NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Computing and Informatics

Department of Informatics, Journalism and Media Studies

Recommending a Machine Learning Model to Detect the Fatigue State for Employees at Namdeb

Thesis submitted in partial fulfilment of the requirements for the degree of

Master of Data Science

at the

Namibia University of Science and Technology

By

Samuel Nghidengwa Nakale

Student Number: 221141995

Main Supervisor: Prof. Fungai Bhunu Shava

Co-Supervisor: Prof. Gloria Iyawa

Submission Date: April 2024

METADATA

TITLE:	Mr
STUDENT NAME:	Samuel Nghidengwa Nakale
SUPERVISOR:	Prof. Fungai Bhunu Shava
CO-SUPERVISOR:	Prof. Gloria Iyawa
DEPARTMENT:	Informatics, Journalism and Media Studies
QUALIFICATION:	Master of Data Science
SPECIALISATION:	Data Science
STUDY TITLE:	Recommending a Machine Learning Model to Detect the Fatigue State for Employees at Namdeb
KNOWLEDGE AREA:	Data Science
KEYWORDS:	Fatigue detection, machine learning, mining industry
TYPE OF RESEARCH:	Applied research
METHODOLOGY:	Quantitative research
STATUS:	Final thesis
SITE:	Windhoek campus
DOCUMENT DATE:	12 April 2024
SPONSOR:	NONE

Declaration

I, SAMUEL NGHIDENGWA NAKALE hereby declare that the work contained in the thesis for the degree MASTER OF DATA SCIENCE project, entitled: "RECOMMENDING A MACHINE LEARNING MODEL TO DETECT THE FATIGUE STATE FOR EMPLOYEES AT NAMDEB", is my own original work and that I have not previously in its entirety or in part submitted it at any university or other higher education institution for the award of a degree.

I further declare that I have fully acknowledged any sources of information I have used for the research in accordance with the Institution rules.

Signature: Date: 12 April 2024

Signature of the Supervisor

I, Fungai Bhunu Shava ,herewith declare that I supervised this research project.

Signature: Date: 12 April 2024

Retention and Use of Thesis

I, Samuel Nghidengwa Nakale being a candidate for the degree of Master of Data Science accept the requirements for the Namibia University of Science and Technology relating to the retention and use of thesis deposited in the Library and Information Services.

In terms if these conditions, I agree that the original of my thesis deposited in the Library and Information Services will be accessible for purpose of study and research, in accordance with the normal conditions established by the Librarian for the care, loan or reproduction of the thesis.

Signature: Date: 12 April 2024

Acknowledgements

Firstly, I would like to thank the Lord our God for the blessings bestowed on my life and for His sufficient grace and mercy. Thank you, Lord, for giving me wonderful parents who sacrificed so much to make sure that I get a good educational foundation.

I would also like to extend my gratitude to my supervisors, Prof. Bhunu Shava and Prof. Iyawa for their time, patience, insightful guidance, encouragement and most importantly for being great mentors to me throughout this journey. I look forward to collaborating with you in the near future.

Special thanks to Namdeb for allowing me to use their data. Mr Kandukira, Mr Du Plessis and Ms Blackburn, thank you for the time spent explaining how the ADAS works and for providing the data and all the necessary support.

To my family and loved ones who supported me throughout this journey, thank you so much for the support and most importantly for believing in me.

This thesis is dedicated to my lovely daughter, Amenenge.

Abstract

Workplace fatigue is one of the major risk factors across different industries and it negatively impacts productivity and workplace safety. Thus, fatigue detection and monitoring is essential to promote occupational health and safety. The advancements in data collection technologies have made it possible for industries to develop data driven solutions by developing Artificial Intelligence (AI) and Machine Learning (ML) based fatigue monitoring and detection systems. The Advanced Driver Assistance Systems (ADAS) is one of the technologies that has been adopted in various industries to improve safety for drivers. Namdeb implemented the ADAS in recent years and they have identified the need for a system to detect and classify employees' fatigue state using data from the ADAS. In this study, an ML based fatigue detection system was proposed. Facial behavioural fatigue features were used to detect fatigue. The proposed system deployed some of the commonly used ML classification algorithms and it was evaluated on a simulated dataset, the Yawning Detection Dataset (YawDD), and a real-world dataset, data from the Namdeb ADAS. The results showed that most of the supervised ML classifiers achieved a fatigue prediction accuracy above 90% for both datasets. The Random Forest (RF) based fatigue detection-based model was found to be the best model. The k-Means which is an unsupervised ML classifier exhibited the worst performance. However, the reliability and generalisability of the results based on the real-world dataset can be improved by using a larger dataset. The major challenge to developing behavioural based fatigue detection systems for real world setting like the mining environment is face detection accuracy which is affected by factors such as low image resolution due to poor and variable lighting conditions, face orientation to camera and proximity of face to the camera. The significant contribution of this study is the use of real-world dataset to test the proposed fatigue detection system. Overall, the study contributes to the promotion of the eighth Sustainable Development Goal (SDG) of promoting safe working environments.

Key words: *Fatigue classification, fatigue detection, fatigue state, machine learning, mining industry, occupational health and safety, video processing*

Table of Contents

List of Figures

List of Tables

List of abbreviations

ADAS	Advanced Driver Assistance System
AI	Artificial Intelligence
CNN	Convolutional Neural Network
DT	Decision Trees
EAR	Eye Aspect Ratio
FRMS	Fatigue Risk Management System
GDP	Gross Domestic Product
GNB	Gaussian Naïve Bayes
SDG	Sustainable Development Goals
ILO	International Labour Organisation
KNN	K-Nearest Neighbours
LR	Logistic Regression
MAR	Mouth Aspect Ratio
MOR	Mouth Opening Ratio
ML	Machine Learning
MLIREC	Ministry of Labour, Industrial Relations and Employment Creation
MME	Ministry of Mines and Energy
NSA	Namibia Statistics Agency
NUST	Namibia University of Science and Technology
PERCLOS	Percentage of time eyes are Closed
PMO	Percentage of Mouth Opening
RF	Random Forest
RNN	Recurrent Neural Networks
SVM	Support Vector Machines
WHO	World Health Organisation
XGBoost	EXtreme Gradient Boosting
YawDD	Yawning Detection Dataset

CHAPTER ONE: INTRODUCTION

This chapter provides the context of the study and it covers the background, problem statement, research objectives, research questions, scope of the study and its significance.

1.1 Background

Health is defined as "… a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organisation, 1948, p. 100). Work activities and employee health are inextricably linked. Therefore, in order to promote the physical and mental well-being of workers, the International Labour Organisation (ILO) has developed a number of international labour standards and instruments relating to occupational health and safety (International Labour Organization, 2003).

Fatigue, which is defined as "… a suboptimal psychophysiological condition caused by exertion" (Phillips, 2015, p. 53), is one of the workplace hazards and a risk factor across different industries impacting negatively on employee performance and safety in the workplace (Hooda et al., 2022; Yadav et al., 2020; Nasirzadeh et al., 2020). Fatigue and driver distraction are among the key driving safety risks (Perkins et al., 2023). Additionaly, fatigue is considered to be one of the main contributing factors to occupational injuries and accidents in hazardous industries such as the mining industry (Bauerle et al., 2021; Pelders & Nelson, 2019). Consequences of occupational injuries and accidents include loss of life, economic and social costs[1], loss in productivity due to time lost, functional limitations as well as mental and emotional trauma (Ajith & Ghosh, 2019; Phillips, 2015). In addition, fatigue can lead to increased employee turnover due to job dissatisfaction which can result in increased hiring costs (Grządzielewska, 2021).

The mining industry is among the most hazardous industries and it is characterised by high occupational risks (Bauerle et al., 2021). Generally, the mining environment makes mineworkers more susceptible to fatigue (Bauerle et al., 2018). Thus, mining workers may suffer a variety of physical and psychological effects that influence the onset of fatigue (Saputra & Purwitasari, 2022). Employee fatigue is considered to be the main cause of occupational injuries and fatalities in the mining industry (Drews et al., 2020), particularly for drivers. It is estimated that about 65%

[1] These include medical treatment expenses, compensation costs and possible reduction in income and livelihood for the injured person and their dependents.

of accidents involving truck drivers in open pit mines are due to fatigue (Fatigue Science, 2020). Therefore, it is important to have proper data driven fatigue detection and monitoring systems to accurately monitor mineworkers' fatigue and to ensure that fatigue incidences can be resolved before they become detrimental (Yadav et al., 2020).

Occupational safety and health are recognised in both international and national development frameworks and policies. The eighth Sustainable Development Goal (SDG8) which is aimed at promoting decent work for all has a target of promoting safe working environments. At a national level, there is a national occupational health and safety policy that is spearheaded by the Ministry of Labour, Industrial Relations and Employment Creation (MLIREC) (MLIREC, 2021). Among others, the policy mandates the directorate of mines under the Ministry of Mines and Energy (MME) to enforce occupational safety and health in the mining industry.

The advancements in technology have made it possible for industries to adopt artificial intelligence and machine learning based technologies to improve their operational efficiency and safety. The construction, aviation, road transport operators and health industries are among the leading industries in adopting data driven predictive fatigue management technologies (Hooda et al., 2022; Lee et al., 2020, Shalash, 2022). In particular, Chen (2022) found that Machine Learning (ML) algorithms perform well in detecting drivers' fatigue state and Chen posits that there is a great prospect for their application in fatigue detection systems for different domains. Furthermore, Jung and Choi (2021) maintain that the advancements in smart mining technologies have resulted in the availability of large amounts of data in real time, which makes it possible for ML based solutions to be utilised for occupational safety, among other mining operations.

1.1.1 Significance of the Namibian mining sector

The mining industry is a key industry for the socio-economic development of the Namibian economy. The industry is among the top five contributors to Namibia's Gross Domestic Product (GDP), on average contributing 9.5% per annum to the GDP over the past ten years (Namibia Statistics Agency (NSA), 2023). In addition, the industry spent N$ 16.823 billion on local procurement, contributed over N$ 4.4 billion to public revenue and provided 16 147 (8 232 permanent) direct employment in 2022 (Namibia Chamber of Mines, 2023).

As shown in Figure 1-1, the performance of the Namibian mining industry is mainly driven by the diamond mining sub-sector which contributes almost 50% of the annual mining output. This makes

Namdeb[2] an important mine for the Namibian economy. In 2022, Namdeb employed a total of 1 629 permanent employees (Namibia Chamber of Mines, 2023), representing 20% of the total permanent employees in the Namibian mining industry.

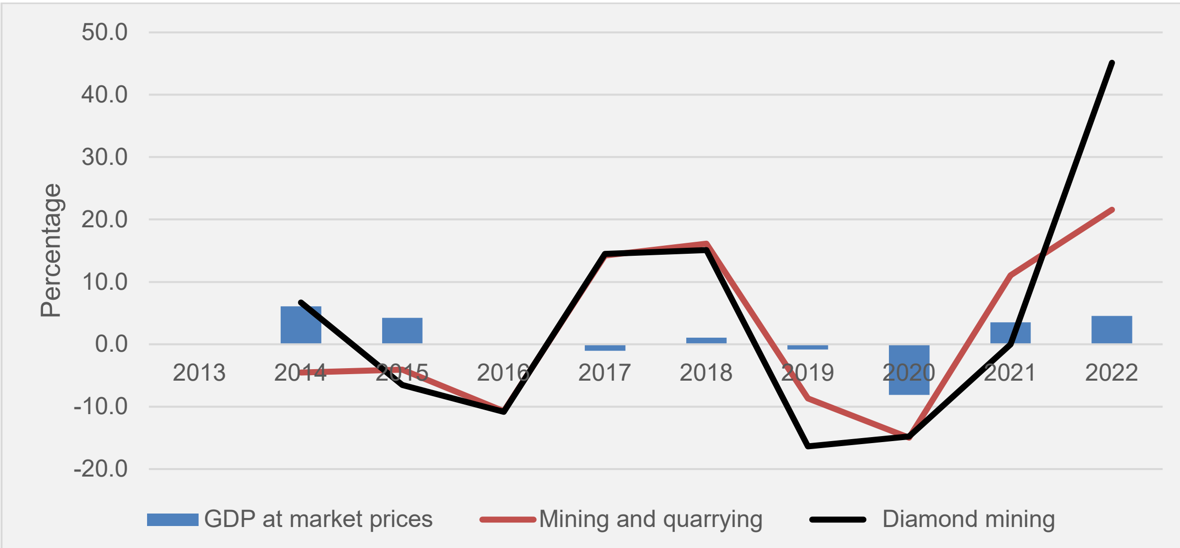


Figure 1-1: Namibian mining industry contribution to GDP

Source: Author compilation using NSA 2022 National Accounts data.

1.1.2 Occupational injuries in the Namibian mining industry

Safety performance in the mining industry is measured by the number of lost day injuries, number of disabling injuries and number of injuries. A report by the Namibia Chamber of Mines (2023) indicates that despite the increases between 2021 and 2022, overall, there has been a reduction in the number of lost day injuries, number disabling injuries and fatalities in the past decade. In addition, one fatality was recorded per year for the period 2018 to 2022 compared to two fatalities per year for the period 2016 to 2017. However, it is important to emphasise that even one fatality is one too many. Furthermore, fatigue does not only result in injuries and fatalities, but it also affects the performance and productivity of employees. Therefore, occupational safety and health still needs to be prioritised in the Namibian mining industry.

[2] Namdeb is a land-based diamond mine with operations in the southern coastal town of Namibia, Oranjemund. It is the largest diamond mine in Namibia and it has been in operation for close to a century.

1.2 Problem Statement

The Advanced Driver Assistance System (ADAS) is a set of intelligent systems and safety innovation technologies designed to assist drivers with safe motor vehicles operation (Ayachi et al., 2021). This technology is installed in motor vehicles and consists of several components including cameras to monitor driver behaviour. Short video clips recorded by the system are transmitted to a control room for monitoring driver behaviour. An analysis of the video data from the ADAS can be used to check and monitor indicators of drowsiness/fatigue. Namdeb has implemented the ADAS in recent years and general improvement in driver behaviour has been reported since the implementation of the system (Namibia Chamber of Mines, 2023). However, the system currently only identifies driver behaviour related to drowsiness/fatigue and does not explicitly detect and classify the fatigue state of drivers. Thus, further analysis of Namdeb ADAS data is required to detect and classify the fatigue state. This study proposes a ML based fatigue detection system for Namdeb drivers using the video data obtained from the existing ADAS. This can enhance Namdeb's fatigue detection and monitoring system in order to improve workplace safety.

1.3 Research Objectives

The main research objective of this study is to recommend an ML model that can be used to detect the fatigue state of employees at Namdeb.

The main research objective was achieved through the following sub-objectives:

- To extract fatigue features which can be used to predict employee fatigue at Namdeb;
- To evaluate and compare the performance of the fatigue detection models based on different ML algorithms; and
- To recommend the key factors to take into consideration when implementing a fatigue detection system for Namdeb employees.

1.4 Research Questions

Main research question:

How can an ML model be developed to detect the fatigue state of employees at Namdeb?

The sub-questions for this study are:

- What are the fatigue features that can be used to detect employee fatigue at Namdeb?

- How do different ML models perform in detecting and classifying employee fatigue?
- What are some of the key factors to take into consideration when selecting a fatigue detection model for Namdeb employees?

1.5 Scope and Delineation

The study only focused on detecting and classifying the fatigue state of truck drivers in the mining industry using Namdeb as a case study. The design of the proposed system was limited by the features which could be extracted from the available dataset.

1.6 Ethical Clearance

Prior to conducting the study, ethical clearance approval was obtained from the Namibia University of Science and Technology (NUST) Faculty of Computing and Informatics research ethics committee.

1.7 Significance of the Study

This study developed a fatigue detection system which can be used to manage fatigue related incidences and thereby minimise the effects of fatigue on mineworkers' performance and productivity. Overall, the study contributes to the promotion of the SDG8 of promoting safe working environments.

1.8 Thesis Outline

The rest of the thesis is organised into six chapters. Chapter two focuses on the literature review and outlines the work that has already been done on the topic, their findings, and the identification of areas where the current study can contribute and fill the literature gap. The methods and techniques that were used to answer the research questions and to achieve the research objectives of the study are detailed in Chapter three. In Chapter four, a description of the proposed system is provided. The implementation and evaluation of the proposed system is discussed in Chapter five. Chapter six concludes the study by checking if the study objectives have been met and if the research questions have been answered. Areas for possible future research are also identified.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of literature on fatigue detection which is an active area of research. The purpose of the review is to provide empirical evidence in current research and to determine gaps in the literature. In particular, the review is aimed at answering the following questions:

- *What fatigue detection technique and predictor features can be used in designing fatigue detection systems for the mining industry*? The aim is to explore the different fatigue detection methods and features in order to determine the ones that are more suitable for the mining domain.
- *What machine learning algorithms can be used in designing fatigue detection systems for the mining industry*? This aims to compare the performance of the different machine learning algorithms used in fatigue detection in order to determine the ones that are more suitable for designing fatigue detection systems for the mining industry.

2.2 Fatigue Detection Systems

Employee fatigue poses a potential risk to the health and wellbeing of employees. Thus, it is important to ensure that Fatigue Risk Management Systems (FRMS) are in place to measure, mitigate and manage fatigue related risks (Sprajcer et al., 2022). Fatigue detection systems are a key component of FRMS which can be used to proactively measure and monitor real time fatigue (Sprajcer et al., 2022). The key factors to consider when developing fatigue detection systems are accuracy, timeliness, efficiency and applicability in real world scenarios (Alharbey et al., 2022; Majeed et al., 2023; Zhu et al., 2022).

The general data flow in fatigue detection systems is displayed in Figure 2-1. The system has four stages, namely: data acquisition, data pre-processing, detection method and fatigue state classification. The following sections provide a discussion of these stages.

