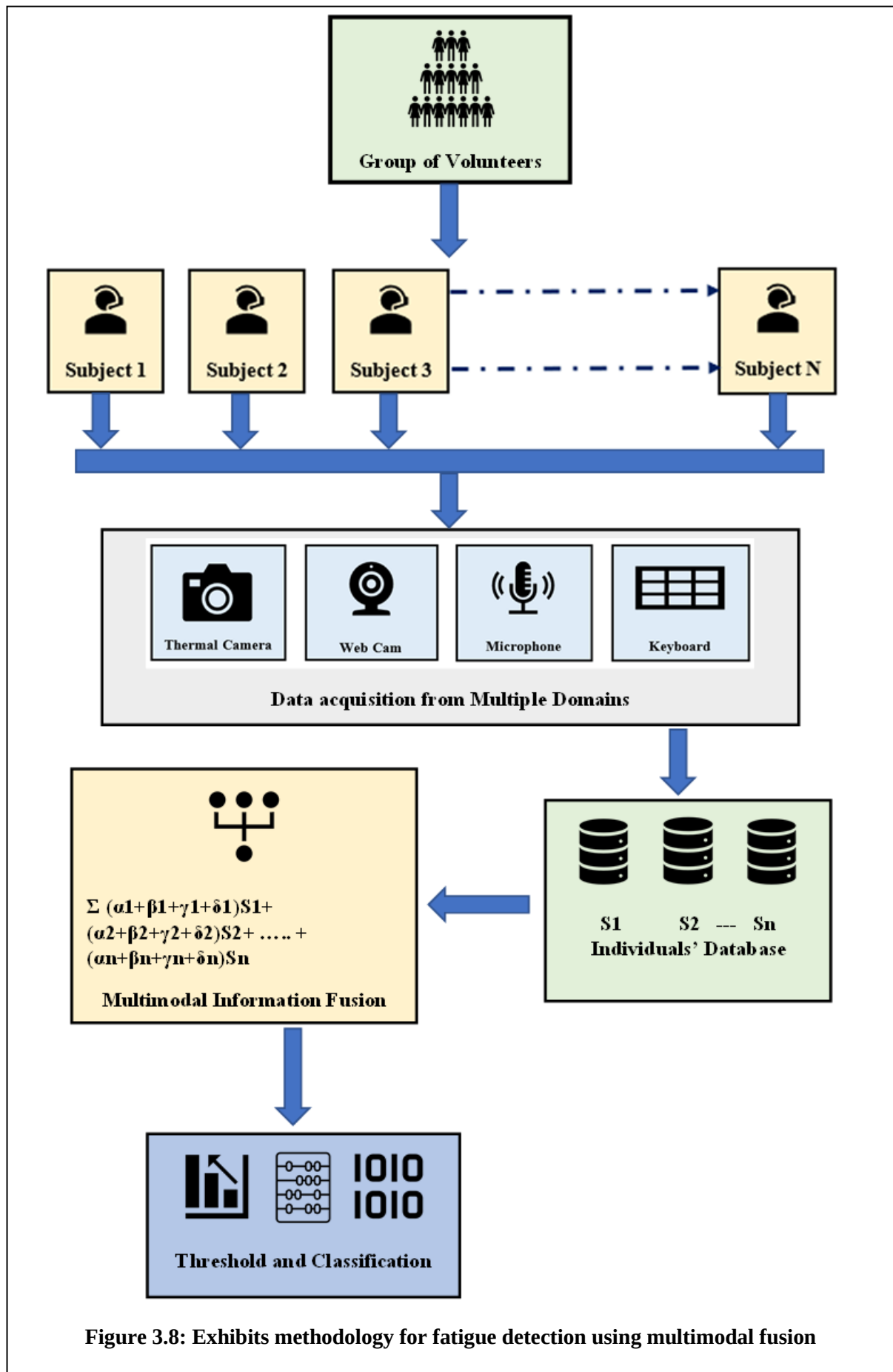
Feature fusion is very effective when more than two domains are involved. In this study, four different domains have been studied. The need for multimodal feature fusion arises to obtain the optimum results. During the experiments, it has been observed that accuracy could be enhanced by combining the features from multiple domains, i.e., feature fusion. So, ten combinations have been evaluated to improve the results, e.g., by combining features from 2 domains and then from 3 domains and assigning different weights to each domain. The results obtained from feature fusion with various combinations are shown in Table 3.5.

Table 3.5: Detection accuracy matrix from the individual domains

S. No.	Domains	Equal Empirical weights	Accuracy	Optimized Empirical Weights	Optimized Accuracy
1	Visual + Thermal	0.5+0.5	75%	0.4+0.6	77.5%
2	Visual + Keystroke	0.5+0.5	70%	0.65+0.35	75%
3	Visual + Voice	0.5+0.5	70%	0.62+0.38	72.5%
4	Thermal + Keystroke	0.5+0.5	72.5%	0.7+0.3	75%
5	Thermal + Voice	0.5+0.5	72.5%	0.55+0.45	77.5%
6	Keystroke + Voice	0.5+0.5	70%	0.31+0.69	72.5%
7	Visual + Thermal+ Keystroke	0.34+0.33+0.33	75%	0.36+0.38+0.26	77.5%
8	Visual + Thermal + Voice	0.34+0.33+0.33	82.5%	0.34+0.36+0.30	87.5%
9	Thermal+ Keystroke + Voice	0.34+0.33+0.33	80%	0.35+0.25+0.30	85%
10	Visual + Thermal + Keystroke + Voice	0.25+0.25+0.25+0.25	82.5%	0.29+0.37+0.16+0.18	92.5%

In this multimodal feature fusion approach, the critical part is predictive modeling through inductive reasoning that involves past evidence to determine the outcome and keeps on adding the dataset after every trial. This is achieved by developing adaptive models trained on past data and making predictions on new data. We took 120 samples (60 Alert + 60 Fatigued) from two classes, i.e., Alert and Fatigued. Initially, we trained our model with 80 samples (40 Alert and 40 Fatigued), tested on the remaining 40 samples (20 Alert + 20 fatigued), and performed cross-validation by using five different classifiers for a baseline of performance. It is explicitly stated that the testing in each fold is done on the data which is not exposed to classifier during its training.

Furthermore, we decided to combine the features from all four domains to improve detection accuracy. In the beginning, features from the four domains have been combined, each with 25 % weight. After some iterations, we implemented the optimization technique to assign the empirical weights (α , β , γ , and δ) to the respective domain with a goal to achieve maximum classification accuracy.



After acquiring features from all domains, features are fused through an adaptive framework for autonomous weighting. The proposed feature fusion system allows the feature-level fusion of signals from multiple domains. Weights are assigned automatically after optimization through the MatLab optimization toolbox using the Multiobjective Optimization Goal Attainment Method. Here, $S_1 + S_2 + \dots + S_n$ represents n number of subjects with weights respective to each domain, i.e., α for Visual spectrum images β for Thermal Images, γ for Keystrokes, and δ for Vocal features. The cumulative result is obtained as a summation of all the empirical weights of respective subjects as explained in equation 3.7.

$$\Sigma (\alpha_1 + \beta_1 + \gamma_1 + \delta_1)S_1 + (\alpha_2 + \beta_2 + \gamma_2 + \delta_2)S_2 + \dots + (\alpha_n + \beta_n + \gamma_n + \delta_n)S_n \quad (3.7)$$

Finally, the features from all four domains are fused to get the optimized results, as shown in Figure 3.8.

3.4.4. Classification

Empirical weights have been assigned to the features to improve the classifier's accuracy. Initially, we assigned 25% weight to all the domains. Afterward, we applied the optimization technique to assign the weights to achieve maximum classification accuracy. We assigned 29% for visual spectrum image analysis, 37% for thermal spectrum image analysis, 16 % for keystroke analysis, and 18% for speech analysis. For this proposed multimodal feature fusion technique, a threshold value is taken as 0.5. It implies that if the output of the proposed multimodal feature fusion technique is less than 0.5, then the subject is alert; otherwise, the subject is classified as fatigued.

Table 3.6: Multimodal Feature Fusion classification accuracy

S. No.	Name of classifier	Accuracy %	Correctly Classified Instances	Incorrectly Classified Instances
1	kNN	85	34	6
2	Random Tree	87.5	35	5
3	Random Forest	87.5	35	5
4	SVM	90	36	4
5	Multilayer Perceptron	90	36	4
6	Proposed Method	92.5	37	3

To optimize the classification through multimodal feature fusion, 80 samples were taken for training (40 fatigued + 40 alert). These were executed through 5 different classifiers to assess the classification accuracy. The first classifier was k-Nearest Neighbors (kNN); it provides an accuracy of 85% with 34 correctly classified instances out of 40. The second classifier is Random Tree, in which accuracy of 87.5% has been achieved with 35 correctly classified instances out of 40. In comparison, 35 cases are correctly classified through Random Forest with an accuracy of 87.5%. Support Vector Machines (SVM) and Multilayer Perceptron gave an accuracy of 90% by detecting 36 correctly classified out of 40 as shown in table 3.6. The highest accuracy of 92.5% was obtained by the proposed method detecting 37 correctly classified subjects out of 40. This was subjected to 3-fold cross-validation giving an average accuracy of 93.33%.

3.5 Results and Discussion

If we discuss individual domain-wise fatigue detection in the first domain, i.e., Visual spectra image analysis, fatigue detection accuracy is 67.5%. This is given in Table 3.7. In the Thermal domain, an upsurge is observed in the detection accuracy. i.e., thermal images, with an accuracy of 75%.

Table 3.7: Detection accuracy matrix from the individual and multimodal domains

Domains	Subject	True	False	Accuracy
1. Visual spectra image	Positive	13	7	67.5%
	Negative	6	14	
2. Thermal spectra image	Positive	16	4	75%
	Negative	6	14	
3. Keystroke dynamics	Positive	14	6	67.5%
	Negative	7	13	
4. Vocal Analysis	Positive	14	6	70%
	Negative	6	14	
5. Proposed Multimodal feature fusion technique	Positive	18	2	92.5%
	Negative	1	19	

However, there is a decline in detection accuracy in the Keystroke domain, i.e., Keystroke analysis with an accuracy of 67.5%. Accuracy of 70% has been observed through the Vocal domain.

Finally, an accuracy of 92.5% is achieved through the multimodal feature fusion technique. This being the highest accuracy of all the techniques tested, is subjected to 3-fold cross-validation.

3.6 Conclusion and Summary

There will be a detrimental effect on the individual's physical performance and capabilities due to poor sleep. Fatigue induced due to sleep deprivation is a significant cause of accidents because fatigue affects awareness and worsens drivers' and operators' vigilance [118]. Also, fatigue assessment is challenging since the clinical factors for the quantification of drowsiness are not well defined. The proposed methodology for fatigue detection is based on feature fusion techniques.

The experiment demonstrates that the performance is optimized after feature fusion from all four domains. The empirical weights have been assigned during feature fusion according to the detection accuracy for classifier optimization.

Table 3.8: Confusion Matrix

Domain	Folds		True	False	Accuracy
Multimodal Feature Fusion	1	Positive	19	1	92.5%
		Negative	2	18	
	2	Positive	19	1	95%
		Negative	1	19	
	3	Positive	18	2	92.5%
		Negative	1	19	
	Average Accuracy				93.33%

The confusion matrix for the proposed multimodal feature fusion technique is shown in Table 3.8. Individually, the thermal domain has a maximum detection accuracy of 75%. The second highest is the vocal analysis, with a detection accuracy of 70%. The Visual spectra and Keystroke dynamics are on the third rank with a detection accuracy of 67.5%. To conclude, the proposed multimodal feature fusion technique exhibits the highest detection average accuracy of 93.33% in 3-fold cross-validation.

Fatigue can manifest differently in different situations or contexts. Multimodal feature fusion enables real-time adaptability by dynamically adjusting the weight or relevance of each modality based on the current environmental conditions or task demands.

Chapter 4

Detecting Fatigue Through Eye Cues Using Adaptive Learning

4.1 Introduction

The majority of people who get an inadequate amount of sleep are prone to mental fatigue due to sleep deprivation. This mental fatigue induced by sleep deprivation is extremely detrimental to persons involved in numerous critical jobs e.g., Air traffic controllers, Mariners, Pilots, Surgeons, and many more. Detecting fatigue due to sleep deprivation through facial expressions involves observing changes in the facial expressions and certain subtle eye cues that may indicate tiredness. Also, detecting fatigue due to sleep deprivation through human eyes involves observing specific signs and changes in eye appearance and behavior.

Eye cues are significant indicators to detect fatigue, e.g. in a fatigued state the eyes may appear partially closed, and the individual may struggle to keep them open. Sleep-deprived individuals may exhibit droopy or heavy eyelids that can give the impression of tiredness and fatigue. Also, sleepiness can lead to increased blinking as the body tries to keep the eyes moist and alert. Excessive blinking may be a sign of fatigue. This also may be reflected in the eye cues, with a lack of alertness or responsiveness. Moreover, lack of sleep can contribute to eye twitching or spasms, which can be a visible sign of fatigue. Furthermore, fatigue can result in slower eye movements and a reduced ability to track moving objects. This can be observed during activities that require visual attention. Fatigue may result in a reduction of facial expressiveness and thus in eye movements i.e., sleep-deprived individuals may display slower eye and eyelid movements. This can be observed in delayed reactions, sluggish responses, and a generally reduced pace eyelid movement.

