A Multi-Modal Machine Learning Approach for Fatigue Detection Using Physiological Signals

Submitted September 2025, in partial fulfillment of the conditions for the award of the degree Msc Computer Science (Artificial Intelligence).

By

Mohd Ali Tahir

20602350

Supervised by Dr. Armaghan Moemeni

School of Computer Science

University of Nottingham

I hereby declare that this dissertation is all my own work, except as indicated in the text

Signature

19th September 2025.

Date

Abstract

Driver fatigue is a leading contributor to road accidents worldwide, making its timely detection a critical challenge for improving transportation safety. This dissertation investigates a multi-modal machine learning approach for fatigue detection using the publicly available SEED-VIG dataset, which contains synchronised electroencephalography (EEG), electrooculography (EOG), and eye-closure (PERCLOS) measures recorded in a simulated driving environment. EEG signals were decomposed into canonical frequency bands and reduced via region-of-interest principal component analysis (ROI-PCA) to capture spatially meaningful patterns, while EOG features quantified blink and saccade behaviour.

A comprehensive experimental framework was developed, incorporating classical machine learning models (Random Forest, XGBoost, Support Vector Machines) and deep learning architectures (CNN-BiLSTM, MLP). Preprocessing steps included median imputation, normalization, and skewness correction, alongside strategies to mitigate class imbalance such as SMOTE and focal loss. Multiple cross-validation protocols were applied. Performance was evaluated using balanced accuracy, macro F1-score, and ROC-AUC, with statistical tests employed to compare models.

Results demonstrate that Classical (RF, XGB, SVM): Acc 0.806–0.810 (max 0.810), BalAcc 0.758–0.820 (max 0.820), Macro-F1 0.777–0.799 (max 0.799), ROC-AUC=0.936–0.939(max=0.939). Deep (MLP, CNN–BiLSTM): MLP- Acc 0.816, BalAcc 0.795, F1 0.798, AUC 0.936; CNN–BiLSTM- Acc 0.924±0.022, BalAcc 0.939±0.013, F1 0.922±0.020, AUC 0.987±0.006 (best overall)ROI-PCA reduced computational complexity, though in some cases at a marginal cost to accuracy. Comparative analysis confirmed the challenges in distinguishing mild fatigue from alert and severe states, consistent with prior literature.

This work contributes an integrated benchmark of classical and deep learning models

for driver fatigue detection on SEED-VIG, highlighting the trade-offs between interpretability, robustness, and efficiency. Future extensions will explore incorporating subjective self-reports and sentiment-based cognitive data, which may provide richer ground-truth signals beyond PERCLOS and enhance real-world deployment.

Keywords: Driver fatigue detection, EEG, EOG, PERCLOS, ROI-PCA, Random Forest, XG Boost, CNN-BiLSTM, cross-validation, class imbalance

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Armaghan Moemeni, for her invaluable support, guidance, motivation, and for always believing in me throughout the course of this dissertation. Her encouragement and constructive feedback have been instrumental in shaping this work.

I am also sincerely thankful to Ms. Alicia Falcón Caro from the School of Mental Health and Clinical Neurosciences at the University of Nottingham, for her advice and guidance while helping me work with EEG signals. Her input provided essential direction for the technical aspects of this research.

My thanks also go to the University of Nottingham, for providing me with the resources and academic environment that made this research possible.

Finally, I would like to extend my heartfelt appreciation to my family, whose love, support, and constant encouragement have been my greatest source of strength throughout this journey.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background and Motivation

Driver fatigue is a leading cause of traffic accidents worldwide and is widely recognised as a major public health and safety concern. The World Health Organisation (WHO) and national transport agencies consistently report that fatigue is implicated in approximately 20% of serious road accidents, with some studies attributing an even higher proportion of crashes to drowsy driving [1], [2]. In the United States, the National Highway Traffic Safety Administration (NHTSA) estimates that drowsy driving contributes to over 90,000 crashes and 50,000 injuries each year, with an economic burden exceeding $100 billion annually [3]. Similarly, the European Transport Safety Council has highlighted fatigue as a primary factor in long-haul and night-time driving accidents [4]. These statistics underscore the urgency of developing effective fatigue detection systems that can proactively monitor drivers and prevent accidents before they occur.

Self-report instruments such as the Karolinska Sleepiness Scale (KSS) are commonly used and have shown moderate validity when compared to objective indicators like EEG and reaction times [5]. However, in real-world or simulated driving environments, subjective reports of sleepiness have been found to not always align with actual driving impairment, suggesting limitations in self-awareness under fatigue. [6]

Vehicle-based measures, including lane departure or steering wheel variability, provide indirect cues of impaired driving performance but often detect fatigue only after dangerous behaviour manifests [7]. Behavioural approaches, such as monitoring eyelid closure (PERCLOS), yawning frequency, or head movements, offer more direct insights but are highly dependent on external conditions (e.g., camera placement, lighting, or driver obstruction) [8]. These limitations highlight that single-modality systems lack robustness and adaptability across diverse driving contexts.

In recent years, advances in physiological monitoring and artificial intelligence (AI) have enabled more objective approaches to fatigue detection. Electroencephalography (EEG), which records electrical activity of the brain, has been

shown to reveal distinctive spectral patterns during transitions from alertness to drowsiness, including increases in theta and alpha power alongside reductions in beta activity [9]. Electrooculography (EOG), which captures eye movement dynamics such as blink duration and saccadic parameters, has been shown to reveal reliable indicators of drowsiness. Studies report that increases in blink duration, delayed eyelid reopening, and decreased saccade velocity correlate with increasing fatigue levels in drivers and healthy individuals in controlled settings. [10][11] Furthermore, computer vision techniques can estimate PERCLOS and other facial features in real time, complementing physiological signals with non-intrusive behavioural cues [12]. By combining these modalities, researchers are moving toward multimodal fatigue detection systems [13] that leverage the complementary strengths of different signals.

Deep learning has significantly advanced driver fatigue detection by enabling automatic feature extraction and robust classification across complex multimodal data. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention-based architectures have demonstrated classification accuracies often exceeding 90% in controlled settings. [14], [15]. However, most systems remain in the laboratory stage and face significant barriers to real-world adoption. There is growing demand for lightweight, real-time fatigue detection solutions deployable on accessible hardware such as smartphones, which can integrate multiple sensors (e.g., wearable EEG, EOG, and camera input) into an inexpensive and scalable driver monitoring tool.)

1.2 Objective

Despite encouraging results in multimodal fatigue detection, several critical gaps remain:

1. **Single-modality limitations**: Many systems depend on a single input (EEG or vision). EEG-based methods, though sensitive to brain state, are highly susceptible to motion artifacts and noise, making robust feature extraction

challenging. Vision-based systems, meanwhile, are vulnerable to environmental factors such as brightness and occlusion, which can reduce detection performance. [16]

2. **Generalizability across drivers**: Physiological responses to fatigue vary across individuals, causing models to perform poorly on unseen subjects. Cross-subject studies confirm this variability, with accuracy dropping under leave-one-subject-out validation [17]. Transfer learning approaches like GDANN improve generalization, reaching >90% accuracy across subjects [18].