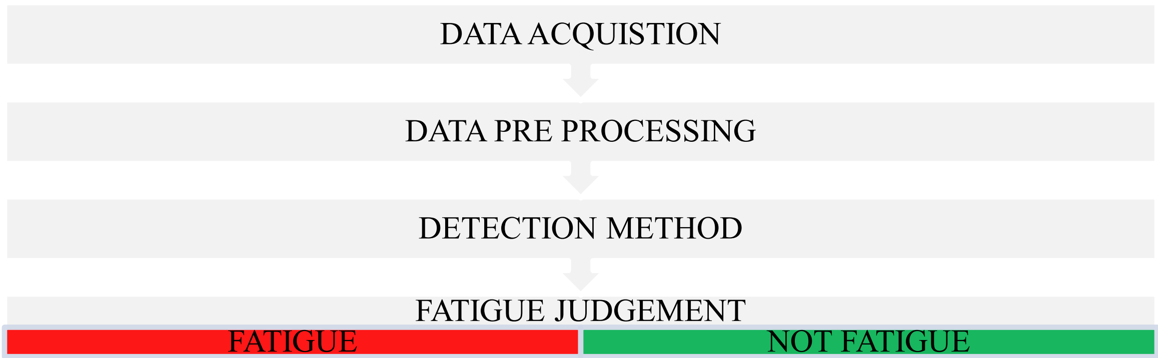


Figure 2-1: General data flow for fatigue detection systems

2.2.1 Data Acquisition

The various sources of data for fatigue detection include video data, biological/physiological signals obtained by using sensors which can be placed on a subject's body and vehicle movements information which is obtained by using sensors attached to the vehicle (El-Nabi et al., 2023). In addition, information can also be collected from subjects where they are required to do self-reporting of fatigue using some defined scales (Golz et al., 2010).

One possible source of data for driver fatigue detection is the visual data obtained from the ADAS which are used to monitor the behaviour of drivers. The ADAS is a driver safety innovation tool that is installed in vehicles which is aimed at improving the safety of drivers. It has a number of functionalities such as cameras which are used for monitoring the driver's behaviour by transmitting short video clips to the control room and proactively sending warning signals to drivers. Incorporating aspects of fatigue detection and monitoring within the functionalities of the ADAS has been explored in a number of studies. Ayachi et al. (2021) and Zhao et al. (2022) proposed fatigue detection systems for ADAS. A similar approach was adopted in this study since Namdeb has already implemented the ADAS which is currently used for monitoring driver behaviour.

2.2.2 Data Pre-processing

The second stage aims to perform pre-processing to the data obtained in the first stage and to extract fatigue features. The required pre-processing is dependent on the data acquisition method applied in stage one. In the case of physiological data, pre-processing is performed in order to reduce the noise in the data (El-Nabi et al., 2023). For image and video data, computer vision algorithms are required for human body detection.

Fatigue feature extraction

In order to optimise the accuracy of any model, it is important to ensure that the best possible features are extracted for use as input in the model. There are five approaches or methods of fatigue detection and they differ in terms of the input features used for fatigue detection.

Subjective reporting

Fatigue detection techniques can be subjective or objective. Subjective measures of fatigue require self-assessment for fatigue by using questionnaires which subjects can complete to assess their levels of fatigue (Golz et al., 2010). Various scales have been proposed for fatigue self-reporting and the Karolinska Sleepiness Scale is the most popular one. Fatigue measurements based on subjective reporting have been found to be more accurate and it is for this reason that they are commonly considered for the validation of fatigue detection models (Perkins et al., 2023). However, the main limitation of using subjective measures is that fatigue detection cannot be done in real time (Kumari et al., 2021; Zhao et al., 2022). Therefore, since real time detection is a key factor in the design of fatigue detection systems, objective measures are more preferred compared to subjective measures.

The three major approaches related to objective fatigue measures are behavioural based, vehicle based and physical based techniques. These are discussed in the following sections.

Behavioural based techniques

The behavioural based techniques, also referred to as camera based or computer vision based techniques, make use of visual data to determine facial movements and expressions which can be used to extract key metrics that are useful for measuring fatigue (Raja Mohana et al., 2021). The behavioural changes which are commonly monitored for fatigue are eye opening or closing, yawning and head nodding (Kumari et al., 2021).

Eye state

The eye closing for a period of time is one of the characteristics of fatigue. This is commonly measured by calculating the Eye Aspect Ratio (EAR), a distance measure of the extent of eye opening or closing. According to Federico (2022), the EAR values normally range between 0.0 and 0.38. However, higher values of EAR have been observed in some studies. The EAR can be used to calculate the percentage of eye closure (PERCLOS) time over a certain period of time as an indicator of prolonged eye closure (Golz et al., 2010; Raja Mohana et al., 2021). PERCLOS has proven to be the most suitable measure of fatigue based on eye condition (Zhao et al., 2022). According to Talebi et al. (2022), the PERCLOS is commonly used in miners fatigue monitoring systems. In addition, the blinking behaviour which is commonly described by blink frequency and blink duration (Karar & Kanumuri, 2023) has been considered for fatigue detection. For example, Islam et al. (2019) proposed a fatigue detection system using two methods, one based on eye closure time and another one based on eye blink rate. The EAR threshold used in their study was 0.3 and eye closure time to define fatigue was a period of more than 5 seconds while a blink rate of less than 6 per minute was used as an indication of fatigue. Using the thresholding method, they reported a detection accuracy of 92.5%. However, their system is not designed for real time fatigue detection.

Mouth state

The characteristics of the mouth used for driver fatigue detection are focused on mouth opening and yawning measures and they include mouth aspect ratio (MAR), mouth openness ratio (MOR), percentage of mouth opening (PMO), and frequency of mouth opening (FOM) (Fan et al., 2007). Jie et al. (2018) advise that mouth openness and yawning detection can be challenging in natural driving scenarios or when the driver covers their mouth during yawning. They propose a driver fatigue detection system that incorporates face touches in order to detect mouth covered yawns. Ayachi et al. (2021) suggest that the best way to make a distinction between yawning and talking is to consider the mouth openness measures together with the yawning frequency measures. They maintain that when the subject is talking, the mouth openness measure will be high but not consecutively while when the driver is yawning, the openness measure will be high for a while to indicate that the mouth has opened consecutively. They, however, did not consider the talking vs yawning aspect in their study. In addition, measuring yawning frequency can be a challenge when the duration of the videos is relatively short.

Nose tracking and head pose estimation

Zhang et al. (2015) suggest that the nose is a more robust measure when the driver is moving their head around which makes it difficult to capture the eye and mouth regions. On the other hand, Mounika et al. (2022) proposed that head pose estimation can be used in cases where it is difficult to detect the driver's eyes or mouth due to issues such as low intensity and lowlight areas. Thus, nose features and head pose estimation parameters are commonly used together to determine whether the head is bending. Under normal circumstances, the nose length ratio (NLR) is expected to range between 0.8 and 1 (Chinthalachervu et al., 2022). Head tilt/movement measures for fatigue were explored in Engineering et al. (2021) and Kumar and Patra (2018) while Yi et al. (2023) considered the Head Euler Angles (HEAs) as a measure of fatigue based on head pose.

Behavioural feature fusion

Due to the shortcomings of the individual behavioural features such as being prone to false detections, the use of multiple facial features has been proposed (Chen et al., 2021; (Chinthalachervu et al., 2022; Savas & Becerikli, 2018; Mounika et al., 2022; Zhu et al., 2022). Yi et al. (2023) proposed a single fatigue evaluation index which is a weighted average of the various feature parameters for the eyes, nose and mouth. The subjective determination of the weights can, however, be a potential limitation for their work. The facial features multi fusion approach has been found to improve the accuracy of the fatigue detection system and to minimise the misjudgement errors. Mounika et al. (2022) proposed a system that can detect fatigue and distraction using the EAR, MAR and head pose estimation. The performance of their system was, however, not evaluated and they also did not provide the threshold values used in their system. In Duan et al. (2023), deep neural networks were used to predict distractions, fatigue and potential hazards for drivers.

Prospects for behavioural based techniques

The behavioural based techniques have become popular due to the advancements in technology for capturing in-vehicle videos and these have been widely applied. They have an advantage in that they are less physically intrusive (Perkins et al., 2023), are cost effective (Hooda et al., 2022b) and are more portable systems (Kumar & Patra, 2018), though this may come at the cost of reduced accuracy.

Behavioural based techniques have a number of shortcomings which makes them to be less accurate and least robust compared to the physiological based techniques (Kumari et al., 2021). The miss detection of faces is the main challenge because these measurements are based on facial features. Junaedi and Akbar (2018) note that drivers tend to look sideways when talking and this can make it challenging to detect the face and calculate fatigue metrics. In addition, behavioural based measurements are affected by external factors such as lighting conditions and occlusions

(Kumari et al., 2021; Perkins et al., 2023). Other challenges include camera movements and the frame rate used to capture images (Ngxande et al., 2017); positioning of the face in the images or videos (Hooda et al., 2022b); noise information such as laughing and talking (Chen et al., 2022); variations in face appearances due to gender, race, facial hair, and face pose among others (Zhang et al., 2015), as well as face mis-detection due to an individual wearing glasses, having a moustache or a different skin tone (Perkins et al., 2023). Perkins et al. (2023) recommend the use of infra-red (IR) cameras in order to mitigate some of the challenges discussed above. In addition, different types of eye shapes such as round or narrow makes it a challenge when determining the threshold values for some metrics (Karar & Kanumuri, 2023).

There is also an issue with the acceptance of such systems due to some privacy concerns. While they are less physically intrusive compared to the other techniques, they are considered to be psychologically intrusive and there are privacy concerns around their implementation since they are based on video data where the identity of the individual can be determined (Perkins et al., 2023). It is, therefore, important to address the privacy concerns and other ethical issues in order to ensure user acceptance for such systems. Martin et al. (2014) proposed a deidentification filter which can be applied on video sequences in order to protect the identity of the individual(s) in the video while preserving information for the facial region which can be used to infer the behaviour and fatigue state.

Rohit et al. (2017) argue that behavioural based fatigue detection methods may not be suitable for the mining environment. This is due to tight space constraints for cameras and the cameras are also subject to occlusions and vibrations which can significantly influence the measurement of fatigue. They may also be difficult to correctly position for different drivers. Nevertheless, Zhao et al. (2022) still maintain that behavioural (or visual features) based fatigue detection methods have great prospects for developing real time fatigue detection products.

Face detection

The face is the region of interest for behavioural based fatigue detection systems. Thus, face detection is a key pre-processing step in the design of behavioural based fatigue detection systems. The core facial points that must be detected are the nose, mouth, left eye and right eye. Various computer vision algorithms such as the Haar Cascade, Viola-Jones, Histogram of Oriented Gradients and Support Vector Machine (HOG+SVM), and Multi-task Cascaded Convolutional Networks (MTCNN) have been proposed for face landmark detection. These facial landmark recognition technologies utilise ML algorithms and deep learning algorithms to model a human face (Mayilvahanan, 2020). Detection accuracy, robustness and computational complexity are some of the key aspects to consider when selecting a facial landmark detector considering that fatigue must be detected in real time and be suitable for real world driving videos where lighting conditions and occlusions as well as proximity to the camera are difficult to control. The most commonly used face detection method is the HOG+SVM detector which detects 68 facial landmarks used to identify key facial points such as the eyes, nose and mouth.

The HOG+SVM face detector has been found to be robust to different illumination, facial expressions and head positioning (Petrellis et al., 2021). Pothiraj et al. (2021) also found it to be robust to tilt in the head movement. However, K. Chen et al. (2021) advise that while the HOG+SVM face predictor is the commonly used detection method due to its high efficiency, it has some weaknesses particularly in that it is sometimes unable to detect the facial features when the face is partially occluded and it is found that the estimated facial landmarks coordinates can deviate from the real positions, thus providing inaccurate estimation for coordinates.

The Viola Jones is found to have a low face detection rate because it requires full view of the frontal face and does not detect face when tilted (Chen et al., 2021). Deep learning methods-based face detectors such as the Convolutional Neural Networks (CNN) have been found to offer improvements over the HOG+SVM, particularly for non-frontal faces at odd angles. However, it is computationally heavy, thereby making it unsuitable for real time fatigue detection (Pothiraj et al., 2021). In addition, it can only detect five facial points landmarks, the left eye pupil, the right eye pupil, the nose tip, left mouth corner and right mouth corner compared to the 68 facial points detected by the HOG+SVM detector. Yu-Hsuan et al. (n.d.) compared the performance of the MTCNN, CNN and Viola Jones. They found that the MTCNN was the best among the three, with the least number of face miss detections. However, the computational speed of the Viola-Jones as measured by the number of processed frames per second was the highest. Thus, it seems like one needs to strike a balance between computational speed and performance in terms of mis-detections.

Physiological based techniques

These techniques are based on human physiology related measurements (Perkins et al., 2023). Wearable devices such as smart caps are used to collect the data (Sabry et al., 2022). The physiological signals include the electroencephalography (EEG), the electrocardiograhy (ECG), the photoplethysmography (PPG), the electrooculography (EOG), the electromyography (EMG) and the heart rate variability (HRV) (El-Nabi et al., 2023). These signals are related to brain electrical acitivity, heart rhythm and rate, heart rate, electrical eye activity, muscle activity and heart beats, respectively. In most of the studies reviewed by El-Nabi et al. (2023) that used physiological based measurements, the EEG was the most frequently used signal for fatigue detection. Multi-physiological feature fusion has been proposed for improved performance (Sikander & Anwar, 2019).

Compared to other techniques, these methods are considered to be a more direct measure of fatigue (Sikander & Anwar, 2019). Overall, physiological based methods for fatigue detection have been found to be more efficient (Shalash, 2021) and the most reliable and accurate compared to the other measures (Kumari et al., 2021; Rohit et al., 2017). Hooda et al. (2022) compared the various physical and biological features used in fatigue detection and found that biological features such as the EEG, EOG and ECG signals achieve relatively high accuracy compared to the physical features such as facial behaviour, eye state and yawning detection.

Nonetheless, the physiological based measurements also have their limitations. Limited applicability to real time fatigue detection due to the time lag between collecting and processing signals has been cited as the major drawback of physiological based methods (Alharbey et al., 2022; Anwar, 2019; Lock et al., 2023; Sikander & Anwar, 2019). In addition, the data collection requires expensive sensors (Hooda et al., 2022; Shalash, 2021). Furthermore, they are intrusive, which means that they can affect the normal operations of the drivers (L. Chen et al., 2021) and this makes them less suitable for commercial products (Perkins et al., 2023). There may also be resistance in accepting them because they can be uncomfortable for individuals who have to wear the devices. High system cost and size have also been cited as some of the disadvantages of fatigue detection systems based on these measures (Kumar & Patra, 2018).

Chen et al. (2021) proposed ML based fatigue detection models for miners in high-altitude and cold environments based on a multi-fusion of physiological measurements. The optimal combination of physiological measurements in their study attained fatigue detection accuracy of 80%. Even though experimental data was used in the study, which may have some potential limitations, this suggests that there is potential for physiological based fatigue detection systems for the mining domain.

Vehicle based techniques

These techniques are based on vehicle movements which are used to infer the driver's behaviour. Compared to the other objective methods discussed above, it is the most simplest in terms of data collection and system implementation (Perkins et al., 2023). Sensors attached to the vehicle are used for data collection and the metrics used for measuring fatigue include lane position deviation, steering wheel movement (SP et al., 2021), brake pattern and vehicle speed (Kumar & Patra, 2018). These techniques have been proven to be viable for commercial products and have been used in the design of driver assistance technologies that are installed in vehicles manufactured by some well-known multinational automobile companies (Perkins et al., 2023; Sikander & Anwar, 2019; Zhao et al., 2022). Some of these system implementations are further discussed in Sowa (2022). However, Perkins et al. (2023) argue that steering position and lane movement may be due to non-fatigue related factors. Thus, vehicle based measurements for fatigue detection are found to be unreliable and have low accuracy (Chen et al., 2021; Kumari et al., 2021). In addition, due to their dependency on road characteristics and conditions, these techniques will only work in limited situations (Ngxande et al., 2017; Rohit, 2016) and thus they lack robustness in different environments (Lock et al., 2023). The mining environment may not be suitable for these techniques due to the surface mine terrains which would make lane maintenance tracking impossible (Rohit, 2016). In addition, steering wheel patterns may also be difficult to monitor for some vehicle types such as those found in the mines.

Hybrid techniques

Stancin et al. (2021) concluded that in order to design reliable fatigue detection systems with high generalisation ability, multiple features and hybrid techniques must be used. This can minimise the shortcoming of the individual techniques (Perkins et al., 2023) and increase the confidence of fatigue detection (Gargan, 2021). However, this approach comes at a cost of increased computational and system costs (Perkins et al., 2023) due to multiple data sources which would be required to implement such systems. Thus, Sikander and Anwar (2019) advise that it is important to select only the effective features and not to fuse all the available features.

2.2.3 Fatigue Detection and Classification

The third stage uses the fatigue features extracted in stage two to train classifiers that can be used to detect fatigue. Whereas the multiple class would be ideal to indicate different stages of fatigue, the binary classification is predominantly used in fatigue detection due to the scarcity of labelled datasets for non-binary classification (Brugman, 2022).

Hooda et al. (2022) classify fatigue detection methods into four categories namely mathematical models, rule-based implementation, ML and deep learning. The advancement in technology for data collection and analysis has resulted in a shift in focus to data driven algorithms i.e. ML and deep learning-based models for fatigue detection. Thus, the discussion here is only limited to ML and deep learning fatigue detection methods. However, since it has also been widely applied, a brief discussion on the thresholding approach is also included.

Thresholding

Thresholding is a rule-based classification approach for fatigue detection, mostly when using behavioural measurements. Using this approach, the fatigue state of a subject is determined by comparing their computed fatigue feature values to set threshold values. According to the systematic review by Perkins et al. (2023), it was the second most popular classification method in fatigue detection studies before 2018. In some studies, this approach has been used to check the correctness of the data labelling. Thresholding tends to be based on a trial and error approach, as there is no systematic approach to threshold determination. Kumar and Patra (2018) attempted to use threshold values computed experimentally from the observations. However, this approach will only work for videos with a longer duration since it requires a calibration phase. The videos used in their experiment were about 30 minutes long. The adaptive thresholding approach where the threshold values are individualised is discussed in Zheng et al. (2023). However, this approach may also not be feasible for real time fatigue detection.