Tiredness can lead to muscle tension in the face or strain around the eyes. It's important to consider that facial expressions and eye cues can vary among individuals, and some people may naturally exhibit certain expressions even when well-rested. Additionally, factors like stress, anxiety, and other health conditions can influence facial expressions.

The proposed research work is based upon an intelligent framework based on self-learning, followed by an adaptive threshold classification. Furthermore, the proposed methodology is non-intrusive in nature, in which the participant is unaware of being monitored during the experimental study; it prevents biasing of results. The novelty of the proposed methodology is the Inter-frame interval of closed eye and open eye acquiring features that leads to the recognizing Alert or Fatigued state through self-learning and adaptive threshold classification. The proposed self-learning adaptive framework is real-time and attained 97.5 % detection accuracy. Since the proposed methodology is self-learning, its accuracy and sensitivity are probable to increase further as the size of the training data set increases over a time period.

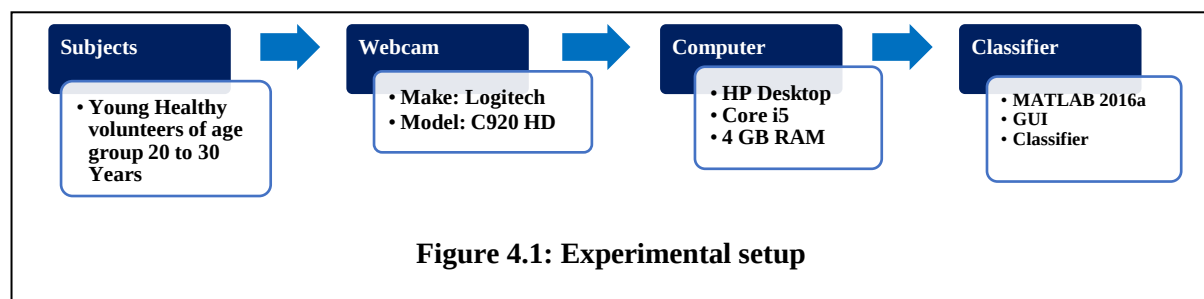
4.2 Novelty

Efforts have been made to make the research advantageous by incorporating the computer vision-based intelligent techniques listed hereafter:

- Eye state detection through image profile
- Adaptive threshold based upon self-learning algorithm.
- The subject remains unaware of being monitored to avoid any bias.
- Five new features have been proposed, which are helpful for the recognizing open-close rate of an eye.

4.3 Material

The experimental setup (shown in Figure 4.1) includes a webcam of make Logitech, Model- C920, used for video capturing purposes. A desktop computer unit of make HP, Model- R-15, is utilized for video processing and computation task. MATLAB, programming software version: 2016a; in-built video acquisition and image processing toolbox is used for coding and classification purposes.



4.4 Subjects

A group of 60 healthy volunteers was instructed to remain awake for at least 18 hours and come for the data acquisition. The same set of volunteers, when they had taken long sound sleep of more than 6 hours and were feeling fresh, called again for the data acquisition. Also, data were acquired in the morning between 8 to 11 AM so that data may not be affected by circadian rhythm. Data acquisition includes the video capturing process required for dataset preparation. Visual parameters from the facial region are then extracted from the dataset, which is used for developing, testing, and validating the classifier.

4.5 Dataset

Preparing datasets is a foundational step in any research. It helps us to avoid errors and inaccuracies during the analysis phase, ensuring more reliable and robust results. Dataset preparation involves ensuring consistency in the data, including standardizing formats, units, and variable names. Consistent data allows us for accurate comparisons and reliable insights, that significantly influence the validity and interpretability of research outcomes.

The image dataset has been obtained from video frames of 60 subjects. A total of 120 samples were acquired, 60 for the alert and 60 for the fatigued states. A two-minute live video clip of high resolution (1024 X 768) is obtained @ a frame rate of 30 fps (frames per second) through the webcam. The captured video stream is then split into frames to get corresponding RGB images, which gives us 3600 frames per sample. Further, open eye and closed eye images are categorized upon visual features for preparing two distinct datasets for open eye and closed eye, respectively. Various morphological operations are carried out to maintain uniformity and reduce the computational complexity for overall performance.

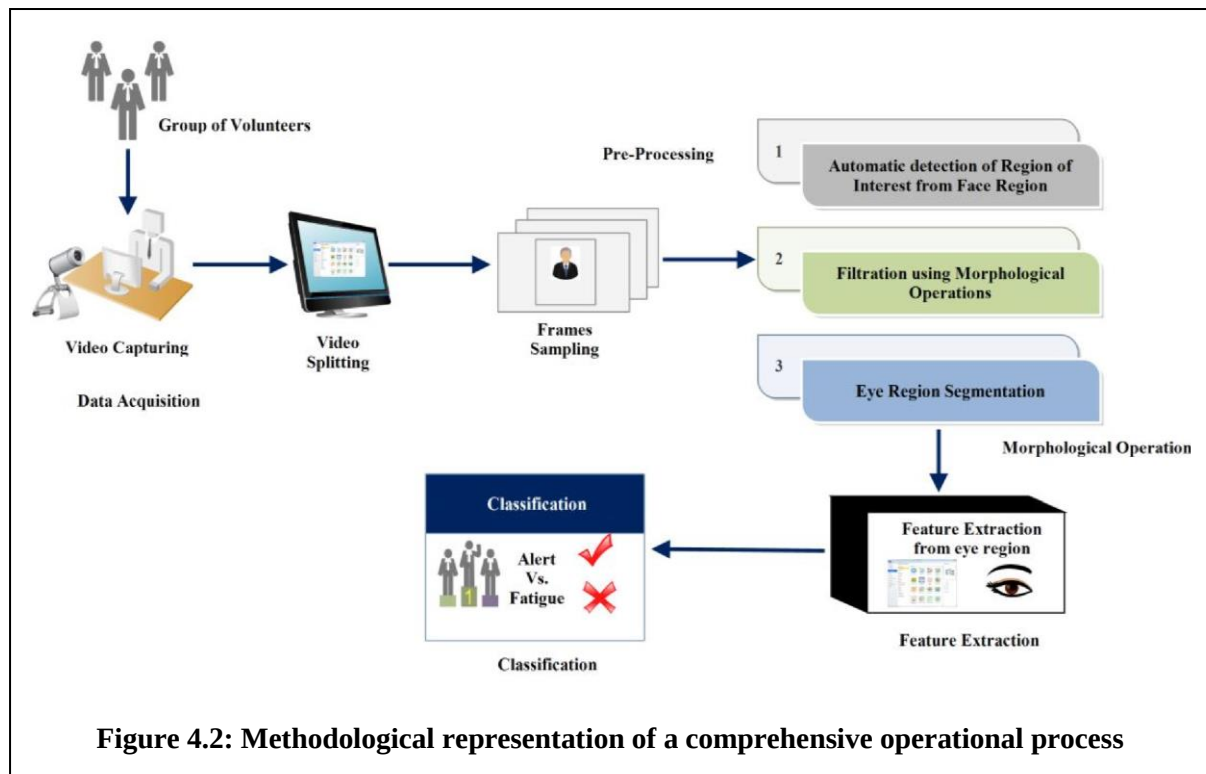
4.6 Methodology

A GUI (Graphical User Interface) is designed using MATLAB to facilitate the entire process, incorporating several sequential phases from the initial frame capturing to the final classification. Once the data is acquired, the methodology is sub-categorized into four sections: pre-processing, morphological operations, feature extraction, and classification. All four sub-categories associated with methodology are discretely described hereafter and depicted in Figure 4.2

4.6.1 Pre-processing

Image preprocessing is a critical step in proposed research work, because of its ability to enhance analysis accuracy, and facilitate more effective extraction of meaningful information. Preprocessing techniques, such as histogram equalization and adjusting the contrast of an image are crucial for improving visibility and highlighting important features, making it easier to distinguish details. Moreover, preprocessing ensures that images are in a consistent format and scale. Normalizing pixel values or standardizing color spaces allows for fair comparisons and ensures that the algorithms are not biased by variations in image characteristics.

In our approach, obtaining discrete frames from the captured video is the foremost step in extracting the ROI (Region of interest) from split frames. Afterward, the facial region is



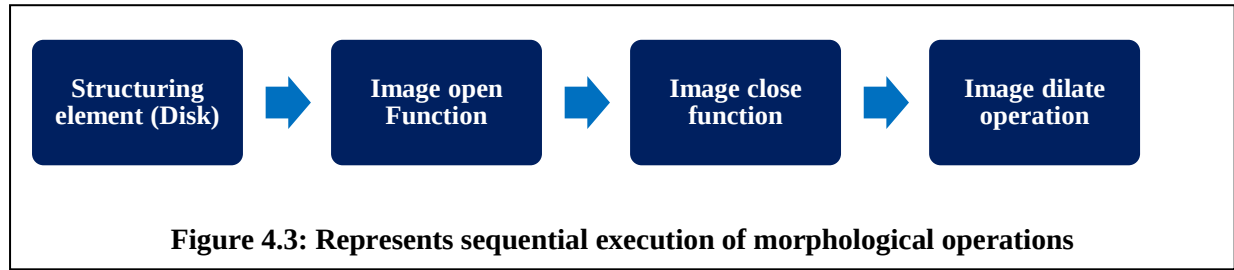
localized from the image frames, and further, the eye pair is recognized over the facial region using the viola-jones algorithm [73]. The Eye-pair region is then cropped and considered a region of interest, rectangular in shape for ensuing pre-processing by converting the color RGB image to a grayscale image. There are two advantages of converting a color image to a grayscale; first, it reduces the processing time and enhances the overall efficiency. Another benefit is the meticulous feature reduction. Afterward, histogram equalization performs image enhancement, followed by image adjustment and binary conversion. Furthermore, image cropping and segmentation provided significant regions for feature extraction.

4.6.2 Morphological Operations

Four morphological operations are performed on the enhanced grayscale eye-pair image to form and structure it for flawless final image segmentation, as represented in Figure 4.3.

Histogram Equalization = S_N

$$S_N = T(r_N) = \sum_{j=0}^N P_r(r_j) = \sum_{j=0}^N \left(\frac{n_j}{n}\right) \quad (4.1)$$



Here, S_N signifies an output image's intensity, corresponding to an input image r_N , N characterizes the grey level, and P represents probability [115]. Now, the preliminary step is structure element function with disk shape in accordance with the edifice eye retina due to its disk-like shape. The second and third steps include an image opening function followed by image closing to form an appropriate image filter. In the morphological opening of P by B , erosion of P by B is followed by dilation of the result by B , i.e., $P \circ B = (P \ominus B) \oplus B$. Here, $\ominus$ it represents erosion and $\oplus$ represents dilation. Congruently, dilation is followed by erosion in the morphological closing of P by B , i.e., $P \bullet B = (P \oplus B) \ominus B$. Subsequently, a dilation operation is performed on the filtered image for thickening. After that, the edge detection operation is performed to distinguish edges from that dilated image.

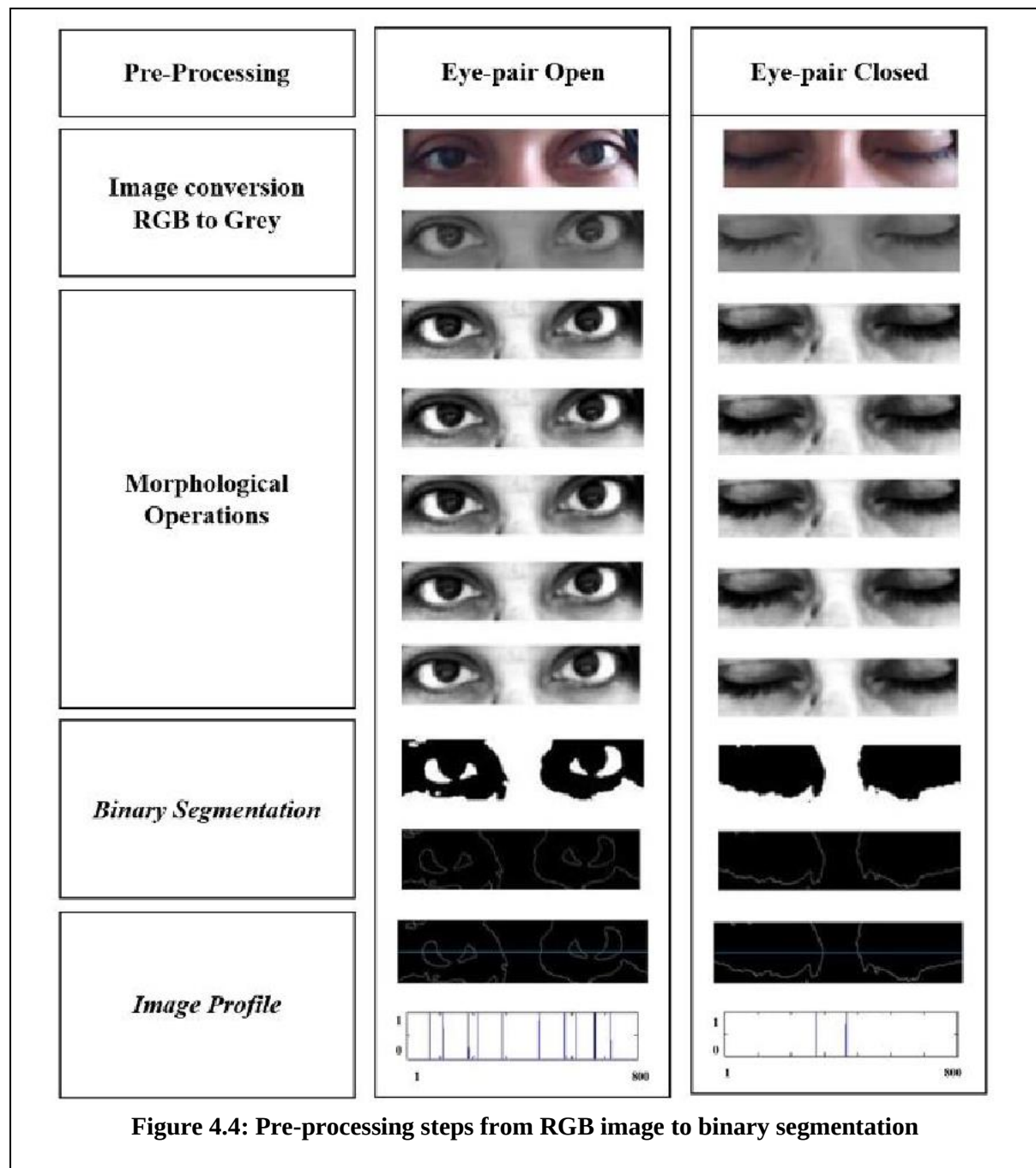
4.6.3 Feature Extraction

Feature extraction is a crucial step in image analysis and signal processing. Feature extraction helped to reduce the dimensionality of data by selecting or transforming relevant features and eliminating redundant or irrelevant information. It improved computational efficiency with fewer features to process and algorithms can operate more quickly, making them more scalable and suitable for large datasets.