3. **Latency and resource constraints**: Deep learning models may achieve high offline accuracy but are computationally expensive, limiting their feasibility for real-time deployment on resource-constrained devices [19].

4. **Imbalanced and noisy datasets**: Fatigue events occur far less frequently than alert states, leading to skewed class distributions that can bias classifiers. He *et al.* (2024) directly addresses this by using a generative attention-based CNN to overcome class-imbalance and data scarcity in fatigue detection models [20].

5. **Limited deployment focus**: While much research emphasizes accuracy in controlled simulations, few studies address deployment feasibility on mobile platforms, which is essential for practical adoption.

This dissertation addresses these gaps by developing a multi-modal fatigue detection system that fuses EEG and EOG features, processes them into compact representations (via region-of-interest PCA and blink/saccade metrics), and implements machine learning models.

1.3 Summary of Approach

This study employs the SEED-VIG dataset [21], which contains EEG, EOG, and PERCLOS labelled fatigue states across 23 sessions. EEG signals are filtered and decomposed into delta, theta, alpha, beta, and gamma bands, followed by ROI-PCA

to condense spatially distributed channels into compact regional features. EOG signals are analysed for blink and saccade metrics, while PERCLOS thresholds are used to define three fatigue classes (alert, mild, severe).

Machine learning models, including Random Forest, XGBoost, and CNN architectures, are trained on these multimodal features. To address data imbalance, strategies such as oversampling (e.g., SMOTE), cost-sensitive learning, and focal loss are explored. Evaluation follows cross-subject protocols to ensure generalizability, and performance is reported using metrics such as balanced accuracy, macro-F1, and ROC-AUC.

1.4 Scope and Data Ethics

The project scope is limited to simulated driving data from the SEED-VIG dataset, ensuring ethical use of pre-collected, anonymized data. No new human experiments are conducted. Privacy considerations are addressed by designing the system for on-device processing, avoiding transmission of sensitive physiological data. Limitations include the absence of real-world road testing and restricted modalities (no heart rate or steering data).

1.5 Research Questions

RQ1: *How effective is a multi-modal machine learning approach combining EEG and EOG features for detecting driver fatigue?*

RQ2: *What is the impact of applying ROI-based PCA on EEG features in driver fatigue detection compared to using raw EEG features?*

Chapter 2: Literature Review

2.1 Introduction to the Literature Review

Driver fatigue is a critical issue for road safety, it has a substantial proportion of contribution towards road accidents worldwide. Traditional approaches to monitor fatigue, such as self-report questionnaires or vehicle-based behavioural measures such as steering wheel variability, lane deviation, they are limited in terms of reliability since they rely on subjective feedback or inclined toward influence of external driving conditions. Similarly Traditional approaches to detecting fatigue rely on subjective questionnaires (e.g., the Karolinska Sleepiness Scale) [22] These methods are easy to administer but depend on the driver's self-assessment and may underestimate fatigue. As a result, recent research has begun focusing on objective physiological and behavioural measurements, since these offer a more direct and reliable way to assess a driver's cognitive and physical state.

This review will cover a range of physiological and behavioural indicators that have been explored for detecting fatigue, including electrophysiological signals such as electroencephalography (EEG) and electrooculography (EOG), other bodily signals (e.g. heart rate, electromyography), as well as external cues like facial expressions, eye movements, and other perceivable actions. Recent studies categorize fatigue detection methods by signal modality, distinguishing uni-modal (single-signal) approaches from multi-modal approaches that fuse multiple data sources [23].

EEG-based methods can directly capture brain-state changes due to drowsiness but often require cumbersome sensors, while camera-based facial monitoring is non-intrusive but sensitive to lighting conditions [24]. By comparing such approaches, the review will identify common weaknesses (such as susceptibility to noise or user discomfort) and pinpoint gaps in the current literature, areas where fatigue detection is still unreliable or where certain contexts (like real-world driving conditions) are not well addressed. Notably, prior comprehensive reviews have concluded that no single fatigue metric is sufficient, fatigue is a complex phenomenon that often requires

multiple features to detect reliably, meaning systems that integrate both “physical” cues (eye, face, head position) and “biological” signals (EEG, heart rate, etc.) tend to perform better [25]. This literature review will thus pay special attention to such hybrid approaches, examining how multi-modal fusion has been used to overcome the limitations of individual signals.

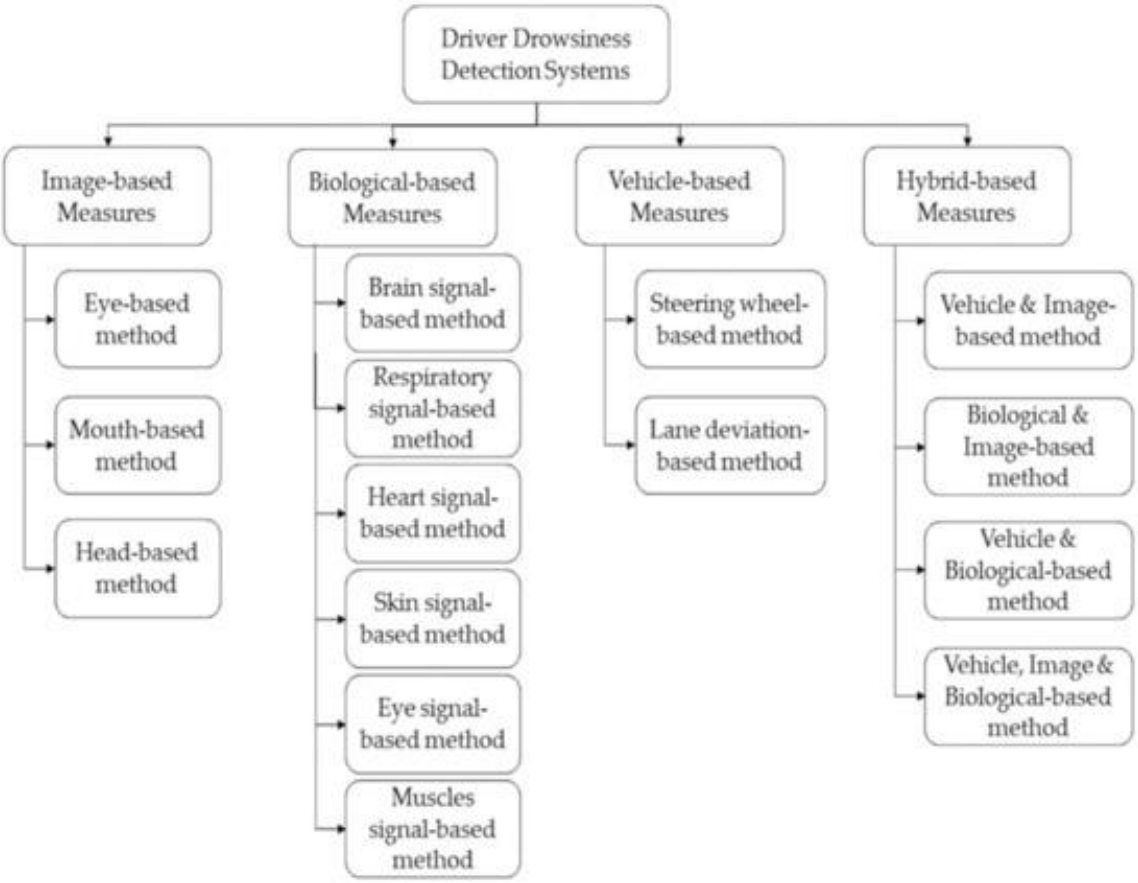

Figure 1: Driver drowsiness detection measures.[38]