Pothiraj et al. (2021) adopted the thresholding method in their study. Fatigue was detected when the EAR is below 0.28 for 40 consecutive frames. They argue that the threshold of 0.28 is a more approximate value for around 70% of the people. However, they do not report any performance metric of their system. Kumar and Patra (2018) and Chinthalachervu et al. (2022) both considered the thresholding approach before using ML classifiers for fatigue detection. The studies recognise that eye sizes are different for different people and this has implications for threshold determination/setting. Thus, subject specific thresholds were defined. However, they also do not report the performance metrics for their system. Kumari et al. (2021) proposed a fatigue detection system using the thresholding method. The threshold values used in their study were 0.3 for EAR and 20 for MAR/lip distance, for 30 consecutive frames to define fatigue. They, however, also do not report the evaluation performance of the system. Lock et al. (2023) used a deep learning method for face detection and using 2 623 images they classified the fatigue state using a threshold value of 20% for the PERCLOS. They reported an F1 score of 86.58%.

Dewi et al. (2022) also maintains that individuals have different eye sizes and advise recalculating EAR threshold values. This can, however, be a challenge for the real time fatigue detection system and also increases the computational time.

Traditional supervised ML models

Most of the work done on fatigue detection systems has been generated in controlled environments and the target is usually labelled. Thus, supervised ML classification algorithms have been used for fatigue detection. Supervised ML models are preferred because they are easier to interpret compared to deep learning models (Perkins et al., 2023). Some of the supervised ML algorithms used for fatigue detection include the Support Vector Machine (SVM), the Linear Discriminant Analysis (LDA), k-Nearest Neighbours (KNN) and the Random Forest (RF) (Perkins et al., 2023). In a review of 130 fatigue detection systems done by Perkins et al. (2023), the SVM was found to be the most commonly used traditional ML algorithm/classifier for fatigue detection. It generally outperforms other supervised ML classifiers.

Rohit et al. (2017) proposed a fatigue detection system using the spectral analysis of EEG signals for possible use in mine vehicles. The study used experimental data for 23 subjects collected using in-lab driving simulator using lightweight EEG sensors to measure brain activity. They extracted 12 features and compared the performance of the LDA and the SVM classifiers. They found that SVM outperformed the LDA, with a cross-subject validation accuracy of 87%. In addition, they compared the SVM classifier using the EEG based features and the one using the blink characteristics. Their findings suggest that using physiological measurements yields higher performance compared to behavioural measurements. In another study, Savas and Becerikli (2018) reported a 97.93% accuracy for a behavioural based driver fatigue prediction using the SVM classifier. Chen et al. (2021) proposed an ML based fatigue detection model for miners in high-altitude and cold environments based on multi-fusion of physiological measurements. They used experimental data for a sample of 45 randomly selected miners. Five physiological measurements were considered and the SVM and RF classifiers were used for fatigue detection. They reported that the best combination of physiological measurements provided an accuracy of 80% for the SVM classifier. In a study by Chinthalachervu et al. (2022), a sensitivity score of 96% and a 100% specificity score was reported when using the SVM classifier trained on multiple behavioural features such as the EAR, MOR, and NLR. Kumar and Patra (2018) proposed a low-cost fatigue detection system for drivers based on three behavioural features, EAR, NLR and MOR. They used simulated data consisting of 30 minutes videos and tested the system on video data for six drivers. They used three classifiers, Bayesian classifier, Fisher's LDA and the linear SVM. The SVM classifier outperformed the other two classifiers with an accuracy score of 95.8%. Savas and and Becerikli (2018) developed a behavioural based fatigue detection model using the SVM classifier. They used multiple facial features such as the PERCLOS, yawn count, blink count, internal zone of the eye opening and head detection and achieved an accuracy of 97.93%. In another study, Darbandy et al. (2020) predicted the physical fatigue for employees in the manufacturing and construction sectors using features extracted from heart signals and the KNN classifier. Their model achieved an accuracy of 78%.

While most fatigue detection studies have used individual ML algorithms, ensemble ML algorithms have also been considered in some work in order to improve the performance of the individual classifiers. Kumari et al. (2021) used various ensemble models such as the gradient boost, Adaboost and RF. Bustos et al. (2023) found that Gradient boosted trees performed well compared to single classifiers. Using physiological features, Ramos et al. (2022) developed ensemble models for fatigue detection. Other fatigue detection studies that considered ensemble models include Sedighi et al. (2020).

Deep learning

In a review of fatigue detection systems done by Perkins et al. (2023), they found that there has been a transition from traditional supervised ML fatigue detection models to deep learning models. Artificial Neural Networks (ANN), the Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) and various modified neural networks have been used for fatigue feature extraction and detection. The deep learning models have also been trained with a wide range of feature parameters including behavioural based features, biological features and vehicle based features (Sikander & Anwar, 2019).

Ayachi et al. (2021) proposed a fatigue detection system for ADAS using the CNN to detect and classify driver fatigue based on the PMO and PERCLOS. The proposed system was evaluated on the NTHU-DDD and achieved an accuracy rate of 96.05%. Alharbey et al. (2022) used a combination of ML and deep learning models to develop models to detect drivers' fatigue. Using physiological measurements, the SVM outperformed other traditional ML classifiers, with an accuracy above 90%. The CNN was used for detecting fatigue using video data and it provided 99% detection accuracy. Raja Mohana et al. (2021) used the Multi-task Cascaded CNN (MTCNN) model for face detection and feature extraction to measure the PERCLOS and POM. Their model achieved an accuracy over 97%. Lock et al. (2023) proposed a YOLOv2 model which uses the deep learning method and based on the PERCLOS metric based on images and they achieved an accuracy of 87%.

Overall, deep learning methods have been found to outperform traditional ML methods regardless of the measurements used (Nasirzadeh et al., 2020; Stancin et al., 2021; Shalash, 2021). In addition, deep learning models are also preferred because of their ability to jointly undertake pre-processing, feature extraction, feature selection and fatigue classification (Perkins et al., 2023). In a review of 25 studies on behavioural based drowsiness detection systems, Ngxande et al. (2017) found that the performance of the CNNs was statistically significant higher compared to the performance of SVMs.

However, deep learning models require more computational resources compared to the traditional ML algorithms and they are less interpretable compared to traditional ML based models (Perkins et al., 2023). Additionally, Ayachi et al. (2021) point out that deep learning models require large datasets for training and thus more computational resources. Perkins et al. (2023) recommend the use of lightweight deep learning models as low cost and real time fatigue detection models.

Unsupervised machine learning models

While the fatigue detection literature is dominated by the use of supervised ML algorithms, the fact that most real-life data are unlabelled means that it is also important to explore unsupervised ML algorithms. Balaskas and Siozios (2021) maintain that supervised ML algorithms are limited in the sense that they require labelled data which may be rare in some of the real-world data sets. This is particularly one area which is not well explored in the fatigue detection literature considering that there is lack of consensus on the labelling of the subject's fatigue state (Perkins et al., 2023). Since the labelling of the subject's fatigue state has a significant influence on the accuracy of the system, inconsistency in fatigue labelling has been cited as one of the key challenges facing the development and design of accurate fatigue prediction models. Ekener and Ekenstam (2023) point out that it's time consuming, as well as difficult and expensive to collect high quality labelled fatigue detection data which can be used for training supervised ML models for fatigue detection. It is therefore not surprising that most of the studies in the literature tend to rely on experimental and simulated data for the design of fatigue detection systems.

However, there are a few studies that used unsupervised ML algorithms for fatigue detection. Using the K-means algorithm to determine the blink threshold, Chao et al. (2019) reported an accuracy score of 92.5%. Gurudath and Riley (2014) argue that in real world driving scenarios, it is difficult to know the fatigue class to which a particular subject belongs. They used the K-means to classify driver fatigue using features derived from the EEG signals.

2.3 Chapter Summary

Fatigue detection is an active area of research and the application of ML in developing fatigue detection systems has been explored in different domains such as the transport and construction industry, with a particular focus on driver fatigue. Various fatigue detection techniques using different fatigue predictor features have been explored such as the behavioural based, physiological based and vehicle-based techniques. The choice of the fatigue detection technique is dependent on the data acquisition method which can be visual data, physiological signals data or vehicle monitoring data. Fatigue features extracted from biological/physiological signals have been found to be more accurate. However, the intrusive nature of such techniques makes them less suitable for fatigue detection. Behavioural based techniques using features extracted from facial characteristics have been predominantly used. Deep learning models have been found to outperform the traditional ML algorithms when developing fatigue detection systems. However, the former requires a lot on computational resources and are less suitable for real time fatigue detection.

Despite the increasing number of studies on fatigue detection, there are still some issues that have not been addressed in practical situations. Specifically, the subjectivity and lack of coherence in the labelling of fatigue has been identified as one of the key issues. In addition, most of the systems are evaluated on simulated data which may not capture some of the realities of real world driving in environments such as the mining industry environment.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the research methodological approach that was used to answer the research questions and achieve the research objectives of the study, in line with Saunders et al.'s (2019) research onion model. This layered model helps to describe the different decisions in terms of the data required, the data collection techniques and the data analysis procedures. The research onion model with the methodological choices used in this study are summarised in Figure 3-1 and discussed in detail in the following sections of this chapter.

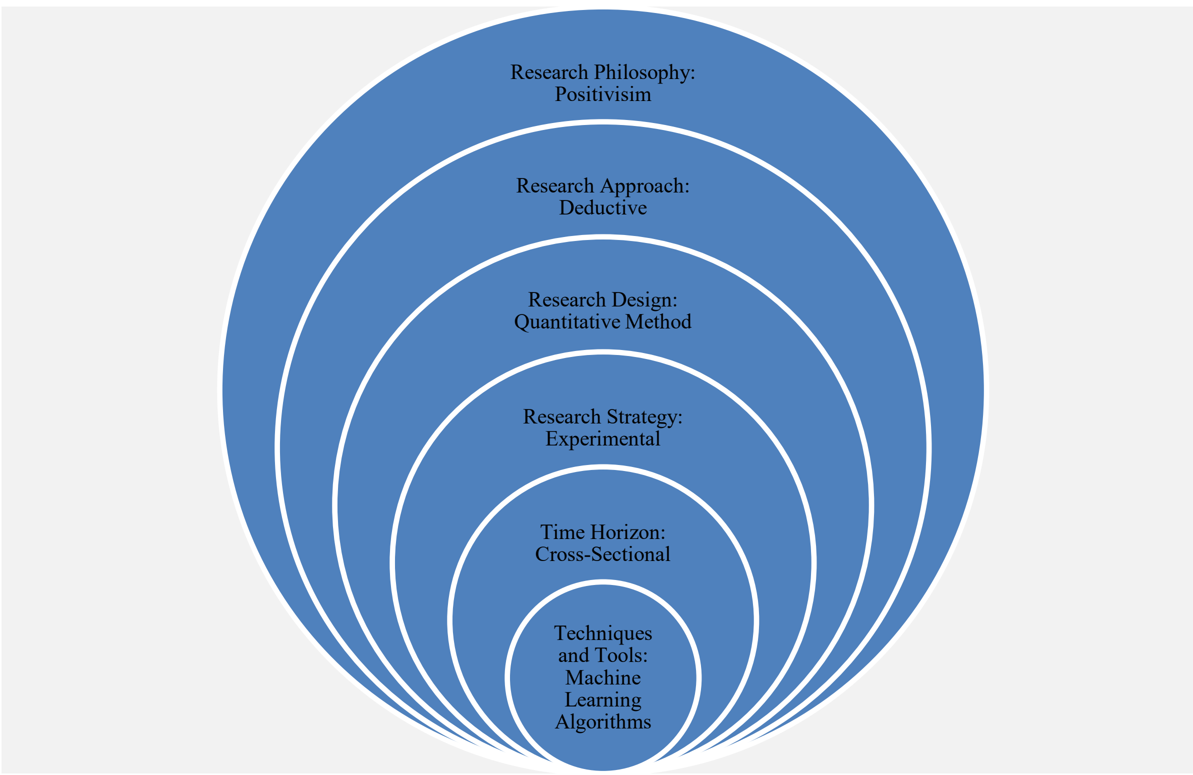


Figure 3-1: Research methodology

3.2 Research Philosophy

The first layer of Saunders et al.'s (2019) research onion is the research philosophy which relates to the development of knowledge and the researcher's assumptions about the nature of that knowledge and how the world works. The four research philosophies are pragmatism, positivism, realism and interpretivism. Saunders et al. (2019) present a comparison of the four research philosophies in terms of how they operate on ontological, epistemological and axiological assumptions and by referring to the data collection techniques used.

In this study, explanatory and predictive analysis were conducted to detect and classify employee fatigue state. With regards to the ontological view, there is only one reality and it can be measured objectively and independently. Epistemologically, the study collected data and analysed it to identify causal relationships. From an axiological point of view, the researcher was independent of the data and maintained objectivity in the data analysis. In addition, a combination of structured and unstructured data were used for this study. Thus, a positivism view for research was appropriate for this study.

3.3 Research Approach

The three main approaches to theory development in research are inductive, deductive and abductive. According to Saunders et al. (2019), research approaches are based on the reasoning adopted in research. In this study, the research approach was deductive. In deductive approach, a set of known premises are used to logically derive the conclusion (Saunders et al., 2019). This approach is usually associated with the positivism research philosophy. In addition, Saunders et al. (2019) explain that a key characteristic of this research approach is that it aims to explain causal relationships between concepts and variables which are usually measured quantitatively. This approach was, therefore, a good choice for this study since the research builds on existing literature on fatigue detection and used mining operational data to detect employee fatigue.

3.4 Research Strategy

A research strategy can be defined as a researcher's plan of action to answer the research questions (Saunders et al., 2019). There are several research strategies and Saunders et al. (2019) guide that the experiment and survey strategies are the ones that are usually linked to the quantitative research design. They further explain that the experimental strategy is used to determine causal relationships between independent variables and a dependent variable. Thus, the experimental strategy was well suited for this study since independent/predictor variables/features were used to predict employees' fatigue state, the dependent variable.

3.5 Data Collection Methods

The next layer of the research onion is about how the data used to achieve the research objectives and answer the research questions is gathered. The data collection method applied was a multiple source cross-sectional secondary data. The study used video data selected from the open source Yawning Detection Dataset (YawDD) (Abtahi et al., 2014). This dataset contains over 300 videos of individuals in different states including yawning, normal, talking, talking/yawning/singing and with some drivers wearing glasses or sunglasses. The videos were recorded using in-car camera and were taken under varying illumination conditions. In this study, a subset of videos was selected from the YawDD. The data consisted of 90 subjects, 43 females and 47 males under different conditions. In total, the selected dataset consisted of 207 videos, with 89 videos of subjects in the normal/non-yawning state, 96 videos of subjects in the yawning state and 22 videos of subjects talking. Additionally, subjects in 117 videos were not wearing glasses whereas in 68 and 22 videos, the subjects were wearing glasses and sunglasses, respectively. In addition, a dataset consisting of 46 videos sourced from the Namdeb ADAS was used to test the performance of the fatigue detection system when using real world driving data. The dataset consisted of drivers who showed drowsiness-related driving behaviour. They were recorded under different real-world driving conditions in a mining environment. Some of the key similarities and differences between the two datasets are summarised in Table 3-1.

Table 3-1: Similarities and differences between YawDD and Namdeb datasets

	YawDD	**Namdeb**
Number of videos	207	46
Recording set/environment	Simulated data in stationary vehicle.	Real world driving video data in the mining environment.
Duration	12 to 25 seconds	12 seconds
Image sizes and distance to camera	Relatively large and close to camera.	In most videos the image sizes are relatively small and far from the camera.
Data labelling	Labelled as yawning, normal or talking.	Data not labelled.
Lighting conditions	Most videos recorded in daylight.	Some videos recorded at night.

The YawDD has been widely used in fatigue detection studies. However, the data was captured in a static environment which means that it can present a challenge for designing systems that are able to detect fatigue in real world driving. Additionally, non-spontaneous and unnatural yawns is one of the criticisms levelled against the use of the data (Chen et al., 2021; Jie et al., 2018). Hence the need to complement the YawDD with additional data as advised by Chen et al. (2021).

The YawDD was used to achieve sub-research objectives 1, 2 and 3 while sub-objective 4 was achieved using the Namdeb data.

3.6 Techniques and Tools

This study deployed various ML algorithms, including supervised and unsupervised ML algorithms. The purpose of using the different algorithms was to determine which one achieves the best performance.

Supervised ML classifiers include Support Vector Machine (SVM), Logistic Regression (LR), Gaussian Naïve Bayes (GNB), k-Nearest Neighbours (KNN) and Decision Trees (DT). In the fatigue detection literature, the most frequently used classifiers are the SVM and KNN. In addition, one ensemble model, the Random Forest (RF), was considered because ensemble models improve on the performance of individual ML classifiers.

Table 3-2: Brief descriptions of the ML algorithms

ML algorithm	Description
SVM	The SVM is a commonly used ML algorithm where the classification is grounded on the concept of decision planes to define the boundaries separating the various classes (Kantardzic, 2011). It can be used for binary or multi class classification. The decision function can use different kernel functions such the linear, polynomial, sigmoid and radical basis functions thus allowing for separation in complex boundaries. It has been widely and successfully used in fatigue detection studies and its performance compare relatively well with deep learning models.
LR	LR algorithm can be used to model the probability of an event occurring as a linear function of predictor variables (Kantardzic, 2011). For example, it can be used to predict the likelihood that an employee is fatigued. If the predicted probability is above a set threshold i.e. 0.5, the model predicts a fatigue event, otherwise it predicts a non-fatigue event. It was outperformed by the DT and RF in a fatigue detection study by Saputra and Purwitasari (2022).