In proposed research, feature extraction process involves multiple steps, firstly the RGB eye frames are converted to a grayscale image for performance enhancement. Further, these frames

have been pre-processed, and morphological operations are performed to make the frames appropriate for feature extraction for eye state discrimination, as shown in Figure 4.4.

To differentiate closed eye frames and open eye frames, the 'image profile' function is applied on the segmented edges, obtained after morphological operations. The 'image profile' function



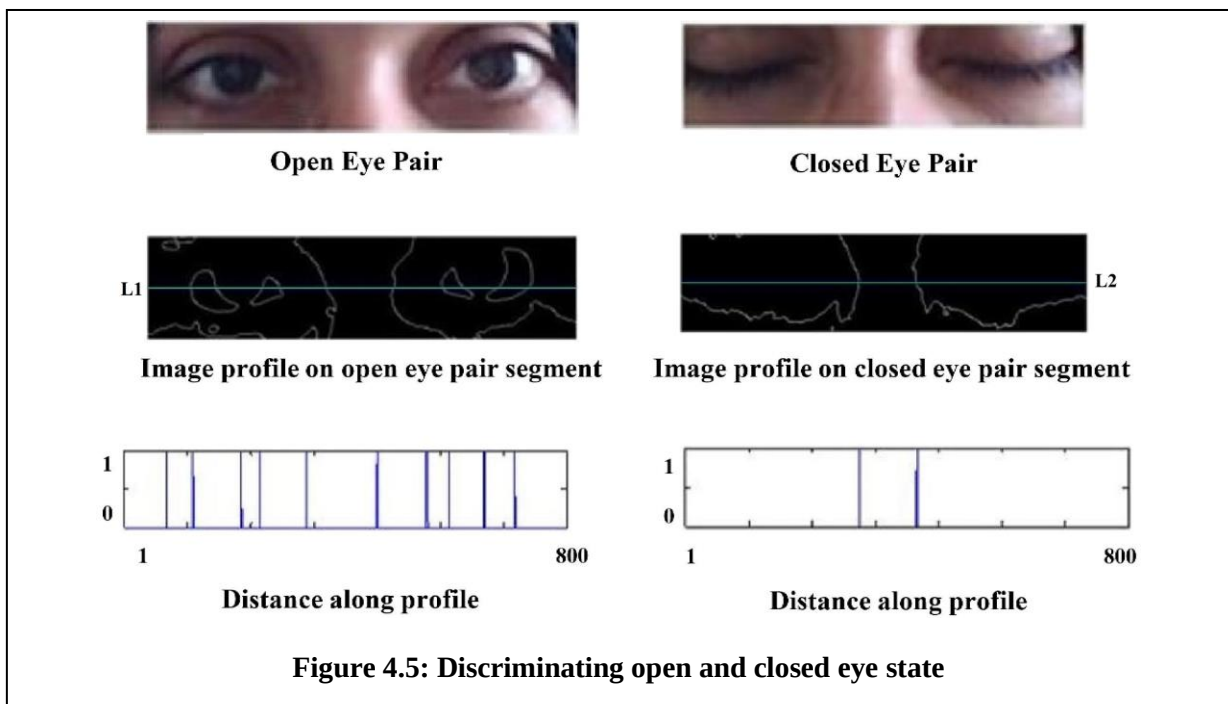
gives a piece of information about the value of pixels along a line. A reference row is selected from the processed image matrix to obtain an 'image profile', and then pixel values are plotted. Based on pixel values, a threshold limit is considered as a set point, and hence a binary classifier is structured to differentiate open and closed eye pair frames from a sample video clip.

4.6.4 Detection of open eye pair and closed eye pair

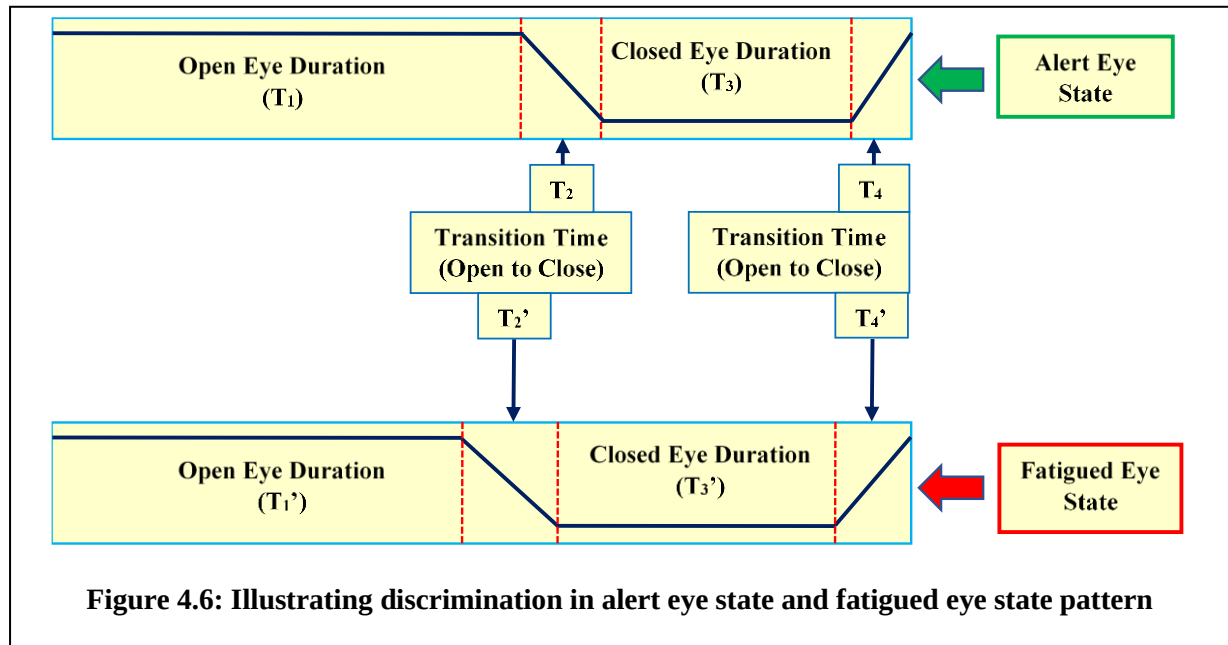
An algorithm developed for binary classification determines the open-eye pair and closed-eye pair. It can be observed from Figure 4.5 that for open-eye images, there are more than eight vertical lines in 'image profiles'. These vertical lines represent white pixels with value one across the lines L1 and L2. Pupil and nose boundaries create more than eight vertical lines in open eyes. However, when the eyes are closed, the binary segmentation only observed nose boundaries; therefore, only two vertical lines appear in an 'image profile'. Image profile gives a set of intensity values taken from regularly spaced points along a line segment or multi-lines in the image. This line graph plots the pixel values along the line segment of the image. A line profile plots the variations of intensity along a line. It returns the grayscale values of the pixels along a line and graphs it. An algorithm for the same has been developed to automatically detect the open or closed eyes from the number of vertical lines.

During the frame's analysis process, after determining open and closed eye frames, including frames in the transition phase, a typical pattern is observed in eye-opening and closing activity.

- a) $T1 > T1' \rightarrow$ Alert state
- b) $T2' + T4' > T2 + T4 \rightarrow$ Fatigued state
- c) $T3' > T3 \rightarrow$ Fatigued state



This transition state further helps determine a specific subject's alert and fatigue state by consequent eye-opening-closing trend. Figure 4.6 indicates that longer eye open duration in an alert state, whereas closed eye duration, is higher in fatigued subjects. Transition time is also elevated in the case of a sleepy subject.



4.6.5. Classification of Alert and Fatigue State

After distinguishing the open and closed eye states, the subsequent step is to discriminate between the fatigued and alert states from segmented frames. The significant features have been identified and extracted from the segmented frames, and data analysis has been performed. In total, 11 features have been extracted, and further feature reduction is performed based on the FDR (Fisher discrimination ratio) and box plots. The FDR with high value means the feature influences to differentiate two classes. Equation 4.2 has been considered for computing the FDR for the n th feature [116].

$$FDR_n = \left| \frac{\mu_{an} - \mu_{bn}}{\sqrt{(\sigma_{an}^2 + \sigma_{bn}^2)}} \right| \forall n \in [1, 11] \quad (4.2)$$

After eliminating less significant features, the five best features with higher FDRs have been considered for distinguishing alert and fatigue states. The last step is to differentiate between fatigue and alert state from the pre-processed images, done by the algorithm developed for binary classification. The developed intelligent algorithm and classification technique is a

novel approach for detecting eye-opening and closing and further for classifying subjects for fatigue and alert state.

This technique is also helpful for other applications requiring eye-opening and closing detection. The main emphasis is on detecting drowsy eyes through frames' analysis, with a higher degree of accuracy for detecting fatigue-induced due to sleep deprivation.

Here, the Alert state (mentioned as class 'a') and Fatigue state (mentioned as class 'b') are classified based on the features' discriminative value and their normalized weight (WF).

Table 4.1 Significant features after feature reduction through FDR

S. No.	Features	Description	FDR
1	Open Eye Duration (OED)	Total time duration in which eyes are open in a 2-minute video clip	3.02
2	Closed Eye Duration (CED)	Total time duration in which eyes are closed in a 2-minute video clip	0.97
3	Transition Time Open to Close (TTOC)	Time taken by eye state to change from open-eye to close-eye	2.63
4	Transition Time Close to Open (TTCO)	Time taken by eye state to change from close-eye to open-eye	3.29
5	Total Transition Time (TTT)	Total time taken by eye state to change from close to open and open to close (TTCO+TTOC)	3.00
6	Total Open Eye Frames (TOEF)	Total number of frames with the open-eye state in a 2-minute video clip of 3600 frames	3.02
7	Total Closed Eye Frames (TCEF)	Total number of frames with the close-eye state in a 2-minute video clip of 3600 frames	0.98
8	Inter Frame Interval Open to Close (IFIOC)	Frames count in transition phase from open eye state to close eye state	2.63
9	Inter Frame Interval Close to Open (IFICO)	Frames count in transition phase from close eye state to open eye state	3.28
10	Total Inter Frame Interval (TIFI)	Time is taken from open-eye to close-eye state and vice-versa	3.00
11	Total Blink Count (TBC)	Total number of eye blinks in the 2-minute video clip	1.87

The five most significant features out of eleven are selected and weighted based on a higher FDR value. These 5 significant features are shown bold in Table 4.1 in the similar kind of model was proposed by Singh M. et al. [116].

In our proposed method, the sampled value (x) for a feature is considered by its weight, i.e., weighted FDR (WF). Then normalized weight to be contained within the final value.

This normalized weight is the degree of closeness for the value of a feature to the mean (μ) of that feature. The normalized weight of the feature F1 in the class ‘a’ (alert) is calculated as Z_a .

$$Z_a = (1 - \frac{\mu_{a1} - x}{\sigma_{a1}}) \quad (4.3)$$

Here, sampled value ‘x’ is from feature F1; we will evaluate that concerning both classes a (using equation 4.3) and b (using equation 4.4). In class, ‘b’ (fatigue) can be computed as Z_b .

$$Z_b = (1 - \frac{\mu_{b1} - x}{\sigma_{b1}}) \quad (4.4)$$

Subsequently, weighted FDR for class ‘a’ will be:

$$WZ_a = \sum_1^n (Z_{a_n} WF_n) \quad (4.5)$$

Here n= 5

Similarly, weighted FDR for class ‘b’ will be:

$$WZ_b = \sum_1^n (Z_{b_n} WF_n) \quad (4.6)$$

Here n= 5

Also, for dynamic threshold value of ‘T’ for sample will be:

$$T = WZ_a - WZ_b \quad (4.7)$$

In the last step, this calculated value T of an unknown sampled value (x), is compared with the threshold band (Th1 to Th2 values).