2.2 Fatigue Detection: Context and Importance

Fatigue (especially driver drowsiness) is a critical safety issue that provides the motivation for extensive research into detection methods. Road accidents caused by fatigued or drowsy driving are a significant concern worldwide. The World Health Organization reports that about *1.3 million people die each year* in road traffic crashes globally [26]. While not all are due to fatigue, sleepiness at the wheel is a major contributor: studies estimate that between *3% and over 30%* of road traffic accidents

globally are related to driver sleepiness or fatigue [26]. In other words, fatigue contributes to a substantial fraction of crashes, especially in certain contexts (e.g. long-haul driving or nighttime driving). In the USA alone, drowsy driving leads to hundreds of fatal crashes annually. The National Highway Traffic Safety Administration (NHTSA) recorded 684 fatalities in 2021 from accidents involving drowsy drivers [27]. However, the true scope of the problem is likely much larger, since fatigue often goes underreported. For instance, an AAA Foundation study estimated that around *328,000 crashes per year* in the U.S. involve drowsy driving (about **3**.5 times the number reflected in police reports), with an estimated *6,400 of those crashes resulting in fatalities* [4]. This gap shows why fatigue-related crashes are so hard to identify and record after they happen, underscoring the importance of having proactive systems to detect fatigue before an accident occurs. Beyond the tragic human cost, drowsy-driving accidents also incur huge economic losses NHTSA estimates fatigue-related injury and fatal crashes cost society on the order of $109 billion annually (not including property damage) [4]. Clearly, there is a strong imperative to detect and prevent fatigue-related impairment, especially in transportation, to save lives and reduce costs.

2.3 Traditional methods for assessing fatigue

Traditional methods have significant limitations, which partly explains why modern research is turning toward physiological and multi-modal solutions [28]. Historically, driver fatigue was often monitored through subjective and indirect means. For example, drivers might self-report their level of sleepiness using questionnaires or rating scales (such as the Karolinska Sleepiness Scale). While simple, these self-reported measures are not very reliable, drivers may not accurately perceive their own level of fatigue or may underreport it (whether intentionally or not). In fact, subjective fatigue assessments can be biased or easily manipulated [29]. Another traditional approach is to monitor vehicle behaviour for signs of driver impairment. Many modern vehicles include safety features that watch for patterns like lane deviations, erratic steering corrections, or slow reaction times, and trigger alerts (e.g. coffee cup warning

icons or audible alarms) if drowsiness is suspected. Such vehicle-based metrics (sometimes called performance measures) can detect some symptoms of fatigue; for instance, a tendency to drift out of lane or inconsistent speed control may indicate a lapse in driver attention. However, vehicle behaviour monitoring is an indirect proxy and can be influenced by external factors. Road curvature, weather conditions, and vehicle type can all affect driving patterns independent of the driver's alertness. Research has noted that vehicle-centric features depend on environmental and driver-specific factors (like weather or individual driving style), making them less reliable, especially in dynamic urban environments [25]. In summary, traditional fatigue detection methods, whether self-reports or simple vehicle telemetry, suffer from either subjectivity or context-dependence, and they may fail to provide a timely warning before an accident occurs.

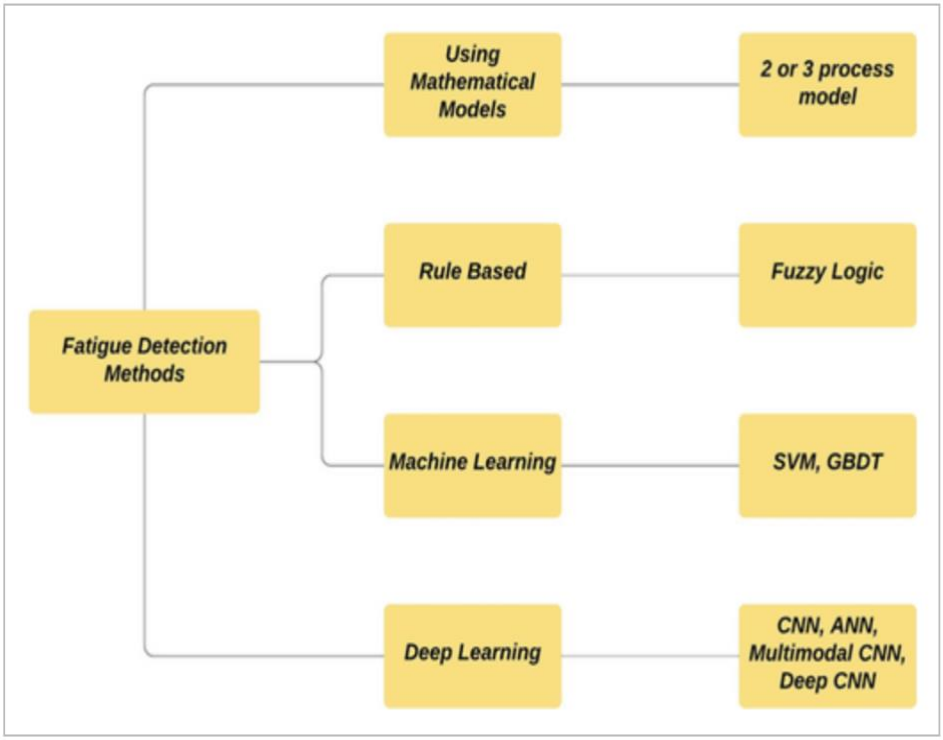

Figure 2: Classification of fatigue detection methods ranging from mathematical and rule-based approaches to machine and deep learning models [25].

Given these shortcomings, recent years have seen a shift toward physiological and multi-modal approaches for fatigue detection. The appeal of physiological signals

(such as EEG brainwaves, EOG eye movement signals, heart rate variability, etc.) is that they provide a more direct, objective measure of the driver's state. These bio signals can reflect the internal changes associated with fatigue (e.g. slow eye blink patterns, changes in brain alpha wave activity, or heart rate slowdown) in real time, without relying on the driver's self-awareness. Wearable or in-cabin sensors can now track many of these indicators continuously. Research has shown that wearable physiological sensors allow for objective monitoring of fatigue, helping to overcome the biases that come with self-reported assessments [23]. For example, an EEG headband can detect the onset of microsleep or lapses in attention that a driver might not consciously notice [30]. Similarly, a camera-based system monitoring eye blink duration or PERCLOS (percent eye closure) can catch early signs of drowsiness [31]. Each of these physiological or behavioural signals has its own strengths and weaknesses (EEG is very sensitive but can be intrusive, camera vision is non-intrusive but affected by lighting, EOG can directly quantify eyelid droop but requires electrode contact, etc.). As a result, there is a growing consensus that no single signal is sufficient, and that multi-modal fusion of several indicators yields the most robust detection [25]. By combining multiple data streams, for instance, integrating brainwave data with eye-tracking and steering behaviour, a fatigue detection system can cross-validate cues and reduce false alarms. Indeed, studies report that multi-modal models outperform single-modal ones, achieving higher accuracy in detecting fatigue than any individual signal alone[25]. Multi-sensor approaches can leverage the complementary nature of different fatigue symptoms: if a driver's EEG indicates drowsiness but their face is momentarily not visible to a camera (or vice versa), the system can still detect fatigue from the other inputs.