DT	DT is a tree based algorithm which is based on the logical classification models methodology (Kantardzic, 2011). The prediction of the target class is performed by using decision rules in a classification tree analysis. The DT algorithm has been used in a number of fatigue detection studies.
GNB	The GNB classifier is a classification algorithm based on the Bayes theorem. It assumes that the predictor variables are independent, given the target class. Thus, the class prediction is based on conditional probabilities. In most fatigue studies where the GNB has been used, it has not achieved accuracy scores above 90%.
KNN	The KNN is a simple but computationally intensive algorithm, especially when the size of the training set is relatively large (Kantardzic, 2011). Classification is based on proximity, where the class label of a new data point is based on the majority class labels of its nearest k data points. It has been used in several fatigue detection studies and has achieved relatively good performance.
RF	RF is an ensemble of DT classifiers where the final classification is based on averaging the predictions of a combination of a number of randomised DT classifiers (Kantardzic, 2011). The ensemble method improves the classification accuracy and reduces the variance found in the individual classifiers (Saputra & Purwitasari, 2022). Thus, ensemble models are known to be more stable and reliable and provide better results compared to individual classifiers.
K-Means	The K-means algorithm is an iterative unsupervised ML algorithm commonly used for clustering problems, especially when working with unlabelled data. The feature points are clustered based on their proximity to the centroids of the different clusters (Chao et al., 2019), where proximity is based on a distance measure. The algorithm aims to minimise the within cluster sum of squares (Gurudath & Riley, 2014).

Unsupervised ML algorithms are used in situations where the target is unlabelled. Unsupervised ML algorithms seek to find the best separated clusters in the data. The most commonly used clustering algorithm is the K-Means. Clustering has rarely been used in the fatigue detection literature. However, owing to the challenge of fatigue labelling, it is considered in this study in order to compare its performance to supervised ML algorithms.

3.7 Ethical Considerations

This research was approved by the Namibia University of Science and Technology (NUST) Faculty of Computing and Informatics research ethics committee (registration number: FREC-04/23) (see Appendix A). In addition, a consent letter and non-disclosure agreement were signed with Namdeb for permission to use their operational data for the purpose of this study (see Appendix B). All NUST ethical requirements and guidelines as well as Namdeb's permission letter conditions were observed and adhered to.

3.8 Chapter Summary

This chapter discussed the methodological approach adopted in this study and the justification for the choices made. The research work followed a positivism philosophy by employing a deductive approach. A quantitative research design was adopted using an experiment strategy and secondary cross-sectional data. The proposed system was an ML based system employing various ML algorithms. The ethical considerations for this study were also highlighted. The next chapter presents a description of the proposed system.

CHAPTER FOUR: PROPOSED FATIGUED DETECTION SYSTEM

4.1 Introduction

In line with the fatigue detection systems general workflow by El-Nabi et al. (2023) as presented in Figure 2-1, the architecture of the proposed fatigue detection system for Namdeb employees is presented in Figure 4-1 and the details of the specific stages are outlined in sections 4.2 through 4.5.

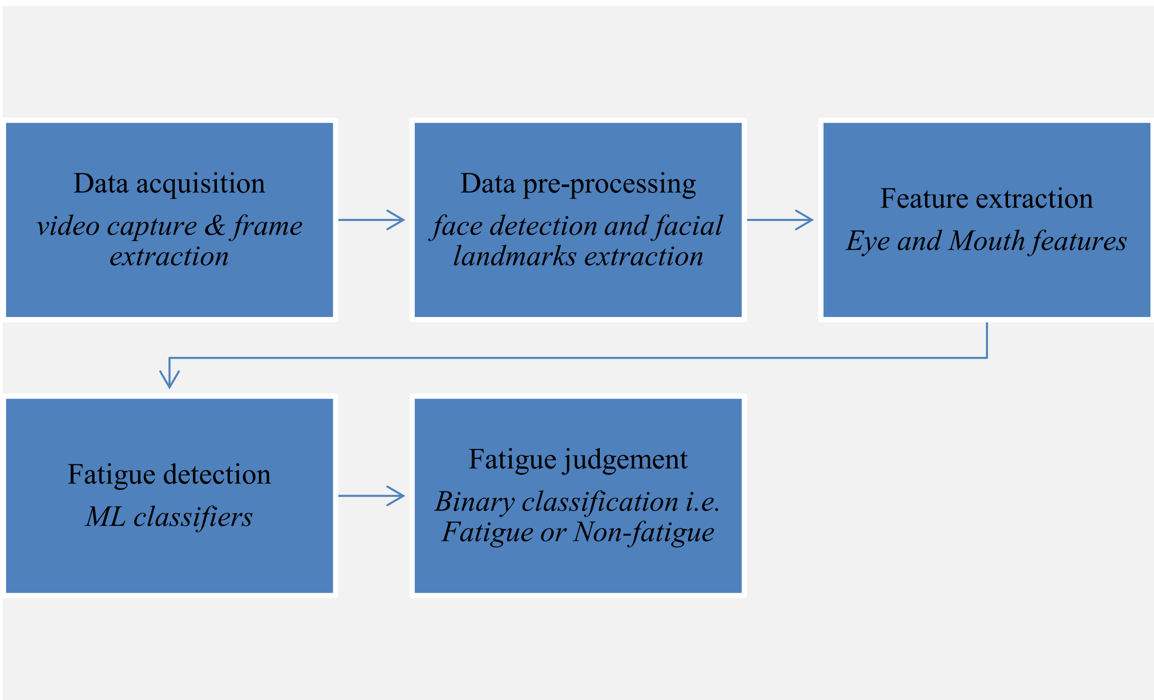


Figure 4-1: Architecture for the proposed fatigue detection system

4.2 Video Data Acquisition

The proposed fatigue detection system is based on behavioural measurements using video data. Thus, the proposed system begins by capturing a pre-recorded video but can easily be adopted for capturing a video from a webcam in order to allow for real time fatigue detection. The videos are captured using the Python OpenCV library which can perform image recognition, among others. Once the video is captured, the video frames are extracted, frame by frame and the image is converted into a two-dimensional array.

The proposed system was trained and evaluated on the YawDD open source video data. Thereafter, the pre-trained model was used for fatigue prediction using the Namdeb video data from the existing ADAS.

4.3 Pre-processing

At the second stage, the frames are transformed to grayscale images and the HOG+SVM detector which is offered by the Dlib Python library is then used to detect the faces in the extracted frames. The face is the region of interest since fatigue detection is based on the facial characteristics/expressions. The HOG+SVM detector describes the entire frame using a HOG, and it uses the SVM to classify any component that is found to be a face or non-face and thereafter, the face box is identified by using appropriate combination of all the classified components as face (Federico, 2022). Once a face has been detected, the HOG+SVM predictor is then used to identify 68 facial landmarks as shown in Figure 4-2.

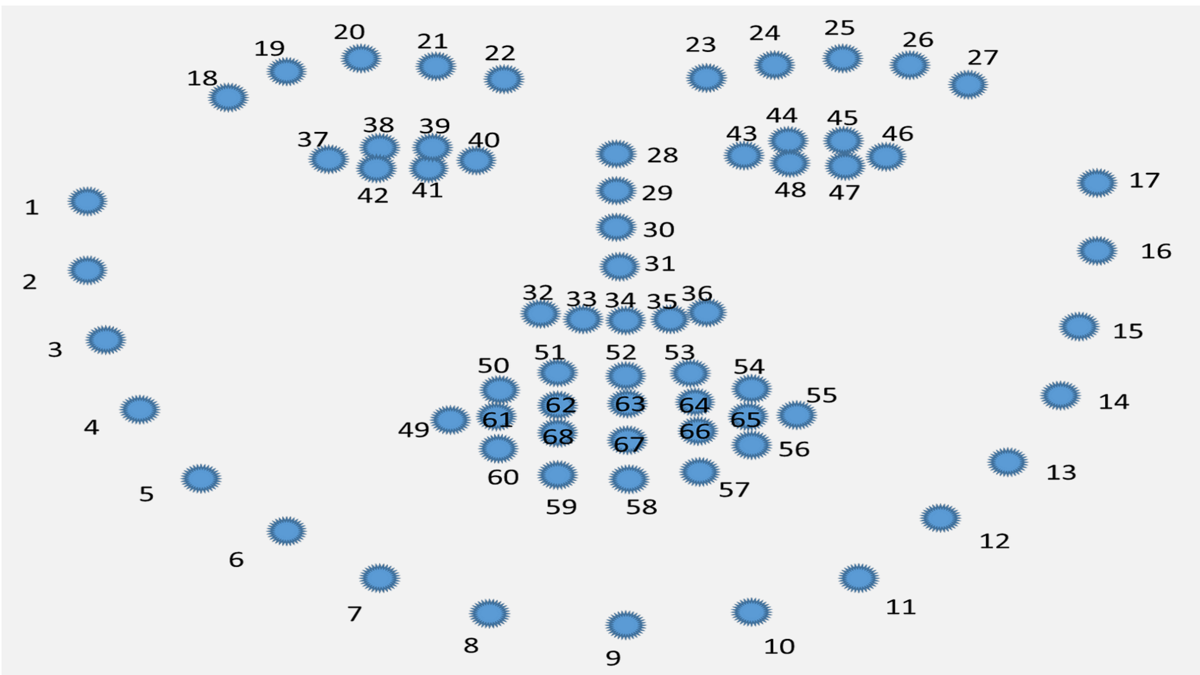


Figure 4-2: Dlib HOG+SVM 68 facial landmarks

Source: Federico (2022)

The Dlib is just one of the face detection solutions proposed in the literature. This face detection algorithm has been used in numerous studies on fatigue detection. Its advantage is that it is able to detect more facial landmarks compared to the CNN detector which can only detect five facial landmark points. However, it has its weakness in that it does not perform well when the images in the videos are small and is found to have a large number of face mis-detections in such scenarios.

4.4 Feature Extraction

Based on the literature presented in Chapter two, the eye condition and mouth state have been proven to be good indicators of fatigue based on facial characteristics. In this study, facial fatigue features, eye and mouth, were considered for fatigue detection.

4.4.1 Eye Features

The EAR is the basis of the eye features considered in this study. The EAR is a distance measure that is used to determine the eye openness state. It is calculated as the ratio of the distance between the horizontal and vertical landmarks for the eye. It is calculated for each facial frame using the Euclidean distance which is available in the Python SciPy package. Since both eyes blink at the same time, the EAR is an average of the EARs for the left eye and right eye. In some studies, for example Islam et al. (2019), only the EAR of one eye is used in order to reduce computation resources since it is believed that both eyes should be in the same state at any given point. This approach can also be suitable in scenarios where the angle of the camera is only able to capture one eye. The following is the EAR calculation formula:

$$EAR = \frac{EAR_{left\ eye} + EAR_{right\ eye}}{2} \quad (1)$$

$$where\ EAR_{left\ eye} = \frac{|p_{38} - p_{42}| + |p_{39} - p_{41}|}{2|p_{37} - p_{40}|} \ , \ EAR_{right\ eye} = \frac{|p_{44} - p_{48}| + |p_{45} - p_{47}|}{2|p_{43} - p_{46}|}$$

and the p_i are the facial landmarks for the eyes as shown in Figure 4-2.

When the eyes are completely closed, the EAR value will be zero or very close to zero. However, there is no clearly defined upper limit for the EAR. For example, Federico (2022) suggests that EAR values range between 0.0 and 0.38, but in this study and other studies, EAR values above 0.38 were found.

The EAR values are then used to compute the PERCLOS which measures the proportion of time the eyes are closed within a given period. This can be calculated by determining the proportion of times the EAR is below a certain threshold. The determination of the threshold value is a topic of contention because it depends on the shape of the subject's eyes (Feredico, 2022). A threshold value of 0.2 was used in this study and the PERCLOS is calculated using the following formula:

$$PERCLOS = \frac{Number\ of\ frames\ with\ EAR\ values\ below\ threshold}{Total\ number\ of\ facial\ frames} x100 \quad (2)$$

4.4.2 Mouth Features

The opening and closing of the mouth is measured to determine whether the subject is yawning and to assess the degree of fatigue. The measure for the level of openness of the mouth used in this study is the MAR. It is derived using the ratio of the distance between the mouth's horizontal and vertical landmarks. It is calculated for each facial frame using the Euclidean distance between the upper lip and the lower lip. The following is the MAR calculation formula:

$$MAR = \frac{|p_{50}-p_{60}|+|p_{51}-p_{59}|+|p_{52}-p_{58}|+|p_{53}-p_{57}|+|p_{54}-p_{56}|}{5|p_{49}-p_{55}|} \qquad (3)$$

where p_i are the facial landmarks for the mouth.

The value of the MAR increases when the mouth is opened for a yawn (Rahul et al., 2022) but there is no defined range for the MAR and different studies used different threshold values. In this study, the threshold value for MAR was set at 0.60 and the PMO is then calculated as:

$$PMO = \frac{Number\ of\ frames\ with\ MAR\ values\ above\ threshold}{Total\ number\ of\ facial\ frames} x100 \qquad (4)$$

4.4.3 Thresholding and Data Labelling

Similar to Sowa (2022), the thresholding approach was used in this study to check the correctness of the data labelling for the YawDD and to define labels for the Namdeb dataset using the computed fatigue features. The re-labelling of the YawDD and labelling of th Namdeb dataset was performed using the conditions set out in equation (5).

Fatigue state:

$$\begin{cases} PERCLOS < PERCLOS_{THRESHOLD}\ \boldsymbol{OR}\ PMO > PMO_{THRESHOLD} = FATIGUE \\ PERCLOS \geq PERCLOS_{THRESHOLD}\ \boldsymbol{OR}\ PMO \leq PMO_{THRESHOLD} = NON-FATIGUE \end{cases} \qquad (5)$$

This pre-processing was necessary because the yawns in the YawDD are non-spontaneous, which means that some of them did not qualify to be classified as fatigue yawns. Additionally, the re-labelling was necessary in order to deal with the videos which were initially labelled as talking. For the Namdeb dataset which initially did not have labels, it was necessary in order to be able to apply supervised ML algorithms. The final labelling consisted of only two classes, fatigue or non-fatigue.

4.4.4 Feature Selection

Statistical analysis of the extracted features is important to ensure that the selection of the best features to use in training fatigue detection models and to maximise accuracy of the model. In addition, feature selection is important to reduce computational and system costs (Perkins et al., 2023). Some of the approaches to feature selection in fatigue detection studies include the Student's t-tests of independence, analysis of variance and correlation analysis (Perkins et al., 2023). Student's t-tests can be performed to determine whether the mean differences are statistically significant for the different fatigue classes. This allows for checking whether there is a statistically significant difference between the means for the fatigue versus the non-fatigue. This test was used in the study by Rohit et al. (2017) using EEG based features. They found that 11 features had statistically significant different means (at 5% level of significance) and only one feature did not. Kumar and Patra (2018) also used the same approach using behavioural features and they were all found to be statistically significant at 5% level of significance.

Correlation analysis can also be performed on the features extracted. This is particularly important when working with large feature spaces where some of them are correlated. In cases where some features are highly correlated, data reduction techniques such as the principal component analysis (PCA) can be used. This approach was used by Kumar and Patra (2018). In this study, the tests of independence and correlation analysis are performed on the features before the classifiers are trained.

In this study, feature selection was based on the Mann-Whitney U Test non-parametric test of independence and the Pearson's correlational analysis. Data reduction techniques were not necessary since the feature space is relatively small.

4.5 Fatigue Detection

A fatigue detection model using visual behaviour measurements and ML algorithms is trained based on the data described in section 3.5. The target variable is a binary variable, therefore, a Boolean model that outputs either 0 for the non-fatigue class or 1 for the fatigue class was the obvious choice. The model's performance is evaluated based on its ability to correctly classify the target variable class.

4.5.1 Fatigue Detection Algorithms

Binary classification was used in this study for fatigue detection. The study deployed a combination of individual popular supervised ML classifiers (DT, GNB, LR, KNN, and SVM), along with one ensemble algorithm (RF) and one unsupervised ML algorithm (k-Means). In this study, the number of clusters for the k-Means is defined and set to two since the training data is labelled with two classes. This was to allow for performance comparison with other algorithms. However, since the k-Means is designed to find patterns in the data by itself, specifying the number of clusters may affect its performance.

4.5.2 Performance Evaluation

Model evaluation is a critical step in the process of developing ML models. It is aimed at assessing the model's performance, reliability and generalisability. There are a number of ML model evaluation techniques. The three that were used in this study are cross validation, hyperparameter tuning and the confusion matrix evaluation metrics.

Cross Validation

Cross validation is a technique that trains multiple models on subsets of the input data and evaluates the models on a subset that was not used in training. There are various cross validation techniques. In this study, a Stratified Shuffle Split cross validation technique was used where the number of splits were set to 5 and the test size was set to 20%. The implementation was done using the Python Sklearn library. This cross-validation technique ensures an even distribution of target class labels between the test and train sets.

Hyperparameter Tuning

Hyperparameters are machine learning model parameters which cannot be directly learned from the data but are important for improving the model performance, complexity and learning rate. These parameters must be specified before the models are trained. In this study, the grid search technique was used for parameter hyper tuning in order to optimise the performance of the models. The implementation was done using the Python Sklearn Grid Search Cross Validation (GridSearchCV) class.

Confusion Matrix Evaluation Metrics

The performance evaluation measures commonly used in classification problems are based on the confusion matrix. The confusion matrix compares the actual classification with the ML algorithm's predictions. For binary classification, four values are obtained from the confusion matrix. The four values are defined as True Positive (TP) when a fatigue case is correctly detected; False Positive

(FP) when a non-fatigue case is incorrectly detected as fatigue; False Negative (FN) when a fatigue case is incorrectly defined as non-fatigue; True Negative (TN) when a non-fatigue case is correctly predicted. These values are used in computing the performance metrics as defined in equations (6) to (9).

Accuracy measures the total number of cases correctly classified as either fatigue or non-fatigue. It is calculated using the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

Recall measures the proportion of fatigue cases that are correctly predicted as fatigue and is calculated using the following formula:

$$Recall = Sensitivity = \frac{TP}{TP+FN} \tag{7}$$

From the pool of all cases that are predicted to be fatigued, Precision finds the proportion of those that are actually fatigued. This is calculated using the following formula:

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

The F1 score is a measure of how well the model is performing by combining the precision and recall. It is calculated using the following formula:

$$F1 - Score = 2x \frac{Precision*Recall}{Precision+Recall} \tag{9}$$

In addition to the performance metrics discussed above, the Area Under the receiver operating characteristic curve (AUC) was also used to compare the performance of the classifiers.