Calculated as:

Th1 = [Mean of (WZ_a) + Standard deviation (from training dataset of Alert samples)]

Th2 = [Mean of (WZ_b) + Standard deviation (from training dataset of Fatigued samples)]

This threshold band has been already calculated from the training dataset. The threshold band is selected using different values of Th1 and Th2 to get the maximum accuracy.

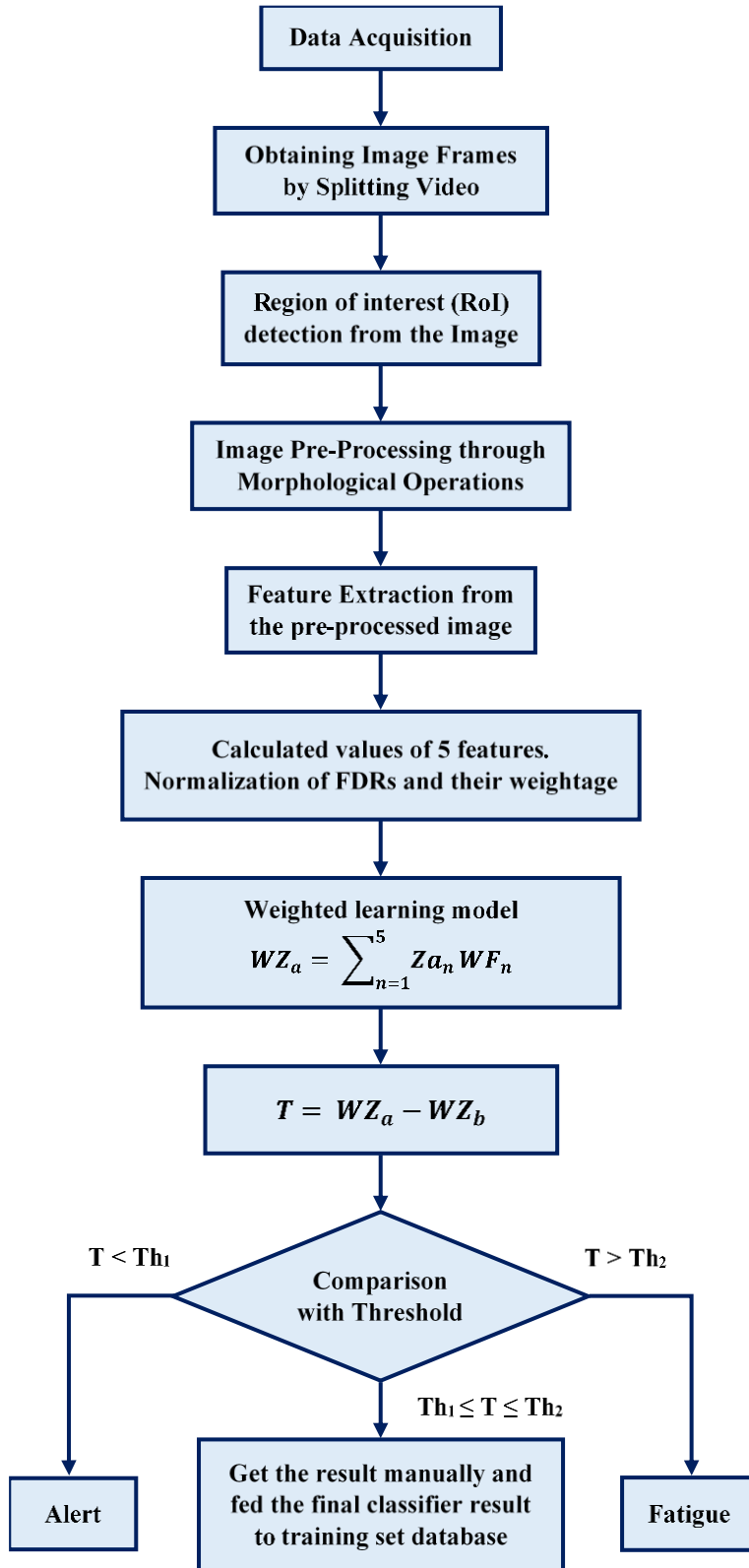


Figure 4.7: Implementation methodology

To get the classification result, T is compared with the threshold.

If $T < \text{lower threshold value Th1}$, then the subject is alert, and if $T \geq \text{upper threshold Th2}$, then the subject will be fatigued. In case It is a grey area, if the value of T is in between Th1 and Th2 then further investigation is required. Figure 4.7 demonstrates a stepwise representation of the implementation methodology.

4.7 Results and Discussion

Experimental study exhibits that sleep deprivation has a major impact on ocular state and decision-making abilities. Several changes occur in the cognitive abilities and ocular state due to sleep deprivation. These decisive deviations reflect as eye cues and associated facial expressions. The sample features for five randomly selected subjects for fatigue and alert state are shown in Table 4.2.

Table 4.2 Alert and Fatigue state sample features of 5 subjects

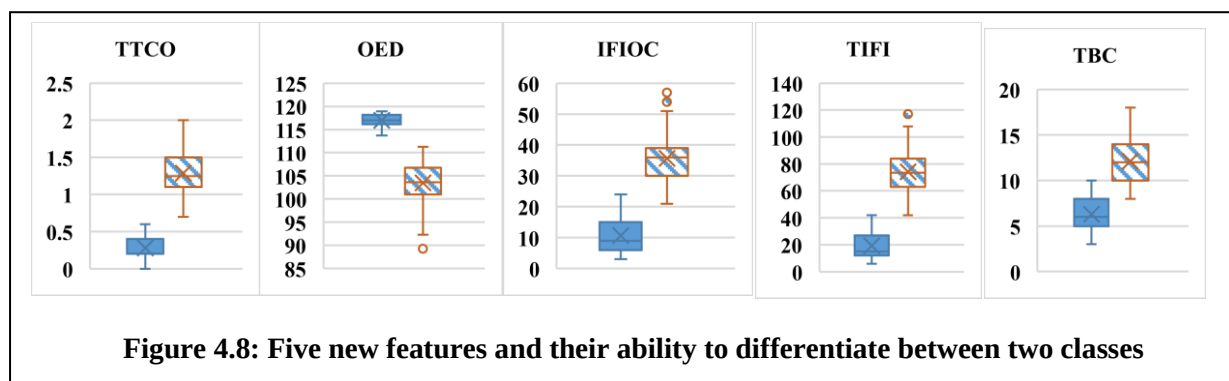
	Alert State					Fatigue State				
Features	OED	TTCO	IFIOC	TIFI	TBC	OED	TTCO	IFIOC	TIFI	TBC
Subject 1	117.9	0.2	9	15	5	95.7	1.7	51	102	16
Subject 2	116.2	0.4	15	27	8	89.3	2	57	117	18
Subject 3	116.7	0.2	9	15	7	103.7	1.40	39	81	13
Subject 4	118.1	0.2	6	12	4	109.6	0.7	21	42	9
Subject 5	115.6	0.5	18	33	6	108	0.9	24	21	11

The first feature is total transition time taken by an eye from close to open state (TTCO), which is elevated in the case of the fatigued subject. The box plot for the TTCO is demonstrated in Figure 4.8. The second feature is open eye duration (OED), which is declined in case of fatigue. A box plot for OED is represented in Figure 4.8. The third feature is the Inter Frame Interval for Open to Close eye (IFIOC) box plot for IFIOC is given in Figure 4.8. It is the interval between several consecutive Open Eye Frames (captured @ 30FPS) and the occurrence of consecutive Close Eye Frames, which is also longer when the subject is fatigued. Total inter-frame interval (TIFI) is the fourth feature, which is higher in fatigue state. The box plot for the TIFI is shown in Figure 4.8. The fifth feature is total blink counts (TBC); TBC's box plot is

exhibited in Figure 4.8. It has been observed that the total number of blinks is increased due to fatigue caused by sleep deprivation.

As a gold standard of fatigue detection, self-declaration by the volunteers is considered as a baseline for class ‘a’ (alert state) and class ‘b’ (fatigue state). The comparative analysis of validation is described in the form of a confusion matrix. Finally, performance analysis and results are discussed along with the concluding summary and future scope.

The proposed algorithm is trained for obtaining the weights of several features used in the classification. This trained algorithm then goes in self-learning mode as it interacts with the environment and gets a penalty for each wrong/undetermined classification. In our study, the



trained framework has been deployed in the Defence Institute of Physiology and Allied Sciences (DIPAS) New Delhi, INDIA, the sponsoring agency of the research work. The pilots (subject) under scrutiny are monitored for this classification. Parallely they also undergo a medical examination. The results of our framework are not only validated by the results of the medical examination, but the thresholds and weights in an adaptive manner to improve the subject-specific tuning, hereby improving the accuracy after each use. Upon achieving satisfactory subject-specific accuracy, the framework is capable of handling the classification without medical supervision.

4.8 Conclusion and Summary

To detect fatigue due to sleep deprivation, trials have been performed on 60 volunteers in two states. A total of 120 samples were taken (60 alert + 60 fatigue). Five novel features based on eye cues have been proposed. The confusion matrix of the obtained results of alert state and fatigue state are shown in Table 4.3. Overall classification accuracy of 97.5% is achieved successfully.

Table 4.3: Confusion Matrix for Fatigued State

True Positive	False Positive
59	1
False Negative	True Negative
2	58

From the experimental study, it has also been concluded that inter-frame interval transition features are useful for classification. Five features proposed in this framework are sensitive to sleep deprivation-induced fatigue. It is mentionable that close to open eye inter-frame interval increases due to sleep- deprived induced fatigue. The same five features have been fed to two other classifiers also.

Table 4.4: Comparative accuracy with different classifiers

% Accuracy	
Statistical Binary Classifier	85.8
kNN; k= 3	91.6
kNN; k= 5	93.3
kNN; k= 7	90.8
Proposed	97.5

The comparison of the proposed framework has been presented in Table 4.4, where traditional classifiers k Nearest Neighbor (kNN) and statistical binary classifiers are compared. The results show that the proposed algorithm, due to its self-learning capability, outperforms other classifiers used for the non-intrusive method. Also, it is observed that by using adaptive learning the results are improved significantly.

Chapter 5

Adaptive Self-Learning for Multimodal Feature Fusion to Detect Sleep Deprived Fatigue

5.1 Introduction

Adaptive self-learning is a machine-learning approach that involves adjusting the learning process of a model based on its performance and the data it encounters. The goal of adaptive self-learning is to improve the model's performance over time by allowing it to adapt to changes in the data distribution and evolving patterns. In traditional machine learning, models are typically trained on a static dataset, and the learned parameters remain constant throughout the deployment of the model. However, in adaptive learning, the model's parameters are updated or adapted as new data becomes available over a period of time. This can be particularly useful in scenarios where the underlying data distribution changes over time or where the model needs to continuously learn from new data. Adaptive self-learning involves selecting the most informative instances from past data to update the database and include in the training process. The model adapts by updated training sets for instances to improve its performance. Adaptive self-learning methods can be powerful tools for addressing challenges in dynamic and changing environments. They enable machine learning models to continually evolve and improve their performance, making them more robust and capable of handling real-world scenarios with varying data distributions.