Furthermore, fusing driver-centric biological signals with vehicle telemetry can cover both the cause and effect of fatigue, one recent review suggests that combining driver physiological indicators with vehicle behaviour data produces a stronger system than either alone [32]. In summary, physiological and multi-modal detection methods are now preferred because they offer objective, real-time insight into driver alertness and

can overcome the limitations of traditional single-source measures. This literature review will delve into these modern approaches, evaluating how effectively they address the critical problem of fatigue-related accidents and where further improvements are needed to ensure drivers stay safe.

2.4 EEG-Based Fatigue Detection

Electroencephalography (EEG) measures brainwave activity and it is widely regarded as one of the most reliable indicators of driver drowsiness [33]. Cognitive fatigue and transitions toward sleep are reflected in characteristic EEG changes, making EEG a direct window into a driver's alertness.

Key EEG features Fatigue is usually marked by a rise in lower-frequency brain activity, particularly in the theta (4-8 Hz) and alpha (8-12 Hz) bands, along with a drop in higher-frequency beta activity (above 12 Hz) [34]. Studies have shown that as drivers become drowsy, theta and alpha wave power significantly increase, while beta (associated with active concentration) drops off [34]. Consequently, simple frequency-domain metrics like the ratio of (θ+α)/β have proven effective for distinguishing drowsy vs. alert states [36]. For example, Eoh *et al.* introduced θ/α and β/α ratios, and Jap *et al.* found that a combined (θ+α)/β index correlated well with driver drowsiness [36]. Other EEG-based features that have been explored include nonlinear metrics (such as fractal dimension), entropy measures, and functional connectivity between different brain regions [34]. However, most recent studies still rely on spectral power features or their variations, usually focusing on the frequency bands that are most sensitive in each brain region [34].

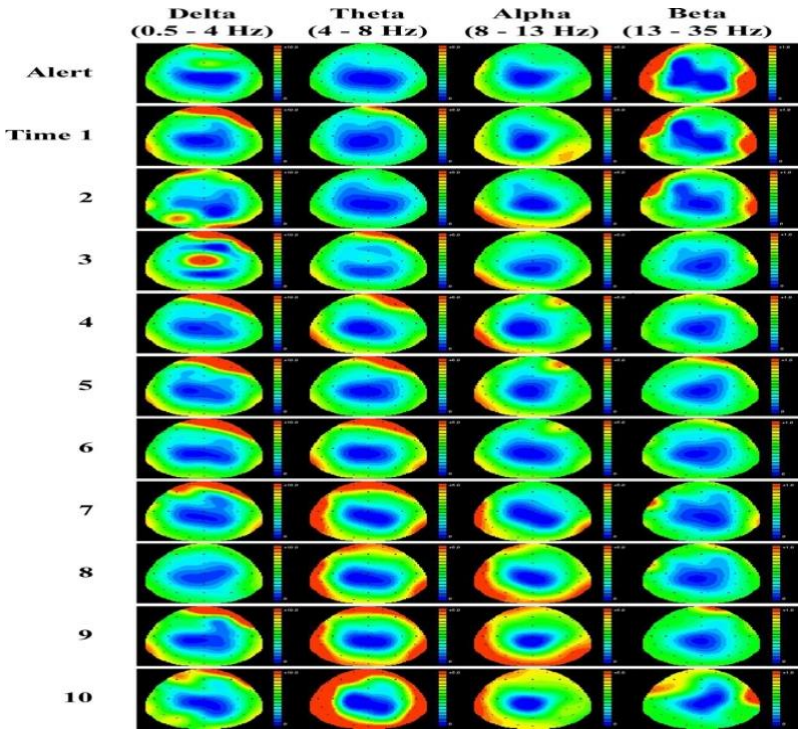

Figure 3: Brain topography of alert baseline [36]

2.4.1Traditional machine learning models: Early EEG-based drowsiness detectors employed classifiers such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) using handcrafted features. For example, power spectral densities in the delta, theta, and alpha bands have been fed to SVM and ANN classifiers to distinguish alert from drowsy epochs. Chan *et al*. validated this approach using a consumer-grade MUSE EEG headband, achieving classification accuracies of 87.8% with SVM and 87.9% with ANN [30] Other studies report even higher accuracy when optimizing feature selection and classifiers: a review by Ibrahim *et al.* notes that combining entropy-based features with an SVM or Random Forest classifier yielded up to 98% detection accuracy on driver fatigue data [37]. Feature selection is critical, recent comparative studies indicate that entropy features and frequency ratios can outperform raw band powers [38]. Nonetheless, a limitation of these classical approaches is the need for extensive feature engineering and tuning for different individuals or contexts.

2.4.2 Deep learning approaches: In the past few years, there has been a surge in deep neural network models that learn discriminative features from EEG signals automatically. Convolutional Neural Networks (CNNs) have been applied to raw EEG or spectrogram images to capture frequency-spatial patterns of drowsiness [38]. Recurrent networks like LSTM have been used to model temporal dynamics of EEG sequences [38]. For example, dua *et al* [39] (as cited in Albadawi *et al.*) converted EEG signals into spectrogram images and used pre-trained deep CNNs (AlexNet, VGG) combined with LSTM, achieving 94.3% accuracy in detecting driver drowsiness [38]. A notable innovation is the use of Graph Convolutional Networks (GCN) to exploit the connectivity between EEG channels (treating electrodes as nodes in a graph). Wang *et al.* proposed an attention-based multi-scale CNN with dynamical GCN (AMCNN-DGCN) that learns both temporal filters and an adaptive adjacency matrix for EEG channels [33]. Their 24-channel EEG model achieved 95.7% accuracy on a simulated driving fatigue experiment, outperforming several baseline EEG models [33]. The GCN component automatically learns inter-channel relationships (reflecting functional brain connectivity changes during fatigue), which improved classification performance [33]. Such end-to-end deep models eliminate manual feature extraction and can adapt to subtle changes in EEG patterns. However, they typically require larger training data and have higher computational cost, an issue for real-time in-vehicle deployment. Researchers are exploring model compression and efficient architectures to address this [23].

2.4.3 Practical considerations and limitations: While EEG is highly sensitive to drowsiness, its use in real cars faces challenges. Standard EEG setups with many wired electrodes are intrusive and cumbersome, deterring real-world drivers. Movement artifacts and electromagnetic interference in a moving vehicle can also corrupt EEG signals. One key development is wearable EEG headbands with dry electrodes, which significantly improve portability. Consumer EEG devices (Muse, NeuroSky, etc.) use 35 electrodes and wireless data transfer, making continuous EEG monitoring more feasible in cars. Studies have validated that these low-density

systems can still detect fatigue-relevant EEG dynamics, though with slightly reduced signal quality. Another limitation is inter-subject variability, EEG patterns of fatigue can differ between individuals. Some recent works apply transfer learning or subject-specific calibration to maintain accuracy across drivers [23][40]. In summary, EEG-based methods provide high accuracy and can predict fatigue at an early stage, as brain activity often changes before any noticeable behavioural signs appear. However, practical limitations remain a challenge. While newer hardware such as wearable dry EEG devices is reducing intrusiveness, maintaining reliable performance in real driving conditions, where motion and noise are common, remains difficult. These limitations highlight the value of combining EEG with other modalities to create a more robust and reliable fatigue detection system, as discussed later.