4.6 Chapter Summary

This chapter detailed the key stages of the proposed fatigue detection system in line with the fatigue detection systems general workflow discussed in the literature. The proposed fatigue detection system is a behavioural based model using visual data. The Dlib HOG face detector was used for face detection and to extract the 68 facial landmark points which are used for facial feature extraction. Fatigue judgement was based on the PERCLOS and PMO. The thresholding approach was used before model training in order to validate the labelling of the YawDD. The fatigue detection system deployed a number of ML algorithms and their performances was evaluated using various model evaluation techniques such as cross validation, hyperparameter tuning and confusion matrix metrics. The next chapter discusses the implementation results.

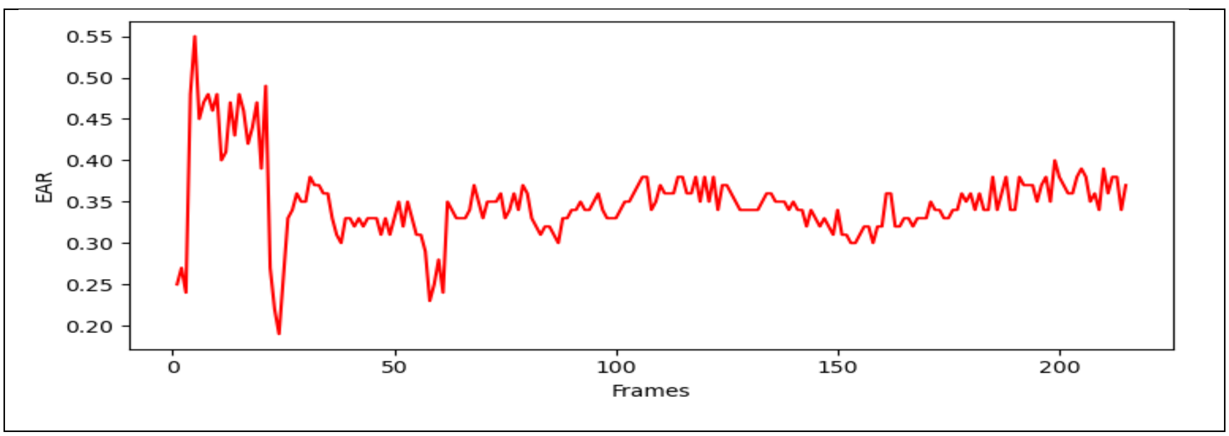
CHAPTER FIVE: IMPLEMENTATION AND EVALUATION

5.1 Introduction

The implementation details of the proposed system and its performance are discussed in this Chapter. The implementation was done in Python version 3 using various libraries and packages.

5.2 Thresholding and Data Labelling

The thresholding method was used to ensure the correctness of the target labels for the YawDD and to label the Namdeb dataset. The initial labelling for the YawDD included 89 videos labelled normal, 96 videos labelled as yawning and 22 labelled as talking. The conditions defined in equation (5) were used to validate the target class labelling of the YawDD. This resulted in a re-labelled data with 152 labelled non-fatigue and 55 labelled as fatigue. Thus, 44 of the videos that were initially labelled as yawning were in actual fact non-fatigue yawning. One example is shown in Figure 5-1 which shows the trend in the EAR and MAR values. Figure 5-2 shows the trend of EAR and MAR values for a video with a correct yawning/fatigue labelling while Figure 5-3 depicts a trend of EAR and MAR values for a video initially labelled as talking and re-labelled as non-fatigue.



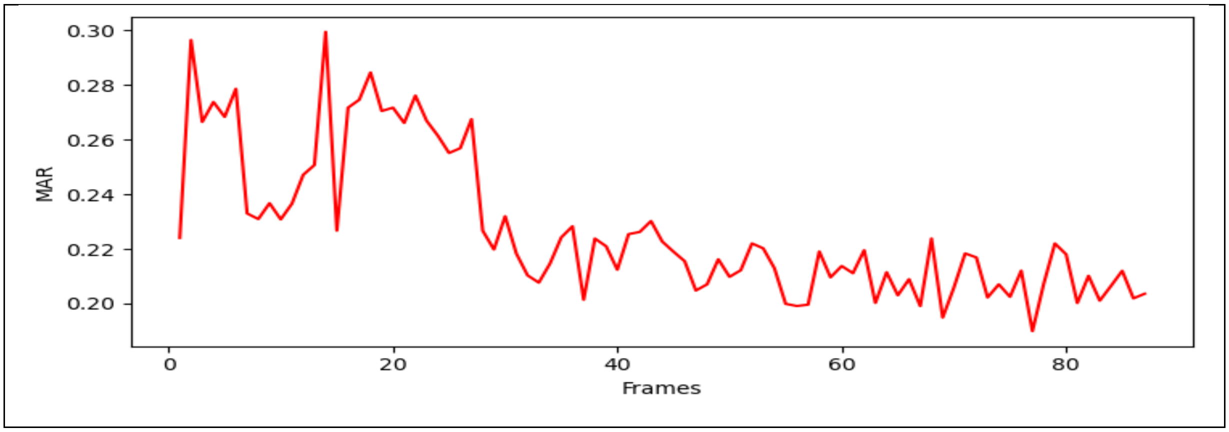


Figure 5-1: Plot of EAR and MAR values for 9-Male-NoGlasses-Yawning video. The video was re-labelled non-fatigue

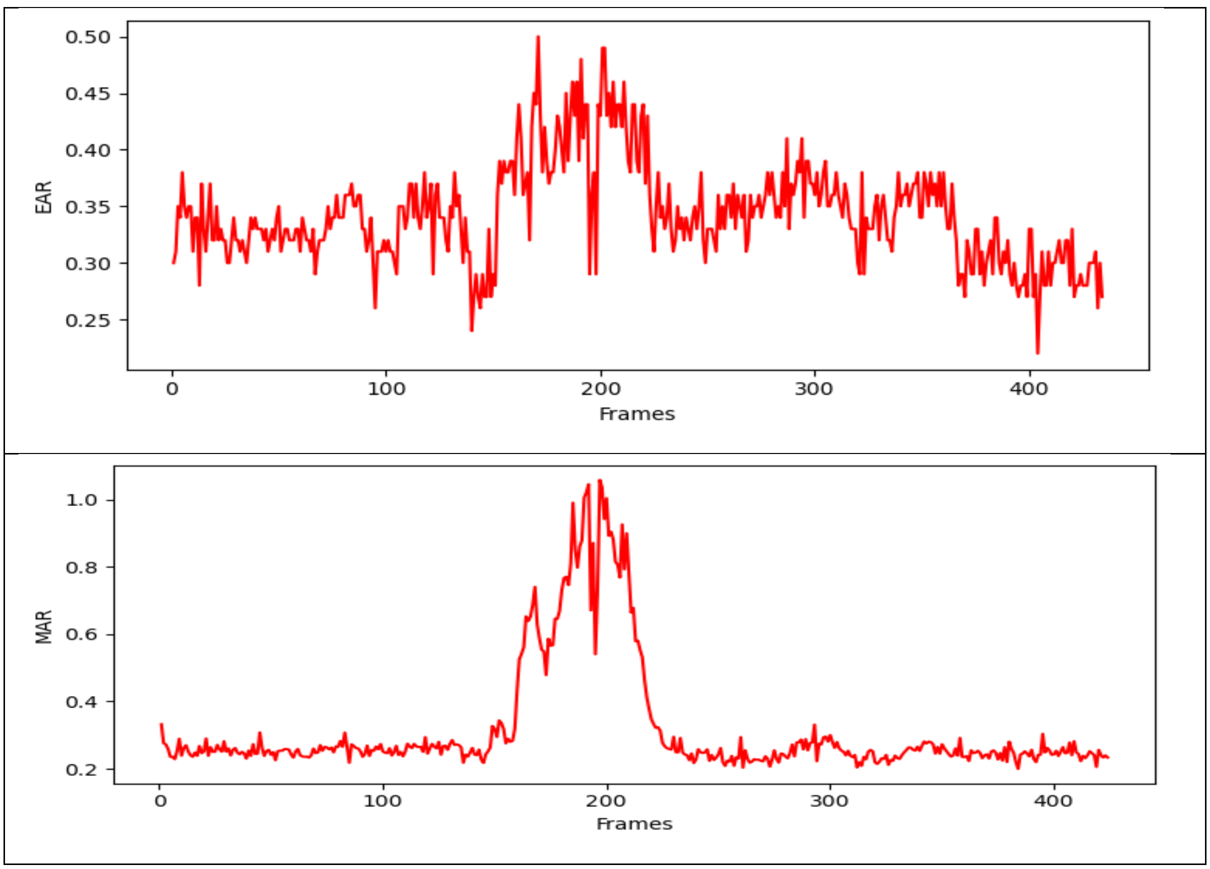


Figure 5-2: Plot of EAR and MAR values for 1-Male-NoGlasses-Yawning video. The video was not re-labelled

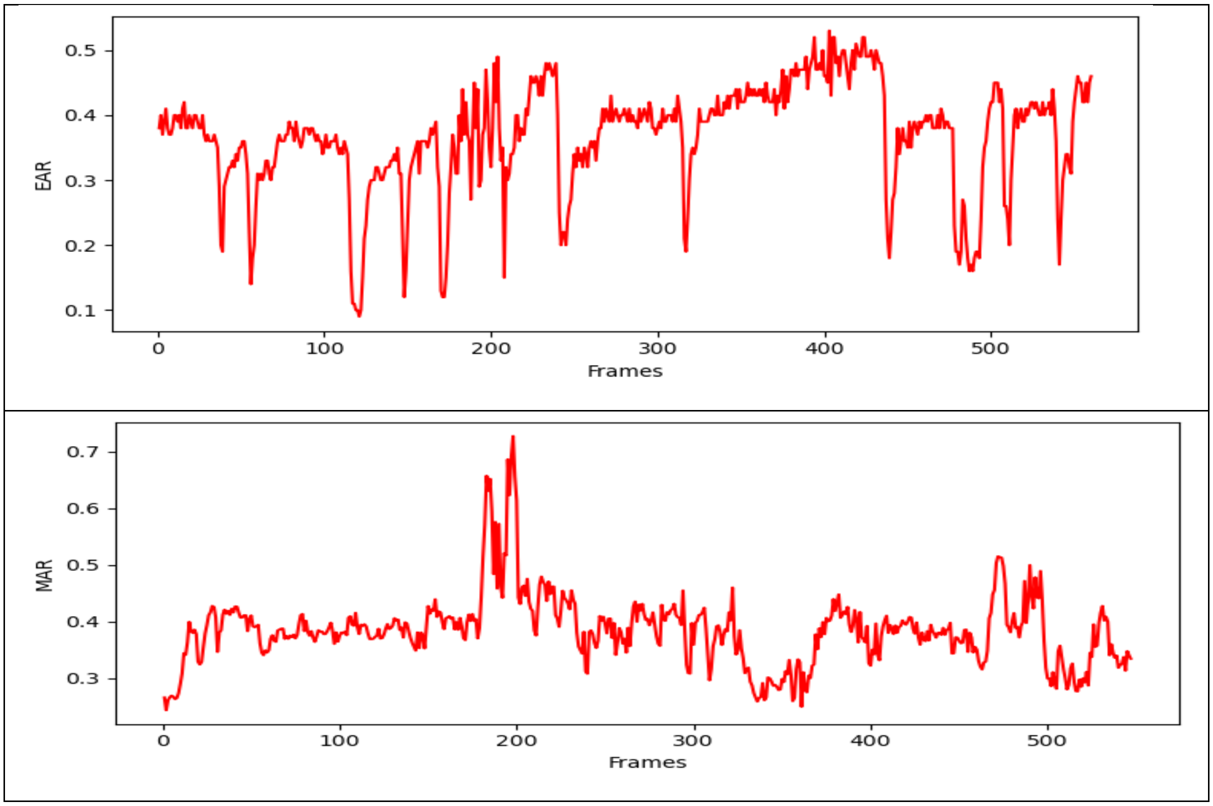


Figure 5-3: Plot of EAR and MAR values for 1-Female-NoGlasses-Talking video. The video was re-labelled normal

The data re-labelling process resulted in an imbalanced data since the percentage of the fatigue cases was only 26.6% of the entire dataset (see Figure 5-4).

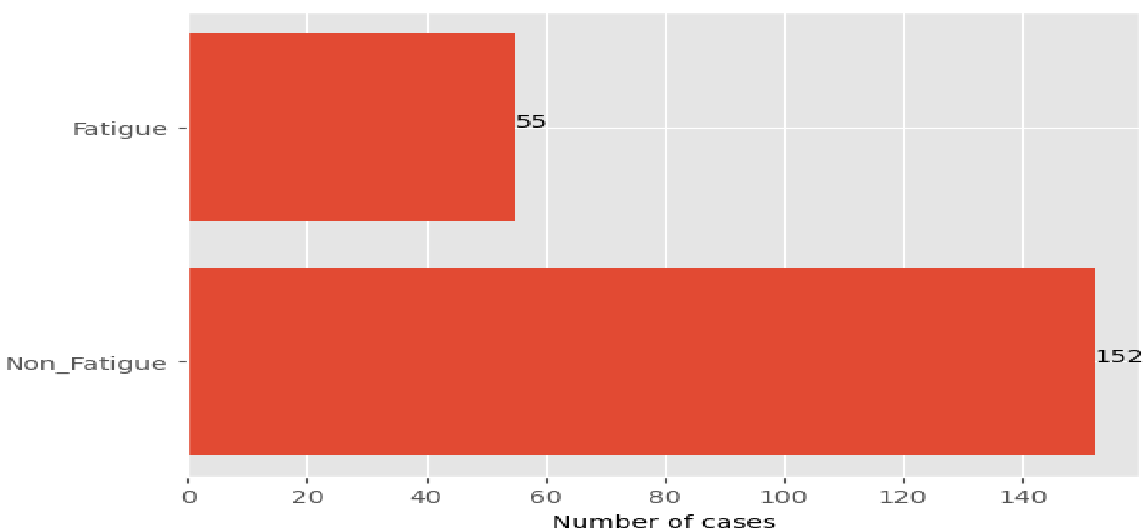


Figure 5-4: Class distribution in the re-labelled dataset

The performance and reliability of ML algorithms is negatively impacted by the use of imbalanced datasets (Anani et al., 2022; Mooijman et al., 2023). Therefore, it was necessary to consider an appropriate technique for dealing with class imbalance datasets. Resampling methods such as the under sampling and the oversampling techniques have been proposed. The under-sampling method removes random cases from the majority class whereas the over sampling method adds random cases from the minority class in order to balance the dataset. Thus, under sampling reduces the majority class and results in loss of information while oversampling uses additional copies of the minority class which can potentially result in overfitting. The oversampling method is preferred when working with a relatively small sample size as is the case in this study. The Synthetic Minority Over-Sampling Technique (SMOTE), which is the commonly used oversampling method, was used in this study. This method works by randomly selecting a case from the minority class and computing its k-nearest neighbours. Synthetic points are then generated and added between the selected case and its neighbours. This method is known to reduce the risk of overfitting since it generates synthetic samples. This technique was similarly used in Saputra and Purwitasari (2022). The imbalanced learn Python module was used to implement the SMOTE method to balance the dataset. As shown in Figure 5-5, this resulted in a balanced dataset of 304 samples, with 152 cases for each class.

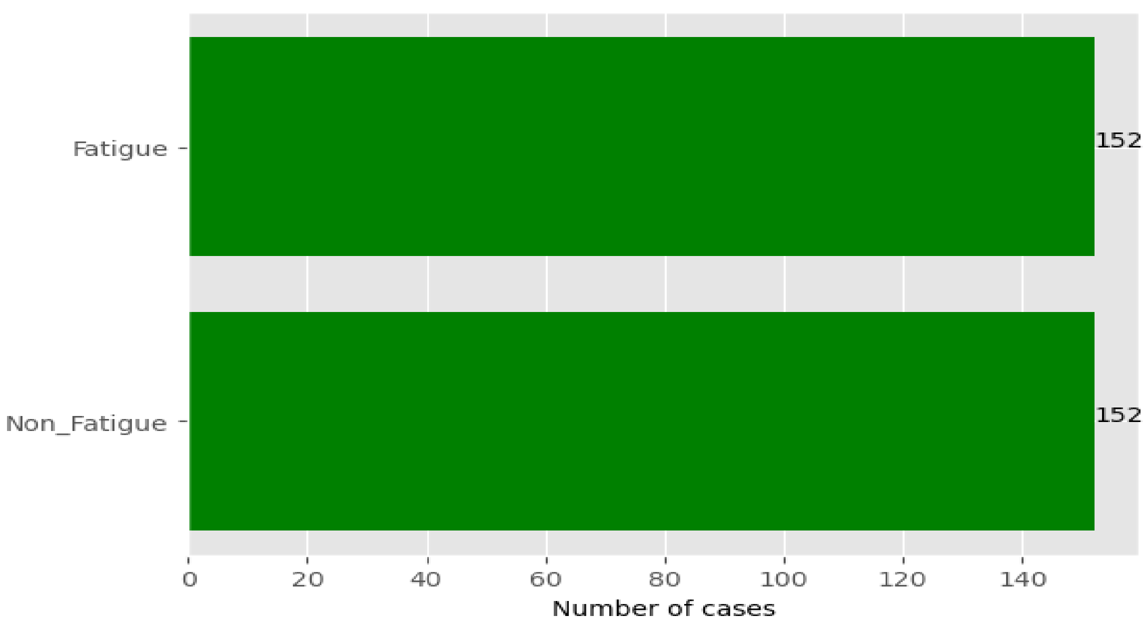


Figure 5-5: Target class distribution in the balanced dataset

5.3 Statistical Analysis of the Predictor Features

The descriptive statistical analysis of the two predictor features was performed on the unbalanced dataset and the results are shown in Table 5-1. Using the Shapiro-Wilk test for normality, the distributions of both features obtained from the YawDD were found to be non-normal. Therefore, the Student's t-test of independence could not be used since it relies on the normality assumption. Instead, the Mann-Whitney U-test which is a non-parametric alternative to the Student's t-test of independence was considered in this study. In Table 5-1, mean values for the non-fatigue and fatigue classes are denoted by $\mu_{NON-FATIGUE}$ and $\mu_{FATIGUE}$, respectively. As expected, the mean values for the fatigue class are greater than those of the non-fatigue class. The Mann-Whitney U-test statistics and p-values in the last column of Table 5-1 confirms that the differences between the means are statistically significant for both predictor features.