5.2 Subjects

We have considered a group of 90 individuals between age group 20 years to 28 years for voluntary participation and provided the informed consent to understand the purpose of the study, its procedures, potential risks, and benefits, and willingly agree to participate without compulsion. They were abstained from the intake of alcohol, coffee, tea, and other psychoactive and psychotropic drugs to restrict the disturbance in sleeping patterns of participants. Also, we have confirmed the availability and accessibility as per the requirements of our experimental study. Furthermore, we have confirmed that the participants should not

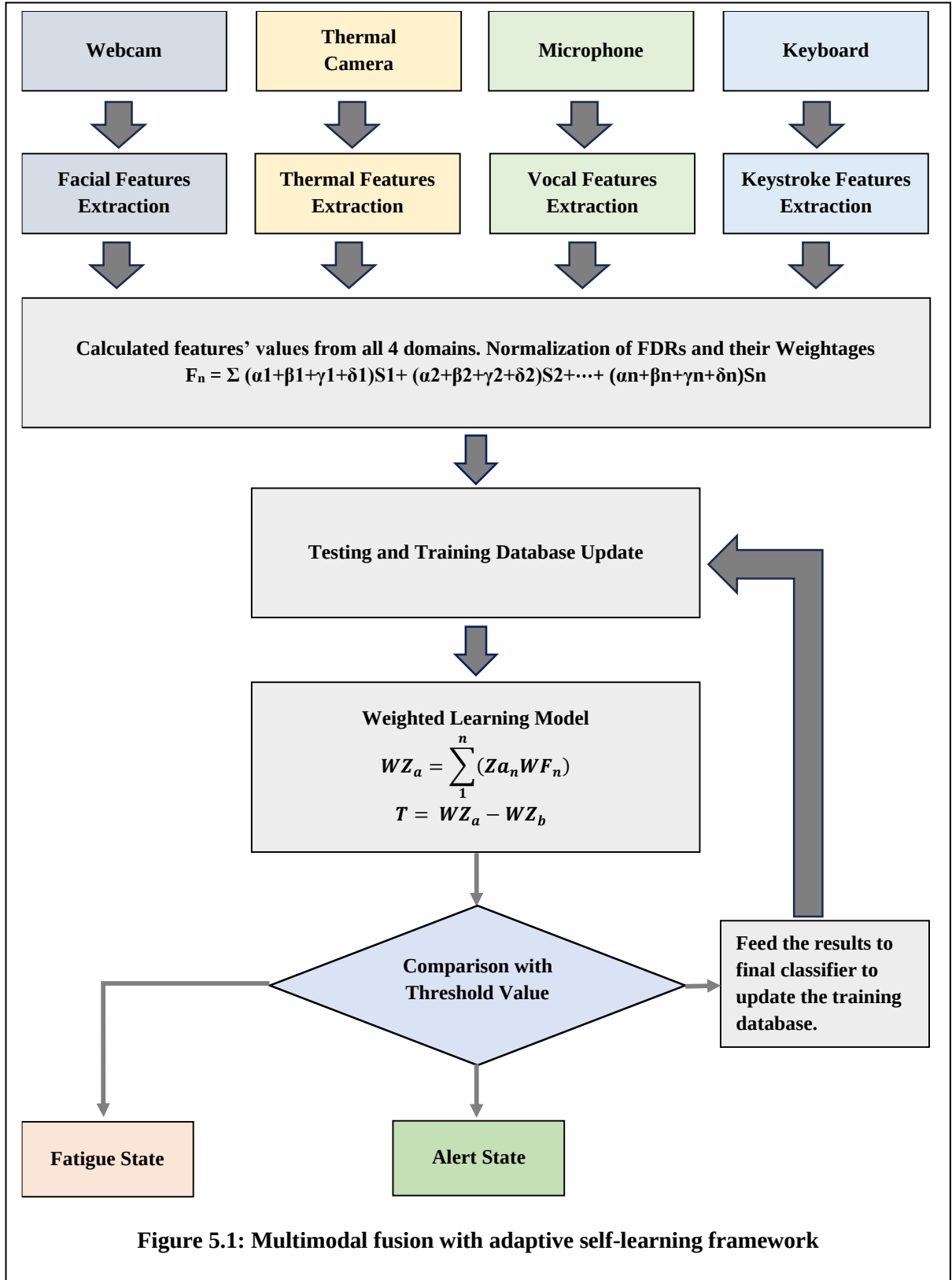
have any kind of pre-existing physiological conditions related to medications, lifestyle or other factors that could have significant impact on the outcomes of our experimental study. outcomes. All the participants were well aware of their obvious right to withdraw from the experimental study at any time without any consequences. Moreover, we have followed the guidelines of our institutional review board (IRB) before the participation of any volunteers for our experimental study. To maintain the consistency in the circadian rhythms the samples were acquired in the morning between 7:00 AM to 11:00 AM. Data is acquired from the four domains, 1. Visual facial images 2. Thermal images. 3. Voice acquisition. 4. Keystroke data acquisition.

5.3 Methodology

During the study it is observed that the results of multimodal feature fusion can be enhanced through a self-learning framework. We acquired the features from all four domains i.e., Visual facial images, thermal images, voice and keystrokes. Feature reduction was performed through FDR and weightages (α , β , γ , and δ) were assigned to respective domain based on FDR normalization. After acquiring features from all domains, features are fused through an adaptive self-learning framework for updating the training set and for autonomous weighting. The proposed adaptive self-learning feature fusion system allows the feature-level fusion of signals from multiple domains through the MatLab optimization toolbox and considered the Multiobjective Optimization Goal Attainment Method. Here, $S_1 + S_2 + \dots S_n$ represents n number of subjects and their normalized FDR weights for respective domain, i.e., α for Visual facial images, β for Thermal spectrum Images, γ for Keystroke dynamics and δ for Vocal features as shown in Figure 5.1 for whole methodology. Cumulative result F_n is obtained as a summation of all the normalized FDR weights of respective subjects as explained in equation 5.1.

$$F_n = \Sigma (\alpha_1 + \beta_1 + \gamma_1 + \delta_1)S_1 + (\alpha_2 + \beta_2 + \gamma_2 + \delta_2)S_2 + \dots + (\alpha_n + \beta_n + \gamma_n + \delta_n)S_n \quad (5.1)$$

Also, we increased the sample size of 180 samples (90 Alert + 90 Fatigued). Trained our model with 120 samples (60 Alert and 60 Fatigued) and tested on the remaining 60 samples (30 Alert + 30 fatigued). Performed cross-validation by using five different classifiers for a baseline of



performance. The highest accuracy of 98.33% was obtained by the proposed method detecting 59 correctly classified subjects out of 60.

5.4 Results and Discussion

For the optimize the classification through multimodal feature fusion, 120 samples were taken for training (60 from fatigued subject and 60 from alert subject). These were performed through five different classifiers to assess accuracy of classification. Accuracy of first and second classifiers i.e., k-Nearest Neighbors (kNN) and Random Tree provided an accuracy of 93.33%. Through both the classifiers 56 instances were classified correctly out of 60. Third classifier we used was classified through Random Forest with an accuracy of 95% ad classified 57 correct instances out of 6. Support Vector Machines (SVM) and Multilayer Perceptron gave an accuracy of 96.66% by detecting 58 correctly classified out of 60 as shown in Table 5.1. The highest accuracy of 98.33% was obtained by the proposed method detecting 59 correctly classified subjects out of 60.

Table 5.1: Comparison of results with other classifiers for adaptive self-learning

S. No.	Name of classifier	Accuracy %	Correctly Classified Instances	Incorrectly Classified Instances
1	kNN	93.33	56	4
2	Random Tree	93.33	56	4
3	Random Forest	95	57	3
4	SVM	96.66	58	2
5	Multilayer Perceptron	96.66	58	2
6	Proposed Method	98.33	59	1

5.5 Conclusion and Summary

There will be detrimental consequences on the individual's physical performance and capabilities due to inadequate sleep and disturbed sleep pattern. Fatigue induced through sleep deprivation is a major cause of fatal accidents because fatigue has a significant impact in the awareness and worsens individuals' vigilance [118]. Moreover, assessment of fatigue induced due to sleep deprivation is extremely challenging because the objective measures for quantification of fatigue and drowsiness level are not well defined. The proposed methodology

for fatigue detection is based on Multimodal feature fusion techniques and adaptive self-learning framework.

The confusion matrix for the proposed adaptive self-learning multimodal feature fusion methodology is shown in Table 5.2. Individually, the Vocal analysis domain has a maximum detection accuracy of 83.33%. Whereas, the second highest is of Thermal spectra images, with a detection accuracy of 76.66%. The Keystroke dynamics is on the third rank with a detection accuracy of 75% and the lowest detection accuracy of 73.33% is from Visual spectra images. To conclude, the proposed adaptive self-learning framework with multimodal feature fusion methodology demonstrated the highest detection average accuracy of 98.33%

Table 5.2: Accuracy of individual and multimodal through adaptive self-learning

Domains	Subject	True	False	Accuracy
1. Visual spectra image	Positive	23	7	73.33.5%
	Negative	9	21	
2. Thermal spectra image	Positive	24	6	76.66%
	Negative	8	22	
3. Keystroke dynamics	Positive	23	7	75%
	Negative	8	22	
4. Vocal Analysis	Positive	26	4	83.33%
	Negative	6	24	
5. Proposed Multimodal feature fusion technique	Positive	30	0	98.33%
	Negative	1	29	

The experimental study exhibits optimized performance after combining adaptive self-learning and fusion of features from multiple domains. The empirical weights have been initially assigned during feature fusion according to the detection to optimize the detection accuracy of classifiers.

Chapter 6

Major Findings and Future Scope

6.1 Major Findings

In our research work the major findings are widely dependent on our methodology and objectives. Major findings suggest directions for future studies based on critical findings. These can help to identify the areas that need further investigation. Some of the advantageous major findings from proposed research study are listed below:

1. Through the findings based upon statistical results we have identified that the vocal features and thermal imaging features are highly discriminatory for the identification of fatigue.
2. The study provides several proxies for fatigue induced due to sleep deprivation. In this study, our experiments indicate that sleep-deprived persons are also highly fatigued.
3. Through analyzing the patterns and trends of eye opening and closing we have identified it as a discriminative factor hence we considered interframe interval as a significant feature.
4. If we discuss qualitative insights, we have discovered that a non-intrusive methodology is comparatively better as compared to intrusive means of detecting fatigue due to extremely low distortion.

Moreover, there is one unexpected discovery as well, this finding was not a part of the original research plan the unexpected finding is described below:

1. Based on caveats and interpretation we observed that fatigue detection through Galvanic Skin Resistance (GSR) is highly challenging task because GSR is very dynamic in nature, and it was challenging to maintain a consistent baseline for measuring the GSR because multiple factors affect the GSR measurements. GSR is affected by the type and amount of diet, liquid intake, sweating, emotional state, and thoughts. Ambient environment factors like sunlight and humidity. Including GSR in multimodal fusion has been dropped due to the inconsistent baseline challenge. These can be particularly interesting and may warrant further exploration.

In addition, there are a few limitations with this approach as fatigue due to sleep deprivation is a complex and subjective phenomenon influenced by various factors. Below are some of the limitations associated with our experimental study:

1. Detecting fatigue due to sleep deprivation is a highly subjective experience, and individuals may perceive and respond to fatigue differently. What one person feels fatigue another may not, making it difficult to establish universal criteria for fatigue detection. Moreover, the requirement of amount of sleep is also subjective.
2. Also due to lack of a gold standard and objective measures it is difficult to quantify the fatigue. Till the present date there is no definitive marker that universally and exclusively quantifies fatigue. stress. While various technologies and assessments can help detect stress (e.g., heart rate monitors, self-report questionnaires), these methods are not foolproof. They may produce false positives or negatives and need validation.
3. Moreover, fatigue induced due to sleep deprivation is multifactorial in nature because it is influenced by a multitude of factors, including personal experiences, environmental circumstances, genetics, and individual coping mechanisms. It is often manifested in various ways, including physical symptoms, emotional changes, and behavioral patterns, making it challenging to isolate a single cause.
4. We have also observed that there is a context dependency as well. The perception fatigue often depends on the context in which it occurs. An individual may feel fatigued in one situation but may not feel fatigued in another. Detecting stress accurately requires considering the context context in which it occurs. For example, an individual with a smart phone with internet may not feel fatigued as compared to an individual reading a book due to an obligation that he dislikes. Though both as at same level of sleep deprivation.

Moreover, accurate stress detection often requires a holistic approach that considers multiple factors, including self-reporting, physiological measures, and contextual information. Additionally, respecting individuals' privacy and ethical considerations is crucial when implementing fatigue detection methods in various settings.

6.2 Challenges

Detecting human fatigue due to sleep deprivation can be challenging because it involves various factors and can manifest differently in individuals.

As the emotive state changes, there is a significant change in several physiological parameters; To detect human fatigue considerable parameters are facial expressions, voice patterns, and body reflexes but there are some key challenges associated with these parameters:

- a. Subjectivity: Fatigue and sleep deprivation can affect people differently. Some individuals may show obvious signs of fatigue, while others may develop more subtle symptoms. Additionally, individuals may have varying levels of tolerance to sleep deprivation.
- b. Masking effects: Some people might attempt to mask their fatigue by using stimulants like caffeine or energy drinks, which can temporarily alleviate the feeling of tiredness.
- c. Psychological factors: Stress, anxiety, and other psychological factors can influence how fatigue is perceived and expressed. People may attribute their tiredness to stress or other external factors, rather than recognizing it as sleep deprivation.
- d. Adaptation: Over time, individuals may adapt to chronic sleep deprivation, making it harder to identify the symptoms. This phenomenon, known as "learned sleepiness," can lead people to underestimate their fatigue levels.
- e. Individual differences: People have different sleep needs and patterns, which can make it challenging to set standard criteria for detecting fatigue. Additionally, underlying medical conditions and lifestyle choices can also impact the manifestation of fatigue.
- f. Contextual factors: Fatigue levels may vary depending on the time of day, type of work, or the presence of other influencing factors.

A few challenges were encountered in detecting fatigue during various project phases, which are described below:

1. Fatigue detection through visual spectra facial images:

In this phase, a significant challenge was the accurate segmentation of the subject's face. The segmentation was not precise in the preliminary attempts due to the ambient light. The problem was resolved by using a diffused light source and a well-defined protocol for settings.

2. Fatigue detection through thermal images:

The main challenge was the variation in the temperature color map on the thermal camera due to nearby high-temperature sources, e.g., computers and power converters. Color map variations lead to false results due to inaccurate thermal image segmentation. This problem was resolved by identifying and removing the nearby energy sources so that the color map of the thermal camera remained consistent.

3. Fatigue detection through voice:

The only challenge faced while taking voice samples was the ambient noise, which was eliminated by selecting a noise-cancellation microphone headset.

There is ongoing research to address these challenges, detecting and managing fatigue due to sleep deprivation remains a complex and multifaceted issue. It requires a comprehensive approach that combines objective measures, subjective assessments, and an individualized understanding of sleep patterns and needs.