2.5 EOG and Eye-Based Fatigue Detection

As drowsiness sets in, a driver's eyes exhibit noticeable changes, longer blinks, frequent closure, slow rolling movements, which can be captured via electrooculography or camera-based methods. Electrooculography (EOG)involves electrodes placed around the eyes to measure cornea-retinal potentials as the eyes move or blink [41]. Camera-based eye tracking analyses video of the driver's eyes and face to extract blink metrics and eye closure information. Both approaches focus on ocular behaviours that are well-known indicators of fatigue.

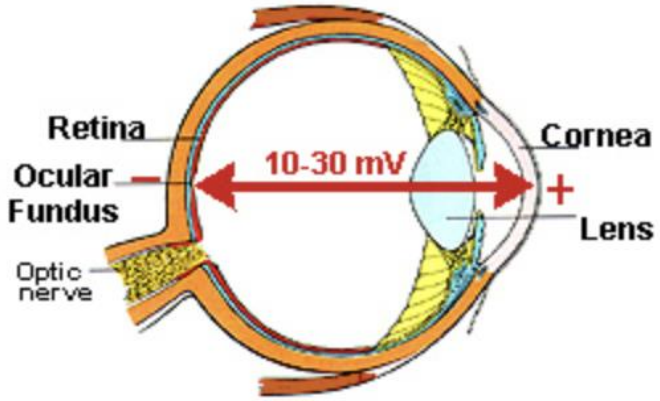

Figure 4: Principle of electro-oculography (EOG). A corneo-retinal potential (10–30 mV) shifts with eye movement and is detected by surrounding electrodes. Principle of electro-oculography (EOG). A corneo-retinal potential (10–30 mV) shifts with eye movement and is detected by surrounding electrodes. [41]

2.5.1 Key eye metrics: A widely used measure is PERCLOS (Percentage of Eyelid Closure), defined as the percentage of time (per minute) that the eyes are at least 80% closed [42]. Extensive evidence shows PERCLOS increases significantly with driver drowsiness [42]. Many modern driver monitoring systems (including commercial ones) use PERCLOS as a primary drowsiness indicator. Other common features are blink frequency (blinks per minute) and blink duration drowsy individuals blink more slowly and often have prolonged eye closures (>0.51 s) termed micro-sleeps [43]. The eye aspect ratio (EAR), a real-time image metric computed from eye landmarks, drops toward zero as the eyes close, and can be thresholded to detect blinks [44]. Patterns of eye movement can also signal fatigue: in early drowsiness, blinks become not only slower but also more erratic, and slow eye movements (SEM) (characterized by rolling motion of the eyes) may occur [45]. EOG signals capture both blinks (spikes in the vertical EOG channel) and lateral eye movements (horizontal channel deflections), enabling computation of blink rate and duration very precisely in the time series [8].

For example, an EOG-based system might count eye-blink peaks and measure their width to quantify slow blinks [8].

Table 1: Wierwille and Ellsworth drowsiness scale [8]

Levels	**Verbal Description**
1	Not drowsy
2	Slightly drowsy
3	Moderately drowsy
4	Significantly drowsy
5	Extremely drowsy

2.5.2 Camera-based eye monitoring: Non-intrusive computer vision techniques have advanced greatly with deep learning, allowing reliable detection of facial landmarks (eyes, pupils, mouth, etc.) even under variable lighting [46]. Many studies use an infrared camera for night vision, positioned on the dashboard. Classic approaches applied Haar cascades or template matching to detect eyes, then measured blink duration and PERCLOS. Recent approaches leverage CNNs: for instance, a two-stream CNN that locates eyes and mouth and extracts both static features (from video frames) and dynamic features (from optical flow) to detect states like eyes-closed or yawning [47]. They combined eye closure duration, frequency of head nodding, and yawning detection, achieving high accuracy (97% on a public drowsiness dataset) in classifying driver state. Overall, camera-based methods can be very accurate under good conditions reported detection accuracies range from 72% up to 99% in research settings [38]. They are also non-intrusive (no body contact) and low-cost (just a camera). The limitations, however, include sensitivity to lighting (poor performance in darkness without IR illumination or if sunlight causes glare) and occlusions (e.g. sunglasses, or if the driver's face turns away). Furthermore, calibration might be

needed to handle different eye shapes or glasses. Despite these issues, vision-based eye trackers are attractive for real vehicles due to their ease of use[38].

2.5.3 EOG-based methods: Using EOG for driver monitoring has the advantage of working in the dark (since it measures electrical signals, not images) and providing continuous, high-resolution measurements of eye activity. Research prototypes have used adhesive electrodes on the forehead and beside the eyes to capture blinks. For example, Kolodziej *et al.* demonstrated an EOG system to detect fatigue in office workers, using blink patterns as features [48]. They found EOG signals could reliably measure prolonged blinks associated with fatigue in screen-based tasks. In driving simulators, EOG-based blink detection has shown high correlation with self-reported fatigue levels. The downside is that electrode setup is somewhat intrusive for drivers (electrodes near the eye can be uncomfortable and drift over time). Motion artifacts (head movement) also affect EOG. Some newer approaches integrate EOG electrodes into glasses or headbands for easier use [49]. Notably, EOG and EEG can be recorded together using the same headset, many EEG headsets pick up eye movement artifacts which can be repurposed as EOG signals. This dual-use concept is valuable for hybrid EEG/EOG monitoring (discussed later). Overall, eye-based measures (whether via camera or EOG) are excellent at detecting the *visible manifestations* of fatigue (eye closure, blinks, yawns), but they might only reflect fatigue once it becomes moderate. A driver could be cognitively fatigued even if their eyes are still open. Thus, eye metrics are often used in tandem with brain or cardiac measures to improve early detection.

2.6 Cardiac signals (ECG/HRV and PPG): Drowsiness is accompanied by changes in autonomic nervous system balance, which affects heart rate. Electrocardiogram (ECG) or photoplethysmography (PPG) sensors can capture heart activity [50].

From these sensors, researchers derive heart rate variability (HRV) metrics, most notably the spectral power in high-frequency (HF) and low-frequency (LF) bands of

the heartbeat interval series, as well as the LF/HF ratio [51]. Fatigue tends to elevate parasympathetic (vagal) tone, increasing HF power and altering the LF/HF ratio. Several studies confirm HRV features correlate with driver fatigue; for instance, Fujīwara et al. (2019) validated an HRV-based detection method against EEG and reported sensitivity as high as 92% in detecting pre-sleep drowsiness. However, achieving reliable performance remains challenging in real driving conditions due to motion artifacts [50]. These results show that heart-based measures can detect fatigue, but typically with lower accuracy than EEG or direct eye measures. The advantages of heart signals are ease of collection (many cars now have driver heartbeat monitors, or one can wear a smartwatch) and low intrusiveness. They can also operate continuously without calibration. The drawbacks include individual variability (heart rate is influenced by fitness, stress, etc.), and slower response heart rate changes may lag behind the onset of drowsiness or be subtle in short driving sessions. In addition, if a driver consumes caffeine or is in a stressful situation, HRV can be affected independently of fatigue [52].