Table 5-1: Descriptive statistics of the two features (using unbalanced dataset)

Feature	**Shapiro-Wilk Test statistic and p-value**	$\boldsymbol{\mu_{NON-FATIGUE}}$	$\boldsymbol{\mu_{FATIGUE}}$	**Mann-Whitney U Test statistic and *p-value***
PERCLOS	0.61 (< 0.0001)	4.02	11.18	2396.5 (< 0.0001)
PMO	0.57 (<0.0001)	1.22	20.75	723.5 (< 0.0001)

The correlation heatmap in Figure 5-6 was used to establish the strength of the relationships between the predictor features and the target variable, denoted by fatigue state, as well as the relationship between the predictor features. Figure 5-6 shows that both fatigue predictor features are significantly correlated with the target variable, with PMO having the strongest correlation. This was to be expected since the data used is a yawning dataset. It also shows that the correlation coefficient between the predictor features is insignificant. Thus, both features can be included in the training set.

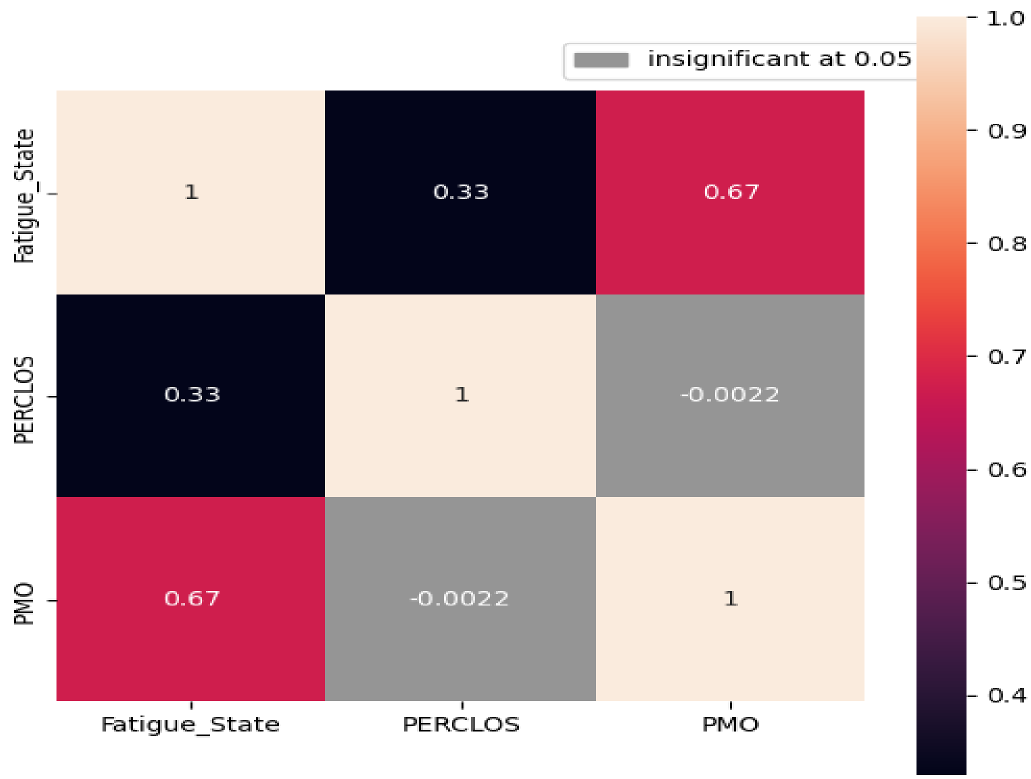


Figure 5-6: Features correlation heatmap

5.4 Experimental Evaluation on YawDD

In the first set of experiments, the YawDD dataset which consisted of 207 videos was used for training and evaluating the ML classifiers to detect fatigue. The Python Scikit-learn package was used for model building and evaluation. The performance evaluation is on the basis of the confusion matrix metrics discussed in section 4.5.

5.4.1 Hyperparameter Tuning

The grid search technique was implemented via the GridSearchCV in Scikit-Learn Python library and used to determine the optimal combinations of hyperparameters of the different classifiers adopted in this study. Table 5-2 shows the resulted hyperparameters used in this study for the different classifiers.

Table 5-2: List of hyper-parameters of adopted classifiers

ML Classifiers	Hyperparameters
DT	ccp_alpha=0.001, max_depth=9, max_features='auto'
KNN	n_neighbors=7
SVM	kernel='rbf', C=10.0, gamma=1
LR	C=0.1, penalty='l2'
k-Means*	n_clusters=2, random_state=42, algorithm='auto', init='k-means++', max_iter=100, n_init=5, tol=0.01
GNB	var_smoothing=0.01
RF	criterion='gini', max_depth=6, max_features='auto', n_estimators=100

* number of clusters manually set at 2

5.4.2 Models Evaluation

The performance evaluation results for each classifier under the different performance evaluation metrics are shown in Tables 5-3 and 5-4, for unbalanced and balanced datasets, respectively. Comparing the results for models trained with an unbalanced dataset and those of models trained with balanced dataset, it is clear that the models trained with a balanced dataset were more superior, underlying the importance of having a balanced dataset when developing classification models.

Table 5-3: Performance evaluation for the employed classifiers (results for 5-fold cross validation with unbalanced data)

ML Classifiers	Accuracy %	F1 score %	Recall %	Precision %	AUC %
RF	86.7	72.0	69.1	78.7	96.3
SVM	90.0	79.7	76.4	84.5	95.5
KNN	89.0	77.1	72.7	84.0	95.4
DT	87.1	76.0	78.2	74.9	84.3
LR	85.2	61.3	49.1	90.2	96.7
GNB	92.4	85.6	85.5	87.1	95.9
K-MEANS	66.2	32.8	32.7	67.8	-

Based on the Stratified Shuffle Split cross-validation, the SVM outperformed the other individual classifiers with an overall accuracy rate of 92.1% followed by the KNN and DT with overall accuracy rates of 91.5% and 91.1%, respectively (see Table 5-4). The SVM also had the highest F1-Score of 92.5% and best Recall value of 97.4%. This confirms that the SVM has great fatigue detection accuracy compared to other traditional ML classifiers. However, the SVM was outperformed by the ensemble model (RF) which achieved an accuracy of 93.8%. The k-Means model exhibited the worst performance.

Table 5-4: Performance evaluation for the employed classifiers (results for 5-fold cross validation with balanced data)

ML Classifiers	Accuracy %	F1 score %	Recall %	Precision %	AUC %
RF	93.8	94.0	97.4	90.9	98.2
SVM	92.1	92.5	97.4	88.5	95.5
KNN	91.5	91.8	96.0	88.3	97.3
DT	91.1	91.2	92.1	90.4	92.2
LR	90.2	89.8	87.4	92.5	96.5
GNB	89.8	89.5	88.1	91.4	96.4
K-MEANS	47.9	53.6	55.6	61.5	-

An ideal fatigue detection system should aim to ensure that a majority of the fatigue cases are correctly classified, thus reducing the number of False Negatives (FN) and resulting in Recall value greater than Precision. Table 5-4 shows that the Recall values for the RF, SVM, KNN and DT models are higher than Precision values, indicating that those models are more tolerant to True Positives (TP) in fatigue detection. However, the opposite is true for the LR, GNB and the k-Means models.

In Figure 5-7, box plots for the range of accuracy scores for the different classifiers are shown reflecting the distribution of the accuracy scores. The central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles of the accuracy scores for each classifier. The small circles represent outliers while the size of the box plot indicates the variability in the accuracy scores. Therefore, Figure 5-7 indicates that the LR, KNN and the SVM had the lowest variability in the accuracy scores since the sizes of their box plots are smaller compared to the other classifiers. This indicates that they were the most stable classifiers in detecting fatigue, even better than the ensemble model. However, the RF has higher values for the median, which indicate that on average the accuracy scores for the RF model are higher than other models. The k-Means has very high variability in the accuracy scores and very low 25th percentile. Therefore, it is not a reliable model for fatigue detection.

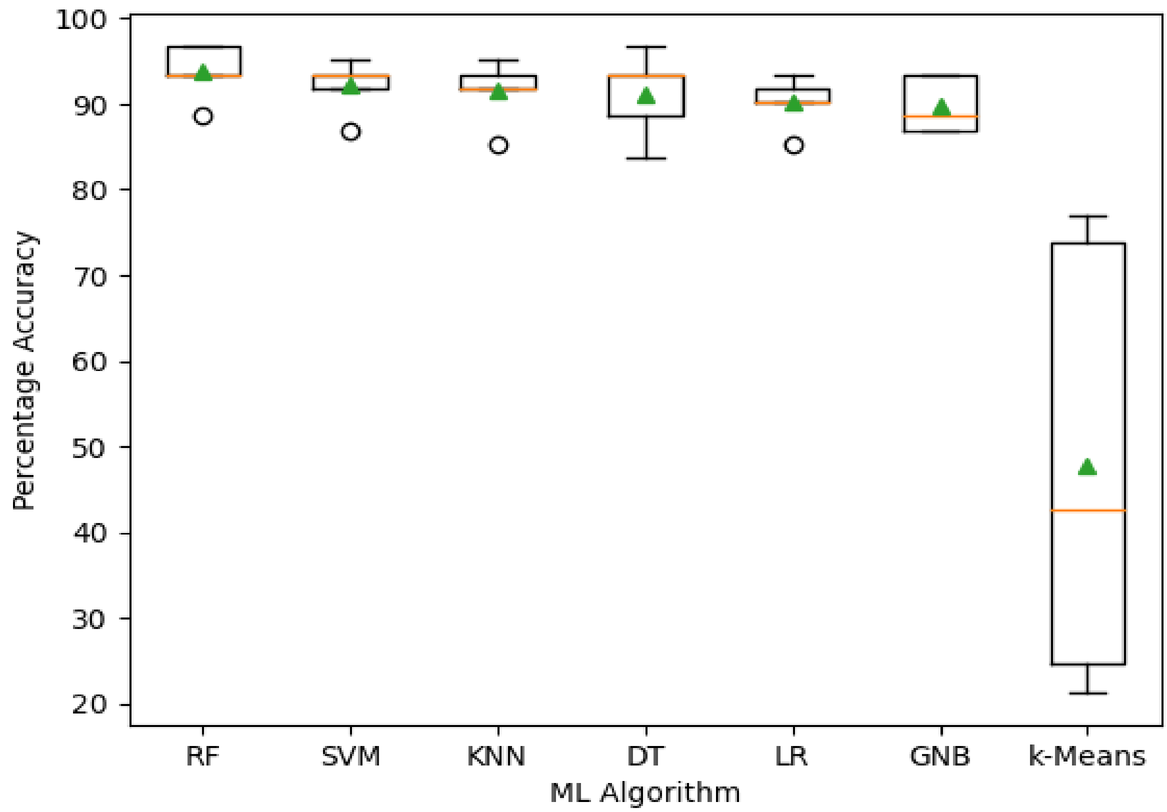


Figure 5-7: Box plot for the accuracy scores for different classifiers

Figure 5-8 displays the confusion matrix for the RF and the SVM which are the best two performing models. The RF model showed marginally higher probabilities of correctly predicting non-fatigue cases and fatigue cases compared to the SVM. This again shows that the RF performed better than the SVM.

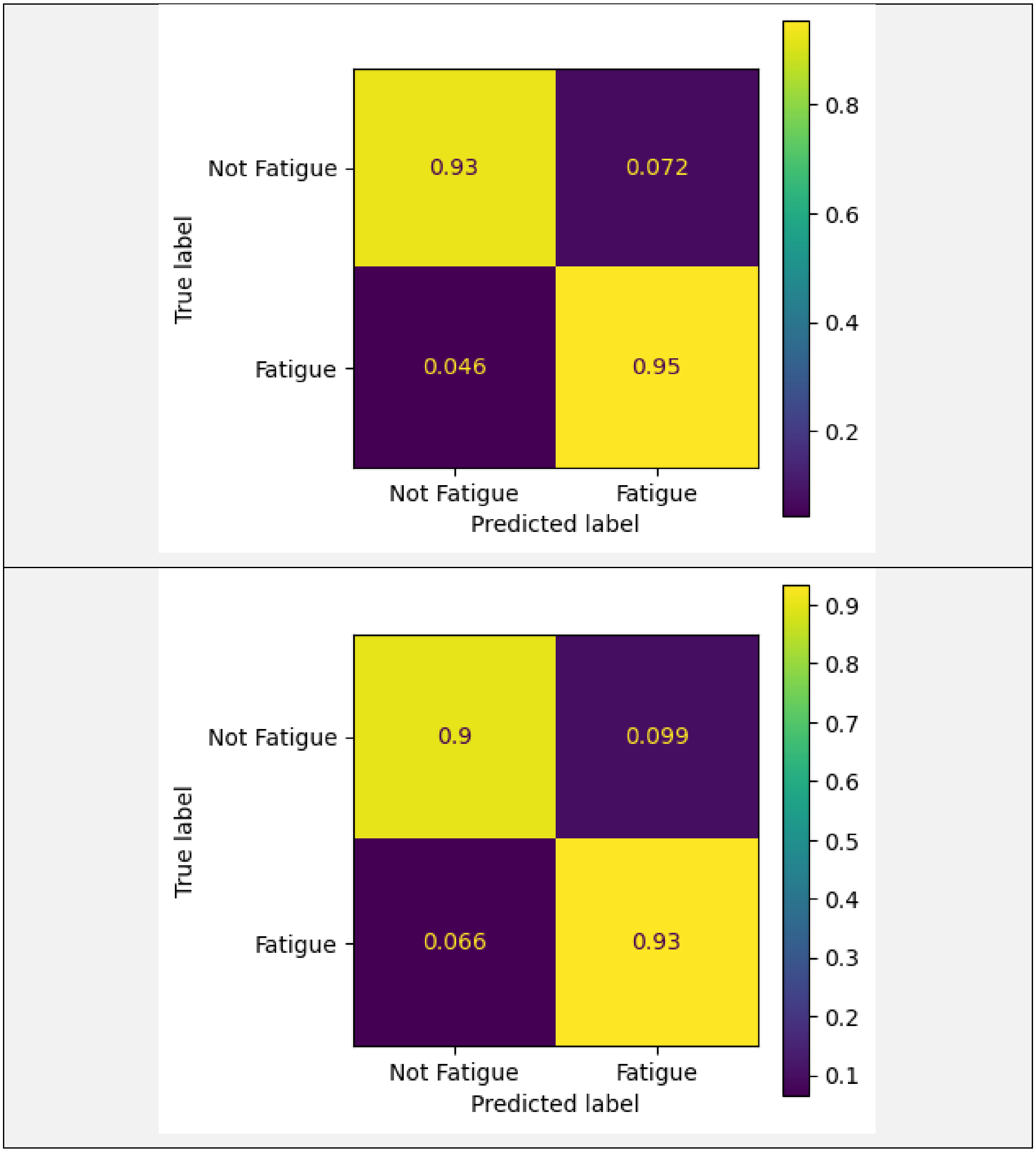


Figure 5-8: Confusion matrices for RF (top), SVM (bottom)

The comparative results showing the performance of the RF model under different conditions is assessed in Table 5-5. The results suggest that the use of glasses or sunglasses affects the performance of the model, resulting in accuracy of about 7 and 17 percentage points lower, respectively. This suggests that when the face is occluded, the model is unable to detect all the facial landmarks, particularly the eyes, thus reducing the model's ability to properly detect the state of the eyes.

Table 5-5: RF Model accuracies in different mouth conditions

Mouth conditions	**Accuracy (%)**
Barefaced	94.01
Glasses	86.76
Sunglasses	77.27

Table 5-6 provides a comparison of the results of this study against similar studies that used the YawDD for fatigue detection. Unlike other prior studies which used deep learning algorithms, the current study proposed a model that employs traditional ML algorithms which are considered to be less complex and more suitable for real time fatigue detection. The proposed model was also trained and validated using a relatively larger number of videos compared to prior studies and will, therefore, have better generalisability. Based on the accuracy of the fatigue detection models, the performance of the deep learning model proposed by Civik and Yuzgec (2023) is only about 0.7 percentage points better than the ML model proposed in this study while the performance of the RNN proposed by Chen et al. (2021) is outperformed by the ML model proposed in the current study.

Table 5-6: Comparison of the proposed system with related studies

Studies	**Number of videos**	**Methodology used**	**Fatigue detection accuracy**
Civik & Yuzgec (2023)	46 videos	Deep learning (CNN)	Accuracy (93.6%) for eye model Accuracy (94.5%) for mouth model
Zheng et al. (2023)	12 videos	Deep learning	Precision (98.8%) Recall (94.2%) F1-Score (94.3%)
L. Chen et al. (2021)	100 videos	Deep learning (RNN)	Accuracy (88%)

Proposed model	207 videos	Traditional ML algorithms	Accuracy (93.8%); F1-Score (94%) Recall (97.4%) Precision (90.9%) AUC (98.2%)

5.5 Fatigue Detection for Namdeb Data

After the fatigue detection model was trained, validated and evaluated on the YawDD, it was tested on the Namdeb ADAS dataset in order to classify the fatigue state of the Namdeb employees. This section presents the results and highlights some of the key factors that can potentially affect the performance of the system on the Namdeb data.

5.5.1 Namdeb Video Data Resizing

The size of the Namdeb ADAS video dataset was 46. Unlike the YawDD, the faces in most of the Namdeb videos were far from the camera. Therefore, it was necessary to resize the Namdeb videos in order to improve the face detection accuracy. This was implemented by using the up-sample parameter of the HOG+SVM face detector. The effects of resizing frames in terms of the average number of faces detected and the number of videos with no faces detected is shown in Table 5-7. The results indicate that without upsizing, the face detection was very low and the majority (about 59%) of the videos had no faces detected, which makes it impossible to determine the fatigue state for those subjects. However, even after frame resizing with various up-sample factors, the average number of facial frames were still low, considering that the total number of frames was between 120 and 122 for each video.