6.3 Future Scope

The challenges and limitations discussed previously can be considered as a future scope. Hence further investigation is required in these areas because there is scope for improvement. Below are a few points which can be specifically considered for further research:

1. Proposed multimodal feature fusion is subject-specific which can be generalized. Many individuals are skilled at concealing their stress, either due to social expectations, fear of judgment, or other reasons. This can lead to underreporting of stress levels, making it difficult to detect accurately.
2. Artificial Intelligence or machine learning techniques can be implemented to eliminate the requirement of a controlled environment and to overcome the limitations and to extract features for GSR for fatigue detection due to sleep deprivation.
3. Addressing adaptive responses is another interesting area of research because the body and mind can adapt to fatigue over time, which may result in individuals not exhibiting typical signs of fatigue. This adaptation can make it challenging to identify fatigue in some cases. Also, due to the variability over time the fatigue levels can vary throughout the day, week, or year. An individual might be fatigued at one moment may not be

fatigued at another time. This temporal variability adds complexity to fatigue detection and challenging areas of research.

4. Quantification of fatigue is also a significant area of research. There is no standard physiological or psychological marker to detect fatigue due to sleep deprivation, and exclusively indicates stress. There is a requirement of objective measures and gold standards for quantification to detect fatigue level fatigue universally.

Effective fatigue detection relies significantly on multimodal feature fusion, as it enables the incorporation of information from diverse sources or modalities. Integration with machine learning and artificial intelligence will further enhance the overall accuracy and robustness of fatigue detection systems. Fatigue, being a complex and multifaceted phenomenon with various behavioral and physiological manifestations, benefits from the comprehensive insights provided by multimodal feature fusion. This approach allows the amalgamation of data from different sources, including facial expressions, eye movements, thermal images, and cognitive performance (reflex actions), offering a more holistic understanding of an individual's fatigue state. The versatility of multimodal feature fusion extends the applicability of fatigue detection systems to real-world, challenging scenarios such as driving, aviation, healthcare, and industrial settings. By combining information from multiple modalities machine learning and artificial intelligence the system becomes more resilient and well-suited for practical deployment in diverse situations.

Annexure - I: Publications (SCI) and Patent

A. Publications (SCI)

1. Title: An intelligent framework for detection of fatigue induced by sleep-deprivation

Authors: Virk Jitender Singh, Singh Mandeep, Singh Mandeep, Panjwani Usha, and Ray Koshik.

DOI: <https://doi.org/10.3233/JIFS-189784>

Journal: Journal of Intelligent & Fuzzy Systems, vol. 42, no. 2, pp. 1223-1233, 2022.

Impact Factor: 1.737 (2022)

Published on: 25 January 2022

2. Title: A Multimodal Feature Fusion Framework for Sleep-Deprived Fatigue Detection to Prevent Accidents

Authors: Virk Jitender Singh, Singh Mandeep, Singh Mandeep, Panjwani Usha, Ray Koushik.

DOI: <https://doi.org/10.3390/s23084129>

Journal: Sensors, Special Issue - Biosignal Sensing and Analysis for Healthcare Monitoring, Biomedical Sensors, vol. 23, no. 8, 4129, 2023.

Impact Factor: 3.847 (2023)

Published on: 20 April 2023

B. Patent

1. Title: Multi-Modal Alertness Assessment System

Application Number: 201811026712 A

Date of filing: 17 July 2018

Journal: The Patent Office Journal No. 04/2020 Dated 24/01/2020, pp. 4698.

DOI: https://ipindia.gov.in/writereaddata/Portal/IPOJournal/1_4835_1/Part-1.pdf

Date of Publication: 24 January 2020

Status: Published

Annexure - II: References

- [1] Jouvet M., “Neurophysiology of the States of Sleep,” The American Physiological Society, Physiological Reviews, vol. 47, no. 2, pp. 117–177, PMID: 5342870, 1967. <https://doi.org/10.1152/physrev.1967.47.2.117>
- [2] Swanson L.M., Arnedt J.T., Rosekind, M.R., Belenky G., Balkin T.J., and Drake, C., “Sleep disorders and work performance: findings from the 2008 National Sleep Foundation Sleep in America poll,” Journal of Sleep Research, vol. 20, no.3, pp. 487–494, 2011. <https://doi.org/10.1111/j.1365-2869.2010.00890.x>
- [3] Scott J.P.R., McNaughton L.R., and Polman Remco C.J., “Effects of sleep deprivation and exercise on cognitive, motor performance and mood,” Physiology & Behavior, vol. 87, no. 2, pp. 396–408, 2006. <https://doi.org/10.1016/j.physbeh.2005.11.009>
- [4] Poudel G.R., Innes C.R.H., Bones P.J., Watts R., and Jones R.D., “Losing the struggle to stay awake: Divergent thalamic and cortical activity during microsleeps,” Human brain mapping, vol. 35, no. 1, pp. 257–269, 2014. <https://doi.org/10.1002/hbm.22178>
- [5] Ji Q., Zhu Z., and Lan P., “Real-time nonintrusive monitoring and prediction of driver fatigue,” IEEE Transactions on Vehicular Technology, vol. 53, no. 4, pp. 1052–1068, 2004. <https://doi.org/10.1109/TVT.2004.830974>
- [6] Liao W., Zhang W., Zhu Z., and Ji Q., “A Real-Time Human Stress Monitoring System Using Dynamic Bayesian Network,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Workshops, San Diego, CA, USA, pp. 70–70, 2005. <https://doi.org/10.1109/CVPR.2005.394>
- [7] Dinges D.F., Rider R.L., Dorrian J., McGlinchey E.L., Rogers N.L., Cizman Z, Goldenstein S.K., Vogler C., Venkataraman S., and Metaxas D.N., “Optical computer recognition of facial expressions associated with stress induced by performance demands,” Aviation Space, and Environmental Medicine, vol. 76, no. 6, B172-B182. PMID: 15943210, 2005. <https://pubmed.ncbi.nlm.nih.gov/15943210/>
- [8] Vural E., Cetin M., Ercil A., Littlewort G., Bartlett M., and Movellan J., “Drowsy Driver Detection Through Facial Movement Analysis,” Human–Computer Interaction,

- Lecture Notes in Computer Science, vol. 4796, pp. 6–18, Springer, Berlin, Heidelberg. HCI, 2007. https://doi.org/10.1007/978-3-540-75773-3_2
- [9] Flores M.J., Armingol J.M., and Escalera A.de la, “Real-time drowsiness detection system for an intelligent vehicle,” 2008 IEEE Intelligent Vehicles Symposium, Eindhoven, Netherlands, pp. 637–642, ISSN: 1931-0587, 4-6 June 2008. <https://doi.org/10.1109/IVS.2008.4621125>
 - [10] Cheon Y., and Kim D., “Natural facial expression recognition using differential-AAM and manifold learning,” Pattern Recognition, vol. 42, no. 7, pp. 1340–1350, ISSN: 0031-3203, 2009. <https://doi.org/10.1016/j.patcog.2008.10.010>
 - [11] Fan X., Sun Y., Yin B., and Guo X., “Gabor-based dynamic representation for human fatigue monitoring in facial image sequences,” Pattern Recognition Letters, vol. 31, no. 3, pp. 234–243, ISSN: 0167-8655, 2010. <https://doi.org/10.1016/j.patrec.2009.08.014>
 - [12] Yao K.P., Lin W.H., Fang C.Y., Wang J.M., Chang S.L., and Chen S.W., “Real-Time Vision-Based Driver Drowsiness/Fatigue Detection System,” 2010 IEEE 71st Vehicular Technology Conference (VTC 2010-Spring), Taipei, Taiwan, pp.1–5, 16-19 May 2010. <https://doi.org/10.1109/VETECS.2010.5493972>
 - [13] Brookhuis K.A., and Waard D.de, “Monitoring driver’s mental workload in driving simulators using physiological measures,” Accident Analysis & Prevention, vol. 42, no. 3, pp. 898–903, ISSN: 0001-4575, 2010. <https://doi.org/10.1016/j.aap.2009.06.001>
 - [14] Benedetto S., Pedrotti M., Minin L., Baccino T., Alessandra R., and Montanari R., “Driver workload and eye blink duration,” Transportation Research Part F: Traffic Psychology and Behaviour, vol. 14, no. 3, pp. 199–208, ISSN: 1369-8478, 2011. <https://doi.org/10.1016/j.trf.2010.12.001>
 - [15] Moore S., and Bowden R., “Local binary patterns for multi-view facial expression recognition,” Computer Vision and Image Understanding, vol. 115, no. 4, pp. 541–558, 2011. <https://doi.org/10.1016/j.cviu.2010.12.001>
 - [16] Xiao R., Zhao Q., Zhang D., and Shi P., “Facial expression recognition on multiple manifolds,” Pattern Recognition, vol. 44, no. 1, pp. 107–116, ISSN: 0031-3203, 2011. <https://doi.org/10.1016/j.patcog.2010.07.017>

- [17] Ji Y., and Idrissi, K., “Automatic facial expression recognition based on spatiotemporal descriptors,” *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1373–1380, 2012. <https://doi.org/10.1016/j.patrec.2012.03.006>
- [18] Sandbach G., Zafeiriou S., Pantic M., and Rueckert D., “Recognition of 3D facial expression dynamics,” *Image and Vision Computing*, vol. 30, no. 10, pp. 762–773, 2012. <https://doi.org/10.1016/j.imavis.2012.01.006>
- [19] Chen S., Tian Y.L., Liu Q., and Metaxas D.N., “Recognizing expressions from face and body gesture by temporal normalized motion and appearance features,” *Image and Vision Computing*, vol. 31, no. 2, pp. 175–185, ISSN: 0262-8856, 2013. <https://doi.org/10.1016/j.imavis.2012.06.014>
- [20] Dornaika F., Moujahid A., and Raducanu B., “Facial expression recognition using tracked facial actions: Classifier performance analysis,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 467–477, ISSN: 0952-1976, 2013. <https://doi.org/10.1016/j.engappai.2012.09.002>
- [21] Forsman P.M., Vila B.J., Short R.A., Mott C.G., and Van Dongen H.P.A., “Efficient driver drowsiness detection at moderate levels of drowsiness,” *Accident Analysis and Prevention*, vol. 50, pp. 341–350, 2013. <https://doi.org/10.1016/j.aap.2012.05.005>
- [22] Tang M., and Chen F., “Facial expression recognition and its application based on curvelet transform and PSO-SVM,” *Optik, Elsevier*, vol. 124, no. 22, pp. 5401–5406, ISSN: 0030-4026, 2013. <https://doi.org/10.1016/j.ijleo.2013.03.116>
- [23] Yun T., and Guan L., “Human emotional state recognition using real 3D visual features from Gabor library,” *Pattern Recognition*, vol. 46, no. 2, pp. 529–538, 2013. <https://doi.org/10.1016/j.patcog.2012.08.002>
- [24] Sharma N., Dhall A., Gedeon T., and Goecke R., “Modeling Stress Using Thermal Facial Patterns: A Spatio-Temporal Approach,” 2013 IEEE Affective Computing and Intelligent Interaction (ACII), Humaine Association Conference, Geneva, Switzerland, pp. 387–392, 2-5 September 2013. doi:10.1109/ACII.2013.70
- [25] Gao H., Yüce A., and Thiran J.P., “Detecting emotional stress from facial expressions for driving safety,” 2014-International Conference on Image Processing (ICIP), Paris, France, pp. 5961–5965, 27-30 October 2014. doi:10.1109/ICIP.2014.7026203

- [26] Eskil M.T., and Benli K.S., “Facial expression recognition based on anatomy,” *Computer Vision and Image Understanding*, vol. 119, pp. 1–14, ISSN: 1077-3142, 2014. <https://doi.org/10.1016/j.cviu.2013.11.002>
- [27] Benitez-Quiroz C.F., Rivera S., Gotardo P.F.U., and Martinez A.M., “Salient and Non-Salient Fiducial Detection using a Probabilistic Graphical Model,” *Pattern Recognition*, vol. 47, no. 1, pp. 208–215, 2014. <https://doi.org/10.1016/j.patcog.2013.06.013>
- [28] Mehta R., Yuan J., and Egiazarian K., “Face recognition using scale-adaptive directional and textural features,” *Pattern Recognition*, vol. 47, no. 5, pp. 1846–1858, 2014. <https://doi.org/10.1016/j.patcog.2013.11.013>
- [29] Naruniec J., “Discrete area filters in accurate detection of faces and facial features,” *Image and Vision Computing*, vol. 32, no. 12, pp. 979–993, ISSN: 0262-8856, 2014. <https://doi.org/10.1016/j.imavis.2014.09.004>
- [30] Owusu E., Zhan Y., and Mao Q.R., “A neural-AdaBoost based facial expression recognition system,” *Expert Systems with Applications*, vol. 41, no. 7, pp. 3383–3390, 2014. <https://doi.org/10.1016/j.eswa.2013.11.041>
- [31] Vaidya A.R., Jin C., and Fellows L.K., “Eye spy: the predictive value of fixation patterns in detecting subtle and extreme emotions from faces,” *Cognition*, vol. 133, no. 2, pp. 443–456, 2014. <https://doi.org/10.1016/j.cognition.2014.07.004>
- [32] Wang S., Liu Z., Wang J. Wang Z., Li Y., Chen, X., and Ji Q., “Exploiting multi-expression dependences for implicit multi-emotion video tagging,” *Image and Vision Computing*, Elsevier, vol. 32, no. 10, pp. 682–691, ISSN: 0262-8856, 2014. <https://doi.org/10.1016/j.imavis.2014.04.013>
- [33] Vitale J., Williams M.A., Johnston B., and Boccignone G., “Affective facial expression processing via simulation: A probabilistic model,” *Biologically Inspired Cognitive Architectures*, vol. 10, pp. 30–41, 2014. <https://doi.org/10.1016/j.bica.2014.11.005>
- [34] Yan J., Zhang X., Lei Z., and Li S.Z., “Face detection by structural models,” *Image and Vision Computing*, vol. 32, no. 10, pp. 790–799, ISSN: 0262-8856, 2014. <https://doi.org/10.1016/j.imavis.2013.12.004>
- [35] Azim T., Jaffar M.A., and Mirza A.M., “Fully automated real time fatigue detection of drivers through Fuzzy Expert Systems,” *Applied Soft Computing*, vol. 18, pp. 25–38, ISSN: 1568-4946, 2014. <https://doi.org/10.1016/j.asoc.2014.01.020>