2.6.1 Electromyography (EMG): Muscle activity can change with fatigue, particularly as drivers become more relaxed and have reduced muscle tone. Surface EMG electrodes can detect such changes. A novel approach by Satti *et al.* (2021) placed EMG sensors on drivers' forearms (muscles used for gripping the steering wheel) [53]. As drowsiness increased over a one-hour simulated drive, they observed a significant *decrease* in grip muscle EMG activity, reflecting the driver's loosening grip and reduced physical engagement [53].

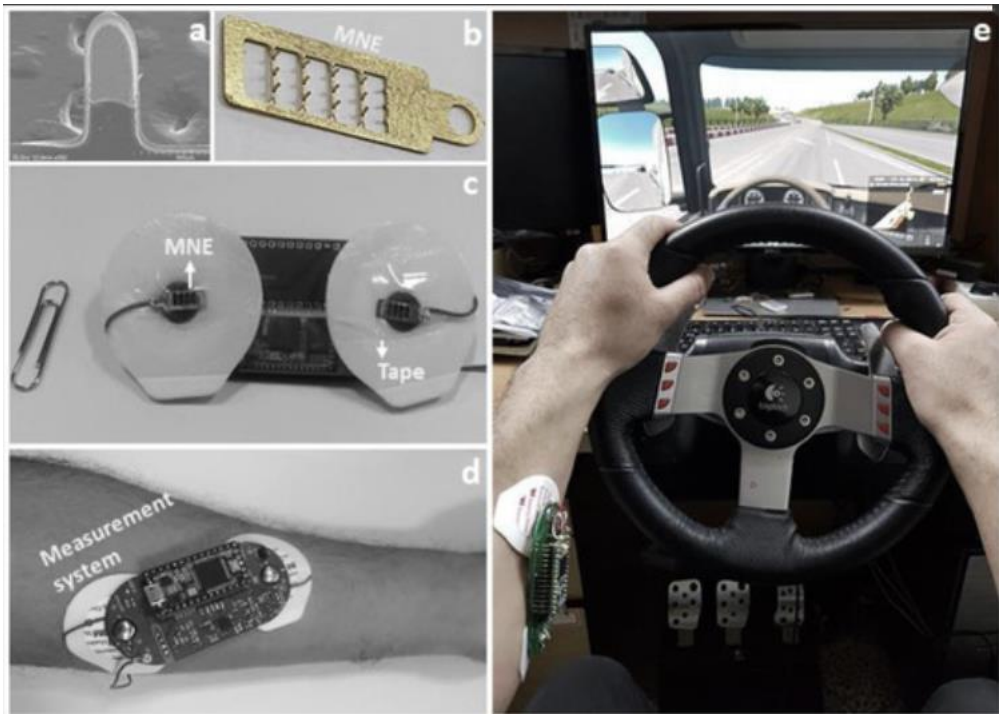

Figure 5: Microneedle array electrode (MNE) system for EMG-based drowsiness detection, showing device, forearm placement, and driving simulator setup [53].

They developed a microneedle based dry electrode system embedded in the steering wheel, which, coupled with a real-time algorithm, could detect drops in EMG amplitude as a trigger for drowsiness alerts [53]. This proof achieved good detection performance and showed that EMG signals (even from non-facial muscles) contain useful drowsiness information. Another potential EMG application is measuring head tilt or nodding through neck muscle activity or monitoring facial muscles (e.g. jaw EMG) for yawning [54]. The advantage of EMG is that sensors can be integrated into existing car interfaces (wheel or seat) unobtrusively. And unlike camera vision, EMG isn't affected by lighting. The downsides are noise and variability, EMG is susceptible to movement artifacts, and different drivers have different baseline muscle tension. It also only captures relatively late-stage fatigue (when muscles relax, or the driver starts to slump).

2.6.2 Skin temperature and other signals: Drowsiness is often accompanied by subtle drops in core body temperature as the body prepares for sleep. Some studies have used infrared thermography to monitor facial skin temperature (e.g. around the nose or inner canthus of the eye) as an indirect sleepiness measure. In theory, a dropping temperature or increasing peripheral skin warmth (due to vasodilation) could indicate the circadian phase where sleep pressure is high. However, results have been mixed, and environmental temperature plays a confounding role [55]. Similarly, galvanic skin response (GSR) or electrodermal activity could reflect reduced alertness (lower sympathetic arousal), but this is more often related to stress rather than fatigue. In practice, these "other" signals (temperature, GSR, respiration rate, etc.) are not strong standalone detectors, but they can add context in a multi-sensor system. [56] For example, a warm cabin environment combined with a lowering heart rate could strongly suggest impending drowsiness. Recent wearable health trackers that measure skin temperature and blood oxygen could feed additional inputs to fatigue detection algorithms, but research in 2020+ has focused more on the primary signals (brain, eyes, heart) for higher fidelity.

In summary, alternative physiological signals like ECG/HRV and EMG provide less intrusive but indirect measures of fatigue. They tend to be easier to integrate into vehicles (steering wheel sensors, seat sensors, wearable bands) and can operate continuously without much user compliance. Their main weakness is lower sensitivity or specificity to drowsiness compared to brain or eye signals.

2.7 Facial and Behavioural Indicators

Visual observation of the driver's face and behaviour is a natural approach to detect drowsiness, like how a passenger might notice a driver yawning or head-nodding. Facial features include eye state (open or closed), mouth state (yawning), head pose (upright or drooping), and facial expression changes. These are extracted via camera-based computer vision. Many modern driver-monitoring cameras (DMCs)

use infrared illumination and inward-facing cameras to continuously analyse the driver's face.

2.7.1 Yawning and mouth movement: Yawning is a strong indicator of fatigue. Vision-based approaches detect yawns by monitoring the mouth opening. One simple metric is the frequency of yawns (counted via detecting when the mouth remains widely open for a certain duration). Researchers have used motion analysis or trained classifiers to distinguish yawning from normal talking. For instance, Abtahi *et al.* developed a yawning detection method using a cascade of regression-tree classifiers on mouth aspect ratio and achieved 95% accuracy in detecting yawns [57]. In recent systems, mouth features are often combined with eye features: e.g. a fuzzy logic system might take PERCLOS, eye closure duration, and average mouth open time as inputs to infer drowsiness level. Deep learning models (like CNNs) can be trained on facial landmark images to automatically learn a "yawn" feature one study used a multi-task CNN to output both an eye state and mouth state, then fused those to decide drowsiness, yielding 93-95% accuracy [58].