Table 5-7: Namdeb video data resizing

Up-sample factors	**Average number of facial frames**	**Number of videos with no facial frames detected**	**FPS**
0	13	27	4
1	34	9	4
2	32	10	4
3	31	8	4

Using the HOG+SVM face detector, the highest face detection accuracy was 88.5%, with almost half (46%) of the videos having face detection accuracy below 10%. Thus, the CNN face detector was considered in order to improve the performance of the system on the Namdeb data. The average face detection accuracy when using the CNN face detector was 72% compared to the average face detection accuracy of 25.6% when using the HOG+SVM face detector. However, the significant improvement in the face detection accuracy when using the CNN face detector comes at the cost of reduced computational speed which makes it a challenge to detect fatigue in real time. The results, however, show that the CNN face detector is a more robust face detector compared to the HOG+SVM, and will be more appropriate when using real world data where the face orientation varies due to a number of factors.

5.5.2 Performance Evaluation

The performance evaluation results of the fatigue detection model based on the RF, SVM and KNN classifiers on the Namdeb data are presented in Table 5-8. Due to the unreliability of results for video data with low face detection accuracy, fatigue classification for the Namdeb data was only done for videos which had at least 80% face detection accuracy. This means that the real-world dataset on which the fatigue detection system was tested is relatively small, thus affecting the reliability and generalisability of the test results.

The results show that all three classifiers achieved over 90% accuracy in fatigue detection. The RF and SVM provide similar results for majority of the metrics and they outperform the KNN.

Table 5-8: Performance evaluation for the Namdeb data

ML Classifiers	Accuracy %	F1 score %	Recall %	Precision %
RF	97.7	97.7	99.6	96.2
SVM	97.6	97.6	99.9	95.5
KNN	92.7	92.5	96.4	90.8

Overall, a comparison of the performance evaluation of the fatigue detection system on the simulated and real-world data clearly suggests that developing behavioural based fatigue detection systems for real world situations presents challenges. Most of these were unexplored in prior studies due to the fact that studies predominantly used experimental datasets to test their systems.

However, real-world driving data, particularly in environments like the mining have a lot of noise. In most cases the camera positioning and hence the face angle in the video is affected by the movements of the vehicle/truck making it a challenge to detect the faces and to extract facial features which can be used to monitor the state of the eyes and mouth in order to make a fatigue judgement. Face detection accuracy is thus a key factor in developing behavioral based fatigue detection systems. The HOG+SVM which is the traditional face detector did not perform well on real world dataset. The CNN face detector offered significant improvement in face detection accuracy and as a result it also resulted in significant improvement in fatigue detection accuracy for the Namdeb dataset. However, this reduces the system's computational speed and increases its complexity. In addition, notable challenges for developing behavioural based fatigue detection systems for real world settings like the mining environment include low image resolution due to poor and variable lighting conditions especially for videos recorded at night, face orientation to camera and proximity of face to the camera.

5.6 Chapter Summary

The implementation results of the proposed fatigue detection system were presented in this chapter. The system was trained, validated and evaluated on the YawDD and subsequently tested on the Namdeb data. The data pre-processing included the re-labelling of the target variable for the YawDD. The proposed system used behavioural based fatigue features for the eyes (PERCLOS) and mouth (PMO). Both features were found to be significantly correlated with the target variable. In addition, the means for the predictor variables were found to be significantly different for the fatigue and non-fatigue classes. The cross-validation evaluation results showed that the ensemble model (RF) had the highest fatigue detection accuracy for both simulated and real-world datasets. The accuracy of the system is, however, affected by face occlusion due to the use of glasses or sunglasses. Test results on the Namdeb ADAS data are promising, particularly when a more robust face detector which has better face detection accuracy for non-frontal faces is used. However, the reliability and generalisability of the results can be improved by using a larger sample size for the real-world dataset in order to evaluate the performance of the system.

CHAPTER SIX: RESEARCH CONCLUSION

6.1 Introduction

Workplace fatigue is one of the major risk factors across different industries. It can negatively affect employee productivity as well as their health and safety. Thus, fatigue detection and monitoring is increasingly being prioritised, particulalry in high risk occupations such as truck drivers and other vehicle operators, among others. Fatigue detection systems can leverage the visual data sourced from the ADAS. The research problem identified in this study is the need for a system to detect and classify Namdeb employee fatigue state. Thus, the study developed an ML model to detect employee fatigue state. The proposed model can be used to enhance Namdeb's employee fatigue detection and monitoring in order to improve workplace safety, in line with the eighth SGD goal.

Fatigue detection is an active area of research. The significant contribution of this study was the use of a real-world dataset to test the proposed fatigue detection system compared to prior studies which only used simulated or experimental data.

Following the development and evaluation of the proposed model as discussed in the preceding chapters, the following sections present a summary of the study by way of answering the research sub-questions. Thereafter, the research limitations and possible areas for further research are discussed.

6.2 Research Summary

The main research objective of this study was to recommend an ML model that can be used to detect the fatigue state of employees at Namdeb. This was achieved through answering the research sub-questions corresponding to research sub-objectives as presented in the coming sections.

What are the fatigue features that can be used to detect employee fatigue at Namdeb?

The ADAS implemented by Namdeb provided video data for drivers showing various driver behaviour, some of which can be used to make a judgement about the driver's fatigue state. Thus, an ADAS fatigue detection system is proposed where the driver's face is detected in the video clip sourced from the ADAS. The eye and mouth state are then monitored for fatigue signs such as frequent or prolonged periods of eye closing as well as eye opening to determine fatigue yawning. The fatigue features extracted are the PERCLOS and PMO and thus, a behavioural based fatigue detection model using facial features fusion was proposed. The approach of using multiple features is known to improve the accuracy of the fatigue detection system. This was presented in sections 4.2, 4.4.1 and 4.4.2.

How do different ML models perform in detecting and classifying employee fatigue?

The proposed system deployed seven traditional ML classifiers to develop a fatigue detection model. The grid search approach was used to determine the optimal hyperparameter combinations for the different classifiers. Using the Stratified Shuffle Split cross validation and the confusion matrix performance metrics, the models were validated and evaluated on the YawDD open source data. Performance evaluation of the models on the YawDD suggests that the ensemble model (RF) achieved a fatigue prediction accuracy of 93.8%, a Recall score of 97.4% and an AUC score of 98.2%, thereby outperforming the individual ML models. The SVM outperformed the individual ML models with an accuracy score of 92.1%, while the k-Means model was the worst performer. A comparison of the RF fatigue detection model in different conditions showed that the use of glasses and sunglasses reduces the accuracy of fatigue detection by about 7 and 17 percentage points. As highlighted in section 5.4, by comparing the results of the proposed model against similar studies that used the YawDD, it was found that the performance of the deep learning models is not significantly different from the RF model, which performed the best in this study. The performance evaluation of the RF, SVM and KNN on the Namdeb ADAS dataset showed that the RF and SVM both achieved a fatigue detection accuracy of 97.7% compared to 92.7% for KNN. While the results of this study have better generalisability because the size of the training and validation set was relatively large than in prior studies, the reliability of the test results can be improved by using a large real-world dataset for testing.

What are some of the key factors to take into consideration when selecting a fatigue detection model for Namdeb employees?

The fatigue detection model for Namdeb employees was trained and validated using the YawDD and subsequently tested on the Namdeb ADAS data.

The results presented in section 5.5 clearly indicate that face detection accuracy is the most important factor for assessing the performance of the system on the Namdeb data. Whereas the HOG+SVM face detector performed well in detecting faces in the YawDD, it performed poorly on the Namdeb data, with most of the faces having a low number of facial frames being detected. This was due to the HOG+SVM face detector only performing well in detecting frontal faces. However, for faces at odd angles, which is the case with most of the videos in the Namdeb data, as the face detector has a low face detection accuracy. Additionally, notable challenges for detecting the Namdeb employees' fatigue state include low image resolution due to poor and variable lighting conditions because some videos were recorded at night, face orientation to camera and proximity of face to the camera. The CNN face detector provided significant improvements in face detection accuracy and fatigue detection accuracy. However, this was at the cost of reduced computational speed, thus affecting the system's ability to detect fatigue in real time.

6.3 Limitations

The major limitation of this study is that the system was trained and validated on an open source data where subjects are in a stationary vehicle. Thus, validation using a different dataset, preferably with drivers in motion is crucial to improve the reliability of the system and its applicability when using real world driving data. In addition, the real-world dataset set was limited, thus affecting the reliability of the test results.

6.4 Areas of Further Research

A number of steps were taken to format the Namdeb dataset in order to align it to the YawDD. However, further exploring the data engineering practices to closely align simulated fatigue/drowsiness detection datasets and real-world datasets such as the one from the Namdeb ADAS can further improve the reliability and accuracy of the test results. Additionally, future work can focus on developing a more robust fatigue detection model using videos with more than one subject. Another potential related area of research would be to develop a system that can detect fatigue and distraction. This is particularly important for the ADAS where a majority of the driver tagging is due to suspected distraction which may not necessarily be because of fatigue.

6.5 Reflections and Lessons Learnt

The advancements in technology present an opportunity for new innovations to solve real world challenges. Among these challenges is workplace fatigue, which has been a risk factor across different industries, including the mining industry which is a backbone of some economies such as Namibia. Fatal and non-fatal injuries as well as various economic and social losses have been attributed to workplace fatigue. Therefore, there is now an increasing interest in innovations and smart technologies to detect, monitor and mitigate workplace fatigue. Some systems such as the

ADAS have been implemented and their functionalities can be enhanced to enhance their real time fatigue detection and monitoring. AI and ML methods have proven to be useful tools in achieving that. However, in order for AI and ML based fatigue systems to have high accuracy and to be effective, they require quality data. However, this can be a challenge in some work environments like in the mines where some factors can be difficult to control. Nevertheless, there are great prospects in the development of such systems, though it would require some investments in data collection equipment. However, by leveraging the existing systems such as ADAS, fatigue detection systems can be tested in real world settings.

Notably, the literature on fatigue detection has predominantly focused on driver fatigue and this is understandable because driving is one of the most high-risk occupations. However, there is a need to explore fatigue detection systems for other occupation categories and in different domains by leveraging the existing work on driver fatigue detection systems. The researcher had this in mind at the initial conceptualisation of this study, however, it did not materialise due to data limitations.

6.6 Thesis Conclusion

This study aimed to address the need for a fatigue detection system for Namdeb employees by evaluating and recommending an ML based fatigue detection model. The study followed a positivism philosophy by employing a deductive approach and adopted a quantitative research design using an experiment strategy and secondary cross-sectional data.

A discussion of the proposed system and the implementation results are presented in Chapters four and five. The data sourced from Namdeb was relatively small, hence the proposed system was trained, validated and evaluated on an open source data (YawDD) and subsequently tested on the Namdeb data. While there are some differences between the two datasets, care was taken in the presentation of the results in section 5.5 to ensure the reliability of the results.

The main contribution of this work is the use of real-world driving data to test the driver fatigue detection systems and to subsequently identify some of the key factors for consideration in the development of employee fatigue detection systems for the mining domain. Prior studies have mostly only evaluated their systems on experimental or simulated data.

The recommended model can be used in conjunction with the ADAS in order to determine if employees tagged with drowsy related behaviour are actually fatigued or not. This can potentially result in improved workplace safety.

APPENDIX A: Ethical Clearance Letter

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY RESEARCH ETHICS COMMITTEE (F-REC)

DECISION/FEEDBACK ON RESEARCH PROPOSAL

Dear Samuel Nghidengwa Nakale (221141995)

RESEARCH TOPIC: DEVELOPING A MACHINE LEARNING MODEL TO PREDICT FATIGUE LEVELS FOR EMPLOYEES AT NAMDEB

Supervisor (if applicable): Prof Fungai Bhunu Shava

Qualification registered for (if applicable): Master of Computer Science

(Reference number of applications: **FACULTY RESEARCH ETHICS COMMITTEE REGISTRATION NUMBER: FREC-04/23**)

Re: Ethical screening application No: **FREC-04/23**

The Faculty of **Computing and Informatics** Ethics Screening Committee of the Namibia University of Science and Technology reviewed your application for the above-mentioned research. The research as set out in the application has been:

Approved	X

(Indicate with an X, and N/A if not applicable and proceed)

We would like to point out that you, as a researcher, are obliged to maintain the ethical integrity of your research, adhere to the ethical guidelines of NUST, and remain within the scope of your research proposal and supporting evidence as submitted to the F-REC. Should any aspect of your research change from the information as presented to the F-REC, which could affect the possibility of harm to any research subject, you are under the obligation to report it immediately to your supervisor or F-REC as applicable in writing. Should there be any uncertainty in this regard, you must consult with the F-REC.

We wish you success with your research, and trust that it will make a positive contribution to the quest for knowledge at NUST.

Any ethical issues that need to be highlighted?	Why are these issues important?	What must/could be done to minimize the ethical risk?
No	N/A	N/A

Recommendation: The application is approved.

Sincerely,

Dr Suama L Hamunyela
Chair: Faculty Ethics Screening Committee
Tel: +264-61-207-2922

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY
Windhoek NAMIBIA
2023 -03- 07
OFFICE OF THE DEAN
FACULTY OF COMPUTING & INFORMATICS

CC: Co-supervisor: Dr Gloria Iyawa

APPENDIX B: Namdeb Consent Letter

NAMDEB

A NAMIBIA DE BEERS PARTNERSHIP

14 April 2023

CONFIDENTIAL

Samuel Ngidengwa Nakale
Namibia University of Science and Technology
Windhoek
Namibia

PERMISSION TO CONDUCT RESEARCH AS PART OF THE REQUIREMENTS FOR THE AWARD OF MASTERS QUALIFICATION

Dear Samuel,

We refer to your email dated 15 March 2023, requesting to conduct master's dissertation titled *"Developing A Machine Learning Model to Predict Fatigue Levels for Employees at Namdeb"*.

I am pleased to inform you that your request has been granted subject to the following conditions:

- You are obliged to observe all ethical guidelines as stipulated by the research ethics committee of the Institute of Higher Learning that you are registered with for the duration of your study period.
- That all information gained from the research should be treated with utmost circumspection, confidentiality, and anonymity.
- You are required to notify the company should there be any substantial change or amendments to your study.
- That Namdeb retains all intellectual property rights to the research project and any development of technology emanating therefrom may not be made without its consent.

Wishing you all the best with your research project and looking forward to partner with you on this journey.

Yours sincerely,

Jurgen Jacob
CHIEF OPERATING OFFICER

I SAMUEL NAKALE, have read and understand the content of this letter and am in full agreement with the content thereof.

SIGNATURE

20/04/2023
DATE

NAMDEB DIAMOND CORPORATION (PROPRIETARY) LIMITED

Tenth floor Namdeb Centre 10 Dr Frans Indongo Street PO Box 1906 Windhoek Namibia
Tel +264 61 204 3333 | Fax +264 61 204 3334 | www.namdeb.com
Incorporated in Namibia | Registration number C1973

APPENDIX C: Language Editor’s Letter

ACET Consultancy
Anenyasha Communication, Editing and Training
Box 50453 Bachbrecht, Windhoek, Namibia
Cell: +264814218613
Email: mlambons@yahoo.co.uk

1 April 2024

To whom it may concern

LANGUAGE EDITING – SAMUEL NGHIDENGWA NAKALE

This letter serves to confirm that a **MASTER OF DATA SCIENCE** research titled **RECOMMENDING A MACHINE LEARNING MODEL TO DETECT THE FATIGUE STATE FOR EMPLOYEES AT NAMDEB** was submitted to me for language editing.

The research was professionally edited and track changes and suggestions were made in the document. The research content or the author’s intentions were not altered during the editing process and the author has the authority to accept or reject my suggestions.

Yours faithfully

NMlambo

PROF. (DR) NELSON MLAMBO
PhD in English
M.A. in Intercultural Communication
M.A. in English
B. A. Special Honours in English – First class
B. A. English & Linguistics

APPENDIX D: Python Code

The proposed system was implemented in Python and the code is available in a GitHub repository and can be accessed using the following link: https://github.com/hinaunye/Fatigue-Detection

REFERENCES

Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., & Hariri, B. (2014). *YawDD*. 24–28. https://doi.org/10.1145/2557642.2563678

Ajith, M. M., & Ghosh, A. K. (2019). Economic and social challenges faced by injured artisanal and small-scale gold miners in Kenya. *Safety Science*, *118*(December 2018), 841–852. https://doi.org/10.1016/j.ssci.2019.05.058

Alharbey, R., Dessouky, M. M., Sedik, A., Siam, A. I., & Elaskily, M. A. (2022). Fatigue state detection for tired persons in presence of driving periods. *IEEE Access*, *10*, 79403–79418. https://doi.org/10.1109/ACCESS.2022.3185251

Anani, A., Risso, N., Nyaaba, W., & Tenorio, V. (2022). Application of machine learning in mine safety: A state-of-the-art review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4314075

Ayachi, R., Afif, M., Said, Y., & Abdelali, A. B. (2021). Drivers fatigue detection using efficientdet in advanced driver assistance systems. *2021 18th International Multi-Conference on Systems, Signals & Devices (SSD)*, 738–742.

Balaskas, K., & Siozios, K. (2021). Fatigue detection using deep long short-term memory autoencoders. *2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 1–4. https://doi.org/10.1109/MOCAST52088.2021.9493378

Bauerle, J. T., Sammarco, J. J., & Dugdale, J. Z. (2021). The human factors of mineworker fatigue: An overview on prevalence, mitigation, and what's next. *American Journal of Industrial Medicine*. https://doi.org/10.1002/ajim.23301

Bauerle, T., Dugdale, Z., & Poplin, G. (2018). Mineworker fatigue: A review of what we know and future directions. *Vision, Innovation and Identity: Step Change for a Sustainable Future - 2018 SME Annual Conference and Expo and 91st Annual Meeting of the SME-MN Section*, *2018-Febru*.

Brugman, S. (2022). *The development of a real-time monitoring system for fatigue detection on truckers*. University of Twente.