- [36] McIntire L.K., McKinley R.A., Goodyear C., and McIntire J.P., “Detection of vigilance performance using eye blinks,” *Applied Ergonomics*, vol. 45, no. 2, Part B, pp. 354–362, ISSN: 0003-6870, 2014. <https://doi.org/10.1016/j.apergo.2013.04.020>
- [37] Ali H., Hariharan M., Yaacob S., and Adom A.H., “Facial emotion recognition using empirical mode decomposition,” *Expert Systems with Applications*, vol. 42, no. 3, pp. 1261–1277, ISSN: 0957-4174, 2015. <https://doi.org/10.1016/j.eswa.2014.08.049>
- [38] Zhang Y., Zhang L., and Hossain M.A., “Adaptive 3D facial action intensity estimation and emotion recognition,” *Expert Systems with Applications*, Elsevier, vol. 42, no. 3, pp. 1446–1464, ISSN: 0957-4174, 2015. <https://doi.org/10.1016/j.eswa.2014.08.042>
- [39] An L., Yang S., and Bhanu B., “Efficient smile detection by Extreme Learning Machine,” *Neurocomputing*, vol. 149, Part A, pp. 354–363, ISSN: 0925-2312, 2015. <https://doi.org/10.1016/j.neucom.2014.04.072>
- [40] Smith R.S., and Windeatt T., “Facial action unit recognition using multi-class classification,” *Neurocomputing*, vol. 150, Part B, pp. 440–448, ISSN: 0925-2312, 2015. <https://doi.org/10.1016/j.neucom.2014.07.066>
- [41] Ouyang Y., Sang N., and Huang R., “Accurate and robust facial expressions recognition by fusing multiple sparse representation based classifiers,” *Neurocomputing*, vol. 149, Part A, pp. 71–78, 2015. <https://doi.org/10.1016/j.neucom.2014.03.073>
- [42] Pawar R., Wamburkar S., Deshmukh R., and Awalkar N., “Driver Drowsiness Detection Using Deep Learning,” 2021 IEEE 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, pp. 1–7, 1-3 October 2021. <https://doi.org/10.1109/GCAT52182.2021.9587519>
- [43] Inthanon P., and Mungsing S., “Detection of Drowsiness from Facial Images in Real-Time Video Media using Nvidia Jetson Nano,” 2020 IEEE 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, pp. 246–249, 24-27 June 2020. <https://doi.org/10.1109/ECTI-CON49241.2020.9158235>
- [44] Hansen J.H.L., and Patil S., “Speech Under Stress: Analysis, Modeling and Recognition,” *Speaker Classification I. Lecture Notes in Computer Science 2007*, vol. 4343, pp. 108–137, Springer, 2007. https://doi.org/10.1007/978-3-540-74200-5_6

- [45] Tawari A., and Trivedi M.M., “Speech Emotion Analysis: Exploring the Role of Context,” *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 502–509, October 2010. <https://doi.org/10.1109/TMM.2010.2058095>
- [46] Fernandez R., and Picard R., “Recognizing affect from speech prosody using hierarchical graphical models,” *Speech Communication*, vol. 53, no. 9–10, pp. 1088–1103, ISSN: 0167-6393, 2011. <https://doi.org/10.1016/j.specom.2011.05.003>
- [47] Schuller B., Batliner A., Steidl S., and Seppi D., “Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge,” *Speech Communication*, vol. 53, no. 9–10, pp. 1062–1087, ISSN: 0167-6393, 2011. <https://doi.org/10.1016/j.specom.2011.01.011>
- [48] Lee C.C., Mower E., Busso C., Lee S., and Narayanan S., “Emotion recognition using a hierarchical binary decision tree approach,” *Speech Communication*, vol. 53, no. 9–10, pp. 1162–1171, 2011. <https://doi.org/10.1016/j.specom.2011.06.004>
- [49] Bozkurt E., Erzin E., Erdem, C.E., and Erdem A.T., “Formant position based weighted spectral features for emotion recognition,” *Speech Communication*, vol. 53, no. 9–10, pp. 1186–1197, ISSN: 0167-6393, 2011. <https://doi.org/10.1016/j.specom.2011.04.003>
- [50] Krajewski J., Schnieder S., Sommer D., Batliner A., and Schuller B., “Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech,” *Neurocomputing*, vol. 84, pp. 65–75, ISSN: 0925-2312, 2012. <https://doi.org/10.1016/j.neucom.2011.12.021>
- [51] Bhardwaj S., Srivastava S., Hanmandlu M., and Gupta J.R.P., “GFM-Based Methods for Speaker Identification,” *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 1047–1058, ISSN: 2168-2275, June 2013. <https://doi.org/10.1109/TSMCB.2012.2223461>
- [52] Caraty M.J., and Montacié C., “Vocal fatigue induced by prolonged oral reading: Analysis and detection,” *Computer Speech & Language*, vol. 28 no. 2, pp. 453–466, ISSN: 0885-2308, 2014. <https://doi.org/10.1016/j.csl.2012.12.003>
- [53] Jaimes A., and Sebe N., “Multimodal human-computer interaction: A survey.” *Computer Vision and Image Understanding*, vol. 108 no. 1–2, pp. 116–134, 2007. <https://doi.org/10.1016/j.cviu.2006.10.019>

- [54] Schleicher R., Galley N., Briest S., and Galley L., “Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?,” *Ergonomics*, vol. 51, no.7, pp. 982–1010, PMID: 18568959, 2008. <https://doi.org/10.1080/00140130701817062>
- [55] Vizer L.M., Zhou L., and Sears A., “Automated stress detection using keystroke and linguistic features: An exploratory study,” *International Journal of Human-Computer Studies*, vol. 67, no. 10, pp. 870–886, 2009. <https://doi.org/10.1016/j.ijhcs.2009.07.005>
- [56] Sharma N., and Gedeon T., “Objective measures, sensors and computational techniques for stress recognition and classification: A survey,” *Computer methods and programs in biomedicine*, vol. 108, no. 3, pp. 1287–1301, PMID: 22921417, ISSN: 0169-2607, 2012. <https://doi.org/10.1016/j.cmpb.2012.07.003>
- [57] Lim Y.M., Ayesb A., and Stacey M., “Detecting cognitive stress from keyboard and mouse dynamics during mental arithmetic,” *2014 Science and Information Conference* pp. 146–152, London, UK, 2014. <https://doi.org/10.1109/SAI.2014.6918183>
- [58] Nazmul Haque Nahin A.F.M., Alam J. M., Mahmud H., and Hasan K., “Identifying emotion by keystroke dynamics and text pattern analysis,” *Behaviour & Information Technology*, Taylor & Francis, vol. 33, no. 9, pp. 987–996, ISSN: 0144-929X, 2014. <https://doi.org/10.1080/0144929X.2014.907343>
- [59] Cui Z., Sun H.M., Yin R.N., Li G., Sun H.B., and Jia R.S., “Real-time detection method of driver fatigue state based on deep learning of face video,” *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 25495–25515, ISSN:1573-7721, 2021. <https://doi.org/10.1007/s11042-021-10930-z>
- [60] Xu T., Wang H., Lu G., Wan F., Deng M., Qi P., Bezerianos A., Guan C., and Sun Y., “E-Key: An EEG-Based Biometric Authentication and Driving Fatigue Detection System,” *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 864–877, ISSN: 1949-3045, April-June 2023, <https://doi.org/10.1109/TAFFC.2021.3133443>
- [61] Ngai W.K., Xie H., Zou D., and Chou K.L., “Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources,” *Information Fusion*, vol. 77, pp. 107–117, 2022. <https://doi.org/10.1016/j.inffus.2021.07.007>
- [62] Zhang J., Yin Z., Chen P., and Nichele S., “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Information Fusion*, vol. 59, pp. 103–126, 2020. <https://doi.org/10.1016/j.inffus.2020.01.011>

- [63] Li R., Chen Y.V., and Zhang L., “A method for fatigue detection based on Driver’s steering wheel grip,” *International Journal of Industrial Ergonomics*, vol. 82, 103083, ISSN: 0169-8141, 2021. <https://doi.org/10.1016/j.ergon.2021.103083>
- [64] Hooda R., Joshi V., and Shah M., “A comprehensive review of approaches to detect fatigue using machine learning techniques,” *Chronic Diseases and Translational Medicine*, vol. 8, no. 1, pp. 26–35, 2022. <https://doi.org/10.1016/j.cdtm.2021.07.002>
- [65] Frost C., Toczko M., Merrigan J.J., and Martin J.R., “The effects of sleep on firefighter occupational performance and health: A systematic review and call for action,” *Sleep Epidemiology*, vol. 1, 100014, 2021. <https://doi.org/10.1016/j.sleepe.2021.100014>
- [66] Tuncer T., Dogan S., and Subasi A., “EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection,” *Biomedical Signal Processing and Control*, vol. 68, 102591, ISSN: 1746-8094, July 2021. <https://doi.org/10.1016/j.bspc.2021.102591>
- [67] Kartsch V.J., Benatti S., Schiavone P.D., Rossi D., and Benini L., “A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems,” *Information Fusion*, vol. 43, pp. 66–76, 2018. <https://doi.org/10.1016/j.inffus.2017.11.005>
- [68] Ye C., Yin Z., Zhao M., Tian Y., and Sun Z., “Identification of mental fatigue levels in a language understanding task based on multi-domain EEG features and an ensemble convolutional neural network,” *Biomedical Signal Processing and Control*, vol. 72, Part B, ISSN: 1746-8094, 2022. <https://doi.org/10.1016/j.bspc.2021.103360>
- [69] Wu N., and Sun J., “Fatigue Detection of Air Traffic Controllers Based on Radiotelephony Communications and Self-Adaption Quantum Genetic Algorithm Optimization Ensemble Learning,” *Applied Sciences*, vol. 12, no. 20, 10252, 2022. <https://doi.org/10.3390/app122010252>
- [70] Siam A.I., Soliman N.F.; Algarni A.D., Abd El-Samie F.E., and Sedik A., “Deploying Machine Learning Techniques for Human Emotion Detection,” *Computational Intelligence and Neuroscience*, vol. 2022, ISSN: 1687-5265, Article ID 8032673, 2022. <https://doi.org/10.1155/2022/8032673>
- [71] Alnuaim A.A., Zakariah M., Alhadlaq A., Shashidhar C., Hatamleh W.A., Tarazi H., Shukla P.K., and Ratna R., “Human-Computer Interaction with Detection of Speaker Emotions Using Convolution Neural Networks,” *Computational Intelligence and*

- Neuroscience, vol. 2022, SN: 1687-5265, Article ID 7463091, 2022. <https://doi.org/10.1155/2022/7463091>
- [72] Tuncer T., Dogan S., Ertam F., and Subasi A., “A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals,” *Cognitive Neurodynamics*, vol. 15, no. 2, pp. 223–237, ISSN: 1871-4099, 2021. <https://doi.org/10.1007/s11571-020-09601-w>
 - [73] Fatima B., Shahid A.R., Ziauddin S., Safi A.A., and Ramzan H., “Driver Fatigue Detection Using Viola Jones and Principal Component Analysis,” *Applied Artificial Intelligence*, vol. 34, no. 6, pp. 456–483, ISSN: 0883-9514, 2020. <https://doi.org/10.1080/08839514.2020.1723875>
 - [74] O’Keeffe K., Hodder S., and Lloyd A., “A comparison of methods used for inducing mental fatigue in performance research: individualised, dual-task and short duration cognitive tests are most effective,” *Ergonomics*, Taylor & Francis, vol. 63, no. 1, pp. 1–12, ISSN: 0014-0139, 2020. <https://doi.org/10.1080/00140139.2019.1687940>
 - [75] Heaton K.J., Williamson J.R., Lammert A.C., Finkelstein K.R., Haven C.C., Sturim D., Smalt C.J., and Quatieri T.F., “Predicting changes in performance due to cognitive fatigue: A multimodal approach based on speech motor coordination and electrodermal activity,” *The Clinical Neuropsychologist*, vol. 34, no. 6, pp. 1190–1214, ISSN: 1385-4046, 2020. <https://doi.org/10.1080/13854046.2020.1787522>
 - [76] Sikander G., and Anwar S., “Driver Fatigue Detection Systems: A Review,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, ISSN: 1558-0016, 2019. <https://doi.org/10.1109/TITS.2018.2868499>
 - [77] Doudou M., Bouabdallah A., and Berge-Cherfaoui V., “Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges,” *International Journal of Intelligent Transportation Systems Research*, vol. 18, no. 2, pp. 297–319, ISSN: 1868-8659, 2019. <https://doi.org/10.1007/s13177-019-00199-w>
 - [78] Zhao C., Gao Z., Wang Q., Xiao K., Mo Z., and Jamal Deen M., “FedSup: A communication-efficient federated learning fatigue driving behaviors supervision approach,” *Future Generation Computer Systems*, vol. 138, pp. 52–60, ISSN: 0167-739X, 2023. <https://doi.org/10.1016/j.future.2022.08.009>