2.7.2 Head pose and nodding: When drivers begin to microsleep, their head often droops or nods briefly. Monitoring head pose (pitch angle in particular) can thus provide an alert. A camera-based system can estimate head orientation via face landmarks or inertial sensors (IMU) on a wearable. Head nodding frequency (short downward movements of the head) has been used as a feature in some systems [59]. For example, Ghourabi et al. (2020) proposed a system that jointly monitors yawning, blinking, and nodding behaviours to detect driver drowsiness. By integrating these complementary facial cues, the system improved detection reliability compared to using any single feature in isolation. Head pose can also serve as a fallback if the driver's face is not fully visible to the camera (e.g. turned away), a large head tilt might still be measurable and indicative of drowsiness [60].

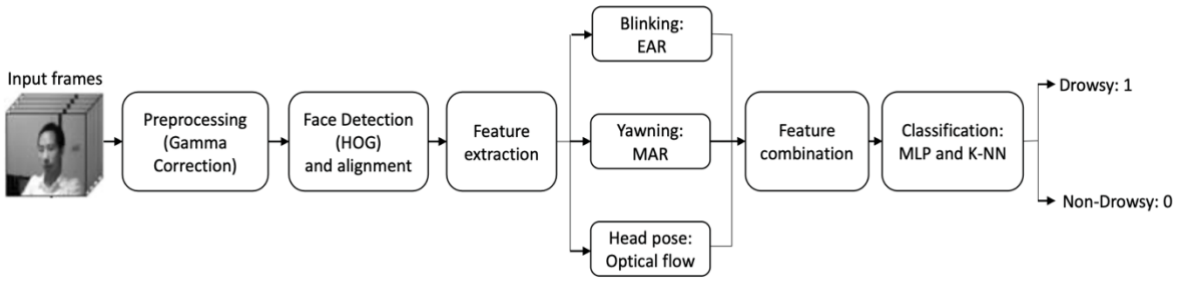

Figure 6: Flowchart of a vision-based drowsiness detection pipeline using eye aspect ratio (EAR), mouth aspect ratio (MAR), and head pose as behavioural indicators. [60]

2.7.3 Behavioural cues beyond the face: Behavioural cues beyond the face have also been studied. Some systems classify driving behaviours such as steering wheel variability, lane deviations, and braking delay as indicators of fatigue. However, these are vehicle-based metrics, and while useful, they are influenced by external factors and are less reliable than direct ocular measures de [61].

2.7.4 Strengths and limitations: Vision-based facial monitoring has become highly effective with modern image analysis techniques. It is non-intrusive and well accepted by users (already incorporated in some commercial driver-assistance systems). High accuracies (90%) are reported, especially when multiple facial features are combined and when evaluated on training datasets [38]. One strength is that camera systems can also serve multiple functions (e.g. identity recognition, distraction detection), making them multi-purpose for driver monitoring. However, a big limitation is robustness, changes in illumination (tunnels, night-time without IR, sun visors casting shadows) can degrade performance. Occlusions like sunglasses or face masks block key features (eyes, mouth). Moreover, vision algorithms may need to be tuned for each individual's appearance and may struggle with diverse ethnicities or facial hair conditions unless well-trained on broad data. Privacy is also a consideration, though most in-car systems process images locally and do not record video. Despite these challenges, facial behaviour monitoring is a cornerstone of fatigue detection due to its intuitive and direct measurement of drowsiness signs. The current trend is to use

deep learning for more reliable detection under varied conditions (for example, using thermal cameras at night to track eyes when visible light fails) [62][63].

2.8 Multi-Modal Fusion Approaches

No single measure is perfect, each has failure modes, so researchers have increasingly explored multi-modal fusion of various sensors to improve detection robustness. Multi-modal fatigue detection systems combine two or more of the mentioned signal types (EEG, EOG, heart rate, facial video, steering inputs, etc.), leveraging the strengths of each. The rationale is that drowsiness manifests in multiple physiological and behavioural channels; by fusing information at multiple levels, one can achieve more accurate and earlier detection than any single channel alone [64].

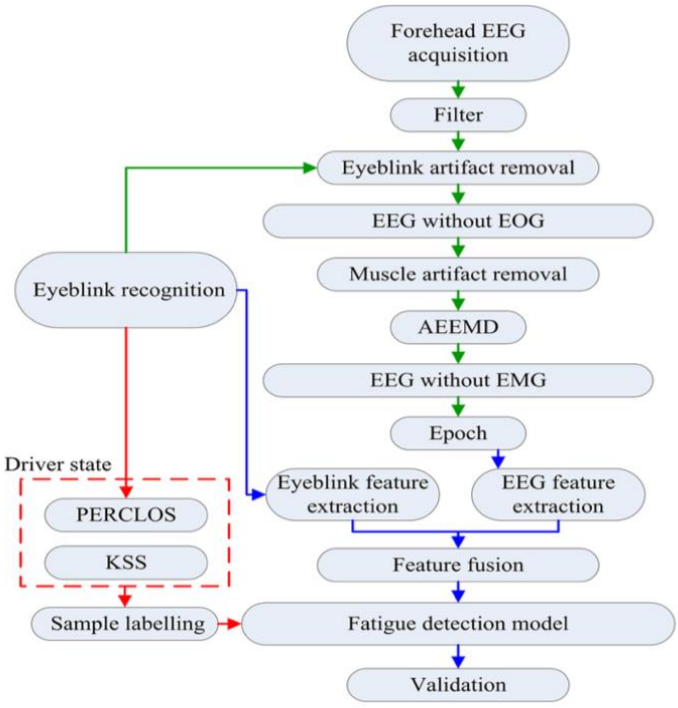

Figure 7: Overall workflow of the proposed multimodal fatigue detection pipeline, fusing forehead EEG and eye-image features for classification [65]

Fusion techniques: Fusion can occur at the feature level (concatenating features from different modalities into one input vector for a classifier) or at the decision level (each

modality yields an independent drowsiness estimate which are then combined via rules or weighted voting). Recently, deep learning models have been used to perform learned fusion, e.g. a neural network with separate subnets for EEG and for eye images whose outputs are merged in a later layer. This allows the network to find complex correlations between modalities. For instance, Min *et al.* (2022) designed a pipeline fusing forehead EEG signals with camera-based facial features in real-time. Their system used a CNN to process EEG and another to process eye images, then a fully connected layer to fuse features, resulting in improved accuracy over EEG-only or vision-only systems (they report 94% accuracy vs 89% for single modality) [65]. Similarly, Zheng *et al.* (2019) combined EEG and EOG entropy features: by fusing these two physiological signals, they achieved higher classification accuracy than using EEG alone, highlighting that eye-blink information (from EOG) complements brain-wave features [66]. Another notable example is by Khushaba *et al.* who fused EEG, EOG, and ECG signals using a wavelet-based feature extraction and employed linear discriminant analysis to classify five levels of drowsiness [67]. Their multi-sensor approach significantly improved sensitivity to mild drowsiness compared to any single signal, because EEG captured early neural fatigue, EOG captured eye closure, and ECG captured autonomic changes, together providing a comprehensive view of the driver's state.