Bustos, D., Cardoso, F., Rios, M., Vaz, M., Guedes, J., Torres Costa, J., Santos Baptista, J., & Fernandes, R. J. (2023). Machine learning approach to model physical fatigue during incremental exercise among firefighters. *Sensors*, *23*(1), 1–13.

https://doi.org/10.3390/s23010194

Chao, L., Changyuan, W., Guang, L., & Lu, S. (2019). *Detection of blink state based on fatigued driving*, *04*(04), 24–29. https://doi.org/10.21307/ijanmc-2019-067

Chen, J., Yan, M., Zhu, F., Xu, J., Li, H., & Sun, X. (2022). Fatigue driving detection method based on combination of bp neural network and time cumulative effect. *Sensors*, *22*(13). https://doi.org/10.3390/s22134717

Chen, K., Zhu, T., Li, S., & Shi, Y. (2021). Real time yawning detection based on machine learning algorithm and time series classification using facial feature points. *International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, 276–280.

Chen, L., Xin, G., Liu, Y., & Huang, J. (2021). Driver fatigue detection based on facial key points and LSTM. *Security and Communication Networks*, *2021*, 1–9. https://doi.org/10.1155/2021/5383573

Chen, S., Xu, K., Yao, X., Zhu, S., Zhang, B., & Zhou, H. (2021). Psychophysiological data-driven multi-feature information fusion and recognition of miner fatigue in high-altitude and cold areas. *Computers in Biology and Medicine*, *133*(April), 104413. https://doi.org/10.1016/j.compbiomed.2021.104413

Chen, Y. (2022). Driver fatigue detection using machine learning methods. *2022 IEEE international conference on artificial intelligence and computer applications, ICAICA 2022*, 906–910. https://doi.org/10.1109/ICAICA54878.2022.9844425

Chinthalachervu, R., Teja, I., Kumar, M. A., Harshith, N. S., & Kumar, T. S. (2022). Driver drowsiness detection and monitoring system using machine learning. *Journal of Physics: Conference Series*, *2325*(1), 012057. https://doi.org/10.1088/1742-6596/2325/1/012057

Civik, E., & Yuzgec, U. (2023). Real-time driver fatigue detection system with deep learning on a low-cost embedded system. *Microprocessors and Microsystems*, *99*(December 2022), 104851. https://doi.org/10.1016/j.micpro.2023.104851

Darbandy, M., Rostamnezhad, M., Hussain, S., Khosravi, A., Nahavandi, S., & Sani, Z. (2020). A new approach to detect the physical fatigue utilizing heart rate signals. *Research in Cardiovascular Medicine*, *9*(1), 23. https://doi.org/10.4103/rcm.rcm_8_20

Dewi, C., Chen, R. C., Jiang, X., & Yu, H. (2022). Adjusting eye aspect ratio for strong eye blink detection based on facial landmarks. *PeerJ Computer Science*, *8*(2020), 1–21. https://doi.org/10.7717/peerj-cs.943

Drews, F. A., Rogers, W. P., Talebi, E., & Lee, S. (2020). The experience and management of fatigue: A study of mine haulage operators. *Mining, Metallurgy and Exploration*, *37*(6), 1837–1846. https://doi.org/10.1007/s42461-020-00259-w

Duan, C., Liu, Z., Xia, J., Zhang, M., Liao, J., & Cao, L. (2023). *Enhancing cross-dataset performance of distracted driving detection with score-softmax classifier*, *14*(8), 1–12.

Ekener, F., & Ekenstam, F. (2023). *Self-supervised pre-training for drowsiness prediction*. Chalmers University of Technology.

El-Nabi, S. A., El-Shafai, W., El-Rabaie, E. S. M., Ramadan, K. F., Abd El-Samie, F. E., & Mohsen, S. (2023). Machine learning and deep learning techniques for driver fatigue and drowsiness detection: a review. *Multimedia Tools and Applications*, 1–37. https://doi.org/10.1007/s11042-023-15054-0

Engineering, C., Engineering, C., & Engineering, C. (2021). Real time driver drowsiness detection using opencv and facial landmarks. *Journal of Contemporary Issues in Business and Government*, *27*(6), 4297–4314. https://doi.org/10.47750/cibg.2021.27.06.054

Fan, X., Yin, B. C., & Sun, Y. F. (2007). Yawning detection for monitoring driver fatigue. *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, ICMLC 2007*, *2*(August), 664–668. https://doi.org/10.1109/ICMLC.2007.4370228

Fatigue Science. (2020). Digging into fatigue in mining. https://fatiguescience.com/blog/digging-into-fatigue-in-mining/#_Toc10983371

Federico, G. (2022). Driver drowsiness detection – AI system. Linkedin. https://www.linkedin.com/pulse/driver-drowsiness-detection-ai-system-giulio-federico

Gargan, I. E. (2021). *Model-based validation of driver drowsiness detection system for ADAS*. https://amslaurea.unibo.it/25716/1/Ivan_Enzo_Gargano_Model%20Based%20validation%20of%20Driver%20Drowsiness%20Detection%20System%20for%20ADAS_Thesis.pdf

Golz, M., Sommer, D., Trutschel, U., Sirois, B., & Edward, D. (2010). Evaluation of fatigue monitoring technologies. *Somnologie*, *14*(3), 187–199. https://doi.org/10.1007/s11818-010-0482-9

Grządzielewska, M. (2021). Using machine learning in burnout prediction: A survey. *Child and Adolescent Social Work Journal*, *38*(2), 175–180. https://doi.org/10.1007/s10560-020-00733-w

Gurudath, N., & Riley, B. H. (2014). Drowsy driving detection by EEG analysis using Wavelet

Transform and K-means clustering. *Procedia Computer Science*, *34*, 400–409. https://doi.org/10.1016/j.procs.2014.07.045

Hooda, R., Joshi, V., & Shah, M. (2022a). A comprehensive review of approaches to detect fatigue using machine learning techniques. *Chronic Diseases and Translational Medicine*, *8*(1), 26–35.

Hooda, R., Joshi, V., & Shah, M. (2022b). A comprehensive review of approaches to detect fatigue using machine learning techniques. *Chronic Diseases and Translational Medicine*, *8*(1), 26–35. https://doi.org/10.1016/j.cdtm.2021.07.002

Hooda, R., Joshi, V., & Shah, M. (2022c). A comprehensive review of approaches to detect fatigue using machine learning techniques. In *Chronic Diseases and Translational Medicine*, *8*(1), 26–35. John Wiley and Sons Inc. https://doi.org/10.1016/j.cdtm.2021.07.002

International Labour Organization. (2003). *Report VI ILO standards-related activities in the area of occupational safety and health : An in-depth study for discussion with a view to the elaboration of a plan of action for such activities Sixth item on the agenda*.

Islam, A., Rahaman, N., & Rahman Ahad, M. A. (2019). A study on tiredness assessment by using eye blink detection. *Jurnal Kejuruteraan*, *31*(2), 209–214. https://doi.org/10.17576/jkukm-2019-31(2)-04

Jie, Z., Mahmoud, M., Stafford-Fraser, Q., Robinson, P., Dias, E., & Skrypchuk, L. (2018). Analysis of yawning behaviour in spontaneous expressions of drowsy drivers. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 571–576. https://doi.org/10.1109/FG.2018.00091

Junaedi, S., & Akbar, H. (2018). Driver drowsiness detection based on face feature and PERCLOS. In *Journal of Physics: Conference Series* (Vol. 1090, p. 012037). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1090/1/012037/pdf

Jung, D., & Choi, Y. (2021). Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. *Minerals*, *11*(2), 1–20. https://doi.org/10.3390/min11020148

Kantardzic, M. (2011). *Data mining: Concepts, models, methods, and algorithms*. John Wiley & Sons.

Karar, S., & Kanumuri, T. (2023). Assessment of driver fatigue and drowsiness based on eye blink rate. *International Conference on Data Analytics & Management*, 311–324.

https://doi.org/10.1007/978-981-99-6550-2

Kumar, A., & Patra, R. (2018). Driver drowsiness monitoring system using visual behaviour and machine learning. *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 339–344.

Kumari, S., Akanksha, K., Pahadsingh, S., Singh, S., & Singh, S. (2021). Drowsiness and yawn detection system using python. In *Proceedings of International Conference on Communication, Circuits, and Systems: IC3S 2020: Vol. 728 LNEE* (pp. 225–232). https://doi.org/10.1007/978-981-33-4866-0_33

Lee, Y. L., Chou, W., Chien, T. W., Chou, P. H., Yeh, Y. T., & Lee, H. F. (2020). An app developed for detecting nurse burnouts using the convolutional neural networks in microsoft excel: Population-based questionnaire study. *JMIR Medical Informatics*, *8*(5). https://doi.org/10.2196/16528

Lock, N. K., Ng, W. M., Jusoh, N. A., Kamarudin, N. H., Ramli, R., & Zulkoffli, Z. (2023). Drowsiness detection for safe driving using PERCLOS and YOLOv2 Method. *Lecture Notes in Electrical Engineering*, *882*, 101–112. https://doi.org/10.1007/978-981-19-1577-2_9

Majeed, F., Shafique, U., Safran, M., Alfarhood, S., & Ashraf, I. (2023). Detection of drowsiness among drivers using novel deep convolutional neural network model. *Sensors*, 1–26.

Martin, S., Tawari, A., & Mohan, M. T. (2014). Toward privacy-protecting safety systems for naturalistic driving videos. *IEEE Transactions on Intelligent Transportation Systems*, 1811–1822.

Mayilvahanan, S. (2020). *Yawn Detection using Support Vector Machine* [Texas A&M University]. https://doi.org/10.22214/ijraset.2020.32227

Ministry of Labour, Industrial Relations and Employment Creation. (2021). NATIONAL OCCUPATIONAL SAFETY & HEALTH POLICY. https://mol.gov.na/documents/53329/0/National+Occupational+Safety+and+Health+Policy+%282%29.pdf/dd661e21-6331-ba2e-4b9c-4eb844c65fbb

Mohanty, S., Hegde, S. V., Prasad, S., & Manikandan, J. (2019). Design of real-time drowsiness detection system using dlib. *2019 5th IEEE International WIE Conference on Electrical and Computer Engineering, WIECON-ECE 2019 - Proceedings*, 4–7. https://doi.org/10.1109/WIECON-ECE48653.2019.9019910

Mooijman, P., Catal, C., Tekinerdogan, B., & Lommen, A. (2023). The effects of data balancing approaches : A case study. *Applied Soft Computing*, *132*, 109853. https://doi.org/10.1016/j.asoc.2022.109853

Morais, C., Ribeiro, J., & Silva, J. (2023). Human factors in aviation: Fatigue management in ramp workers. *Open Engineering*, *13*(1), 1–10. https://doi.org/10.1515/eng-2022-0411

Namibia Chamber of Mines. (2023). *2022 Annual Review*.https://chamberofmines.org.na/wp-content/uploads/2023/04/2022-Chamber-of-Mines-Annual-Review.pdf

Namibia Statistics Agency. (2022). *Annual National Accounts 2021*. https://nsa.nsa.org.na/wp-content/uploads/2022/08/Annual-National-Accounts-2021.pdf

Nasirzadeh, F., Mir, M., Hussain, S., Darbandy, M. T., Khosravi, A., Nahavandi, S., & Aisbett, B. (2020). Physical fatigue detection using entropy analysis of heart rate signals. *Sustainability (Switzerland)*, *12*(7). https://doi.org/10.3390/su12072714

Ngxande, M., Tapamo, J. R., & Burke, M. (2017). Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques. *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, PRASA-RobMech 2017*, *2018-Janua*, 156–161. https://doi.org/10.1109/RoboMech.2017.8261140

Pelders, J., & Nelson, G. (2019). Contributors to fatigue at a platinum smelter in South Africa. *Journal of the Southern African Institute of Mining and Metallurgy*, *119*(3), 313–319. https://doi.org/10.17159/2411-9717/2019/v119n3a11

Perkins, E., Sitaula, C., Burke, M., & Marzbanrad, F. (2023). Challenges of drivers drowsiness prediction: The remaining steps to implementation. *IEEE Transactions on Intelligent Vehicles.*, *8*(2), 1319–1338.

Petrellis, N., Zogas, S., Christakos, P., Keramidas, G., Mousouliotis, P., Voros, N., & Antonopoulos, C. (2021). High speed implementation of the deformable shape tracking face alignment algorithm. *Proceedings - 2021 24th Euromicro Conference on Digital System Design, DSD 2021*, 174–177. https://doi.org/10.1109/DSD53832.2021.00035

Phillips, R. O. (2015). A review of definitions of fatigue - And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour*, *29*, 48–56. https://doi.org/10.1016/j.trf.2015.01.003

Pothiraj, S., Vadlamani, R., & Reddy, B. P. K. (2021). A non-intrusive method for driver drowsiness detection using facial landmarks. In *3C Tecnología_Glosas de innovación* *aplicadas a la pyme* (pp. 71–85). https://doi.org/10.17993/3ctecno.2021.specialissue8.71-85

Rahul, K., Raj, S. G., Rajesh, S., & Udhayakumar, G. (2022). Driver's drowsiness detection by analyzing yawning and eye closure. *International Research Journal of Engineering and Technology*, *May*, 3101–3104.

Raja Mohana, S. P., Manu Vidhya, S., & Reshma, D. (2021). A Real-time Fatigue Detection System using MultiTask Cascaded CNN Model. *Proceedings - 2021 IEEE 10th International Conference on Communication Systems and Network Technologies, CSNT 2021*, 674–679. https://doi.org/10.1109/CSNT51715.2021.9509627

Ramos, P. M. S., Maior, C. B. S., Moura, M. C., & Lins, I. D. (2022). Automatic drowsiness detection for safety-critical operations using ensemble models and EEG signals. *Process Safety and Environmental Protection*, *164*(June), 566–581. https://doi.org/10.1016/j.psep.2022.06.039

Rohit, F. (2016). *Real-time drowsiness detection using wearable , lightweight EEG sensors*. https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=7564&context=etd

Rohit, F., Kulathumani, V., Kavi, R., Elwarfalli, I., Kecojevic, V., & Nimbarte, A. (2017). Real-time drowsiness detection using wearable, lightweight brain sensing headbands. *IET Intelligent Transport Systems*, *11*(5), 255–263. https://doi.org/10.1049/iet-its.2016.0183

Sabry, F., Eltaras, T., Labda, W., Alzoubi, K., & Malluhi, Q. (2022). Machine learning for healthcare wearable devices: The big picture. *Journal of Healthcare Engineering*, *2022*.

Saputra, W., & Purwitasari, D. (2022). Fatigue Management: Machine learning application for predicting mining worker fatigue. *2022 International Conference on Information Technology Research and Innovation, ICITRI 2022*, 117–122. https://doi.org/10.1109/ICITRI56423.2022.9970203

Saunders, M., Phillip, L., & Thornhill, A. (2019). *Research methods for business students* (8th ed.). Pearson Education, Limited.

Savas, B. K., & Becerikli, Y. (2018). Real time driver fatigue detection based on SVM Algorithm. *2018 6th International Conference on Control Engineering and Information Technology, CEIT 2018*, *October*, 25–27. https://doi.org/10.1109/CEIT.2018.8751886

Sedighi, Z., Chen, Y., Baghdadi, A., Lombardo, S., Cavuoto, L. A., & Megahed, F. M. (2020). *A data analytic framework for physical fatigue management using wearable sensors*. *155*. https://doi.org/10.1016/j.eswa.2020.113405

Shalash, W. M. (2021). A deep learning cnn model for driver fatigue detection using single EEG channel. *Journal of Theoretical and Applied Information Technology*, *31*(2).

Sikander, G., & Anwar, S. (2019). Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2339–2352.

Sowa, A. (2022). *Development of a computer vision-based system for recognising fatigue of truck drivers*. http://essay.utwente.nl/92474/1/Sowa_BA_EEMCS.pdf

SP, R. M., Manu, V. S., & Reshma, D. (2021). A real-time fatigue detection system using multi-task Cascaded CNN Model. *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, 674–679.

Sprajcer, M., Thomas, M. J. W., Sargent, C., Crowther, M. E., Boivin, D. B., Wong, I. S., Smiley, A., & Dawson, D. (2022). How effective are fatigue risk management systems (FRMS)? A review. *Accident Analysis and Prevention*, *165*(October 2021), 106398. https://doi.org/10.1016/j.aap.2021.106398

Sri Mounika, T. V. N. S. R., Phanindra, P. H., Sai Charan, N. V. V. N., Kranthi Kumar Reddy, Y., & Govindu, S. (2022). *Driver drowsiness detection using eye aspect ratio (EAR), Mouth Aspect Ratio (MAR), and Driver Distraction Using Head Pose Estimation* (ICT System). Springer Singapore.

Stancin, I., Cifrek, M., & Jovic, A. (2021). *Driver drowsiness detection systems*. https://doi.org/10.3390/s21113786

Talebi, E., Rogers, W. P., & Drews, F. A. (2022). Environmental and work factors that drive fatigue of individual haul truck drivers. *Mining*, *2*(3), 542–565. https://doi.org/10.3390/mining2030029

World Health Organisation. (1948). *Summary reports on proceedings minutes and final acts of the international health conference held in New York from 19 June to 22 July 1946*.

Yadav, N., Banerjee, K., & Bali, V. (2020). A survey on fatigue detection of workers using machine learning. *International Journal of E-Health and Medical Communications*, *11*(3), 1–8. IGI Global. https://doi.org/10.4018/IJEHMC.2020070101

Yi, Y., Zhang, H., Zhang, W., Yuan, Y., & Li, C. (2023). Fatigue working detection based on facial multifeature fusion. *IEEE Sensors Journal*, *23*(6), 5956–5961. https://doi.org/10.1109/JSEN.2023.3239029

Yu-Hsuan, T., Tseng, M.-C., Wang, C.-Y., & Fuh, C.-S. (n.d.). *Detection of driver drowsiness* *using multi-task learning*. https://www.csie.ntu.edu.tw/~fuh/personal/DetectionofDriverDrowsinessUsingMulti-TaskLearning.pdf

Zhang, W., Murphey, Y. L., Wang, T., & Xu, Q. (2015). Driver yawning detection based on deep convolutional neural learning and robust nose tracking. *International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Zhao, G., He, Y., Yang, H., & Tao, Y. (2022). Research on fatigue detection based on visual features. *IET Image Processing*, *16*(4), 1044–1053. https://doi.org/10.1049/ipr2.12207

Zheng, H., Wang, Y., & Liu, X. (2023). Adaptive driver face feature fatigue detection algorithm research. *Applied Sciences (Switzerland)*, *13*(8). https://doi.org/10.3390/app13085074