- [79] Zhang Y., Guo H., Zhou Y., Xu C., and Liao Y., “Recognising drivers’ mental fatigue based on EEG multi-dimensional feature selection and fusion,” *Biomedical Signal Processing and Control*, vol. 79, no. 2, 104237, ISSN: 1746-8094, 2022. <https://doi.org/10.1016/j.bspc.2022.104237>
- [80] Chen J., Yan M., Zhu F., Xu J., Li H., and Sun X., “Fatigue Driving Detection Method Based on Combination of BP Neural Network and Time Cumulative Effect,” *Sensors*, vol. 22, no. 13, 2022. <https://doi.org/10.3390/s22134717>
- [81] Li Y., Zhang S., Zhu G., Huang Z., Wang R., Duan X., and Wang Z., “A CNN-Based Wearable System for Driver Drowsiness Detection,” *Sensors*, vol. 23, no. 7, 2023. <https://doi.org/10.3390/s23073475>
- [82] Singh J., and Arora A.S., “Effectiveness of active dynamic and passive thermography in the detection of maxillary sinusitis,” *Quantitative InfraRed Thermography Journal*, vol. 18, no. 4, pp. 213–225, 2021. <https://doi.org/10.1080/17686733.2020.1736456>
- [83] Rodríguez A.O.R., Riaño M.A., Garía P.A.G., Marín C.E.M., and Crespo R.G., “Initial Approach for Construction of a Public Dataset for Emotional Analysis Through Brain-Computer Interfaces and Second Language Platforms,” *2023 IEEE Global Engineering Education Conference (EDUCON)*, Kuwait, Kuwait, pp. 1–7, ISSN: 2165-9567, 1-4 May 2023. <https://doi.org/10.1109/EDUCON54358.2023.10125173>
- [84] Tan W.R., Chan C.S., Yogarajah P., and Condell J., “A Fusion Approach for Efficient Human Skin Detection,” in *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 138–147, February 2012. <https://doi.org/10.1109/TII.2011.2172451>
- [85] Singh S., and Anand R.S., “Multimodal Medical Image Fusion Using Hybrid Layer Decomposition With CNN-Based Feature Mapping and Structural Clustering,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3855–3865, ISSN: 1557-9662, June 2020. <https://doi.org/10.1109/TIM.2019.2933341>
- [86] Schirra C.O., Bontus C., Stevendaal U., Dössel O., and Michael G., “Improvement of cardiac CT reconstruction using local motion vector fields,” *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 122–130, ISSN: 0895-6111, 2009. <https://doi.org/10.1016/j.compmedimag.2008.10.012>
- [87] Singh S., and Anand R.S., “Multimodal Medical Image Sensor Fusion Model Using Sparse K-SVD Dictionary Learning in Nonsubsampled Shearlet Domain,” *IEEE*

- Transactions on Instrumentation and Measurement, vol. 69, no. 2, pp. 593–607, ISSN: 1557-9662, February 2020. <https://doi.org/10.1109/TIM.2019.2902808>
- [88] Dhanawansa V., Samarasinghe P., Gardiner B., Yogarajah P., and Karunasena A., “The Automated Temporal Analysis of Gaze Following in a Visual Tracking Task,” *Image Analysis and Processing–ICIAP 2022, Lecture Notes in Computer Science*, vol. 13233, pp 324–336, Springer, Cham, 2022. https://doi.org/10.1007/978-3-031-06433-3_28
 - [89] Koh M.-S., Rodriguez-Marek E., and Fischer C., “A new two dimensional empirical mode decomposition for images using inpainting,” *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, Beijing, China, pp. 13–16, 24-28 October 2010. doi:10.1109/ICOSP.2010.5656860
 - [90] Bhardwaj S., Srivastava S., Vaishnavi S., and Gupta J.R.P., “Chaotic time series prediction using combination of Hidden Markov Model and Neural Nets,” *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, Krakow, Poland, pp. 585–589, 2010. doi:10.1109/CISIM.2010.5643518
 - [91] Kumar M., Singh D., and Deepak K.K., “Fractal Analysis of Heart Dynamics During Attention Task,” *Advances in Computational Intelligence Techniques, Algorithms for Intelligent Systems*, Springer, Singapore, pp. 95–107, ISBN: 978-981-15-2619-0, 2020. https://doi.org/10.1007/978-981-15-2620-6_7
 - [92] Farina D., Jiang Y., and Dössel O., “Acceleration of FEM-based transfer matrix computation for forward and inverse problems of electrocardiography,” *Medical & Biological Engineering & Computing*, vol. 47, no. 12, pp. 1229–1236, 2009. <https://doi.org/10.1007/s11517-009-0503-7>
 - [93] VanDam M., and De Palma P., “A modular, extensible approach to massive ecologically valid behavioral data,” *Behavior Research Methods* vol. 51, no. 4, pp. 1754–1765, ISSN: 1554-3528, 2019. <https://doi.org/10.3758/s13428-018-1167-8>
 - [94] Bhatt U., Singh A., Bhadauria H.S., and Kumar M., “Image super resolution based on discrete and Stationary wavelet transform using Canny edge extraction and non local mean,” *2016 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, pp. 1–5, 2016. doi:10.1109/INVENTIVE.2016.7830216

- [95] Gupta J., Gupta N., Kumar M., Ritwik D., Singh J., Chakraborty C., Alazab, M., and Do, D.-T., “Smart and Secure Internet of Healthcare Things,” (1st edition). ISBN: 9781003239895, CRC Press, 2022. <https://doi.org/10.1201/9781003239895>
- [96] Malhotra S., Kumar A., Bansal A., and Kumar M., “Enhanced Face Recognition System Using Age Invariant Features,” 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, vol. 1, pp. 1690–1694, 5-6 July 2019. doi:10.1109/ICICICT46008.2019.8993134
- [97] Fairhurst M.C., and Kaplani E., “Perceptual analysis of handwritten signatures for biometric authentication,” IEE Proceedings: Vision, Image and Signal Processing, vol. 150, no. 6, pp. 389–394, 2003. <https://doi.org/10.1049/ip-vis:20031046>
- [98] Pancholi S., Joshi A.M., and Joshi D., “DLPR: Deep Learning-Based Enhanced Pattern Recognition Frame-Work for Improved Myoelectric Prosthesis Control,” IEEE Transactions on Medical Robotics and Bionics, vol. 4, no. 4, pp. 991–999, ISSN: 2576-3202, November 2022. doi:10.1109/TMRB.2022.3216957
- [99] Nagla K., Singh D., and Uddin M., “Sensor fusion framework for robust occupancy grid mapping,” 2013 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, pp. 1–8, October 2013. doi:10.1109/AIPR.2013.6749330
- [100] Singla S.K., and Arora A.S., “Speaker Verification System using LabVIEW,” IETE Technical Review, Taylor & Francis, vol. 24, no. 5, pp. 403–412, ISSN: 0256-4602, 2007. <https://www.tandfonline.com/doi/abs/10.4103/02564602.10876623>
- [101] Fairhurst M. C., and Kaplani E., “A study of the user interface in biometric verification based on the characteristics of human signature checking,” Knowledge-Based Intelligent Information and Engineering Systems,” KES 2003, Oxford, UK, Proceedings, Part II, vol. 2774, pp. 1270–1276, Springer Berlin/Heidelberg, September 2003. https://doi.org/10.1007/978-3-540-45226-3_172
- [102] Joshi D., Upadhayay M.D., and Joshi S.D., “Robust language and speaker identification using image processing techniques combined with PCA,” 2013 INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATION (ICSC), Noida, India, pp. 213–218, December 2013. doi:10.1109/ICSPCom.2013.6719785

- [103] Koh M.-S., Rodriguez-Marek E., and Talarico C., “Class Projects With Graphic User Interfaces In Matlab,” Paper presented at 2007 Annual Conference & Exposition, Honolulu, Hawaii, ISSN: 2153-5965, June 2007. <https://doi.org/10.18260/1-2--1760>
- [104] Singh S., Tripathy M., and Anand R. S., “Subjective and Objective Analysis of Speech Enhancement Algorithms for Single Channel Speech Patterns of Indian and English Languages,” IETE Technical Review, Taylor & Francis, vol. 31, no. 1, pp. 34–46, ISSN: 0256-4602, 2014. <https://doi.org/10.1080/02564602.2014.890840>
- [105] De Palma P., “Probabilistic Methods in Automatic Speech Recognition,” In Encyclopedia of Information Science and Technology, Third Edition, pp. 253–261. Hershey, PA: IGI Global, 2015. <https://doi.org/10.4018/978-1-4666-5888-2.ch024>
- [106] Koh M.-S., and Rodriguez-Marek E., “Speech enhancement of color noise using empirical mode decomposition,” 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, pp. 1688–1692, November 2012. doi:10.1109/ACSSC.2012.6489320
- [107] Sharma A., and Singh D., “Power Spectral Density Analysis of HRV to Evaluate Changes in ANS During Graded Head-Up Tilt and Head-Reverse Tilt,” In Optimization of Industrial Systems, pp. 61–72, 2022. <https://doi.org/10.1002/9781119755074.ch5>
- [108] Gope J.C., Tabassum T., Mabrur M.M., Yu K., and Arifuzzaman M., “Sentiment Analysis of Amazon Product Reviews Using Machine Learning and Deep Learning Models,” 2022 IEEE International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, pp. 1–6, 24-26 February 2022. doi:10.1109/ICAEEE54957.2022.9836420
- [109] Wijesinghe A., Samarasinghe P., Seneviratne S., Yogarajah P., and Pulasinghe K., “Machine Learning Based Automated Speech Dialog Analysis Of Autistic Children,” 2019 11th International Conference on Knowledge and Systems Engineering (KSE), Da Nang, Vietnam, pp. 1–5, October 2019. doi:10.1109/KSE.2019.8919266
- [110] Xu Z., Guo Y., Chakraborty C., Hua Q., Chen S., and Yu K., “A Simple Federated Learning-Based Scheme for Security Enhancement Over Internet of Medical Things,” in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 652–663, ISSN: 2168-2208, February 2023. <https://doi.org/10.1109/JBHI.2022.3187471>

- [111] Gupta N., Khosravy N., Patel N., Dey N., and Crespo R.G., “Lightweight Computational Intelligence for IoT Health Monitoring of Off-Road Vehicles: Enhanced Selection Log-Scaled Mutation GA Structured ANN,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 611–619, January 2022. doi:10.1109/TII.2021.3072045
- [112] Ding F., Fan B., Shen Z., Yu K., Srivastava G., Dev K., and Wan S., “Securing Facial Bioinformation by Eliminating Adversarial Perturbations,” *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6682–6691, ISSN: 1941-0050, May 2023. doi:10.1109/TII.2022.3201572
- [113] Middya A.I., Nag B., and Roy S., “Deep learning based multimodal emotion recognition using model-level fusion of audio–visual modalities,” *Knowledge-Based Systems*, Elsevier, vol. 244, 108580, ISSN: 0950-7051, March 2022. <https://doi.org/10.1016/j.knosys.2022.108580>
- [114] Tang L., Yuan J., and Ma J., “Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network,” *Information Fusion*, vol. 82, pp. 28–42, 2022. <https://doi.org/10.1016/j.inffus.2021.12.004>
- [115] Lim S.H., Mat Isa N.A., and Ooi C.H., et. al., “A new histogram equalization method for digital image enhancement and brightness preservation,” *Springer - Signal, Image and Video Processing, (SIViP)* vol. 9, no. 3 pp. 675–689, ISSN: 1863-1711, 2015. <https://doi.org/10.1007/s11760-013-0500-z>
- [116] Singh M., Singh S., and Gupta S., “An information fusion based method for liver classification using texture analysis of ultra-sound images,” *Information Fusion*, vol. 19, pp. 91–96, ISSN: 1566-2535, 2014. <https://doi.org/10.1016/j.inffus.2013.05.007>
- [117] Chen C.J., and Miller D.A., “Pitch-Synchronous Analysis of Human Voice,” *Journal of Voice*, Elsevier, vol. 34, no. 4, pp. 494–502, February 2019. <https://doi.org/10.1016/j.jvoice.2019.01.009>
- [118] Dai L., Li Y., and Zhang M., “Detection of Operator Fatigue in the Main Control Room of a Nuclear Power Plant Based on Eye Blink Rate, PERCLOS and Mouse Velocity,” *Applied Sciences*, vol.13, no. 4, 2023. <https://doi.org/10.3390/app13042718>