2.8.1 Benefits of multi-modal systems: The primary benefit is improved accuracy and reliability. As seen in many studies, hybrid systems can achieve very high accuracies (often 95%+). In Albadawi *et al.'s* review, the reported accuracy range for hybrid systems was 79.99%, compared to as low as 62% for single vehicle-based systems and up to 97% for single physiological systems [38]. The fusion helps because if one modality is noisy or momentarily uninformative, the others can compensate. For example, if bright sunlight temporarily obscures the camera (losing eye data), the EEG or heart rate might still indicate drowsiness. Conversely, if EEG signal quality degrades due to motion, the facial cues can take precedence. Multi-modal systems are also more robust to false positives, a blink alone might be just a normal blink, but a blink

accompanied by a theta burst in EEG and a drop-in heart rate is almost surely a sign of fatigue. Hence, fusing modalities can reduce false alarms by requiring concurrent evidence from multiple sources.

2.9 Real-world deployment challenges: Despite their performance advantages, multi-modal systems face several challenges. First, invasiveness/intrusiveness can increase if multiple sensors are required. Wearing an EEG headband plus EOG electrodes plus a chest ECG might be impractical for everyday drivers. Engineers thus look for integrated solutions (e.g. an EEG headset that also captures EOG and perhaps has an PPG sensor on the forehead [68] Second, cost and complexity are concerns, more sensors and more complex data processing (often requiring onboard computation or cloud connectivity) can raise system cost and power consumption [69]. The recent emphasis on IoT and edge computing has spurred architectures where multi-sensor data is streamed to a smartphone or cloud platform for analysis [70]. Abbas *et al.* (2021) discuss multi-sensor smartphone-based platforms that offload heavy deep learning computations to the cloud, which can make in-car implementation easier [70]. However, reliance on connectivity can be a drawback in areas without reliable network coverage. Another issue is synchronization of data streams EEG, camera, and ECG all have different sampling rates and latencies. Ensuring that features are aligned in time (so that, e.g., a blink is matched with the corresponding brain signal at that moment) is non-trivial and requires careful system design. There can also be interference between devices: In multimodal wireless systems, careful configuration is required to prevent potential signal interference or crosstalk between devices, such as EEG and EOG sensors. Finally, user acceptance is a factor while a camera and maybe a smartwatch are acceptable, many drivers would balk at wearing an EEG cap for daily driving. Thus, a lot of current research aims to minimize sensor intrusiveness [71] while still reaping the benefits of multiple modalities.

Nonetheless, multi-modal approaches are arguably the future of driver monitoring. They align with the general trend in automotive safety of using sensor fusion (similar to how ADAS uses fusion of cameras, radar, lidar for reliability). In the context of fatigue detection, the literature clearly shows that hybrids outperform single-sensor systems in diverse conditions [29][38]. A practical implementation might use a camera + wearable (like an ear-clip PPG or a headband) to balance intrusiveness and information gain. Recent prototypes, for example, combine an EEG headband with an infrared camera and have demonstrated high drowsiness detection rates in on-road tests [23]. As vehicles move toward semi-autonomous operation, robust driver state monitoring will be crucial, multi-modal systems provide the redundancy and accuracy needed for such applications.

2.10 Datasets in Fatigue Detection

Here are some of the most widely used publicly available datasets for driver fatigue detection. They vary in scale, sensing modalities, and experimental design, reflecting different trade-offs between ecological realism, sensor intrusiveness, and labelling.

Table 2: Datasets on Fatigue Detection.

Dataset	Participants	Signals/Modalities	Duration / Samples	Labels	Key Notes
SEED-VIG [72] Zheng & Lu, 2017; Dataset	23 subjects	EEG (17 channels, 1000 Hz), EOG (forehead), eye-tracker	2 hrs per subject in driving simulator	PERCLOS (01 vigilance scale)	Benchmark corpus; spectral power + DE features; EEG+EOG fusion improves accuracy;

					strong subject-dependence noted.
DROZY[73] Massoz et al., 2016	14 adults (30h wakefulness protocol)	EEG, EOG, ECG, EMG, NIR video (30 Hz)	10-min psychomotor vigilance tasks at baseline, 20h, 30h awake	Karolinska Sleepiness Scale (KSS)	Genuine fatigue via sleep deprivation; multimodal (brain, ocular, cardiac, visual).
SUST-DDD [74] Dataset	19 drivers (real road, 2,074 clips)	Smartphone front-camera videos (10 s each)	Daytime & nighttime naturalistic driving	Alert vs. Sleepy (rater consensus)	Real-world dataset; behavioural cues (eye closure, yawning, nodding).
NTHU-DDD[75] Dataset	36 subjects	Infrared video (simulator)	Short clips (12 min)	Pre-defined behaviours (normal, blink, yawn, nodding)	Widely used for computer vision; includes day & night conditions; staged scenarios.

2.11 *Identified Gaps and Research Motivation*

2.11.1 Subject and Session Dependence

Models often perform well on the same subject but degrade sharply with new drivers due to EEG non-stationarity and individual variability. Silveira *et al.* (2019) reported a 40% drop in “sleepy” detection when moving from subject-specific to subject-independent testing [76]. Transfer learning, domain adaptation, and zero-calibration CNNs (e.g., Paulo *et al.*, 2021) have been explored, as well as ensemble RVFL approaches [77]. Yet, as Feng *et al.* (2025) emphasize, truly calibration-free systems remain “very challenging,” and many works still rely on subject-dependent validation, yielding overly optimistic results.[78]

2.11.2. Data Imbalance and Rare Events

Real-world fatigue is rare, producing heavily imbalanced datasets. Models often achieve high accuracy by predicting “alert” but fail on drowsy cases [79]. Oversampling (SMOTE, GANs) has been applied [80] as well as cost-sensitive losses like focal loss [81]. However, synthetic samples risk overfitting, undersampling discards data, and extreme drowsiness may be missing entirely [82]. Metrics like F1-score and AUROC are increasingly used, but robustly detecting rare microsleep events remains unresolved.

2.11.3. Sensor Intrusiveness and User Acceptance

EEG/EOG are accurate but intrusive, while vision-based methods are less obtrusive but vulnerable to lighting and occlusion [78][83]. Less-intrusive solutions include high-frame-rate cameras [84], portable single-channel EEG [78], and behind-the-ear wearables [84]. Alternatives like rPPG and steering sensors also exist. Still, wearables suffer lower SNR and motion artifacts [85], and vision raises privacy concerns. The gap is balancing accuracy with comfort and acceptance**.**

2.11.4. Overfitting and Inflated Accuracy Reports

Reported accuracies of 95-99% are often linked to subject-dependent tests on staged datasets like NTHU-DDD [86]. Some studies test on independent datasets to reduce overfitting. Still, many papers lack robust validation, often due to limited data, and no large-scale benchmark exists for real-world, cross-subject testing. This methodological gap continues to inflate reported performance.

Table 3: Comparative summary of representative driver drowsiness/fatigue detection

Study (year) — from your refs	**Domain**	**Approach / model**	**Strengths**	**Limitations**
Zheng & Lu (2017) [66]	Multimodal (EEG+EOG)	Differential-entropy EEG + forehead-EOG features; temporal CRF/CCNF; vigilance annotated via eye-tracker PERCLOS	Clear multimodal gain; compact forehead setup; strong temporal modelling	Simulator setting; forehead-EOG only (no full EOG), potential domain shift.
Choi et al. (2019) [68]	EEG	2-s windows; spectral features; XGBoost; compares **wired vs wireless** EEG for instantaneous drowsiness	Near-instant estimates; shows feasibility with wireless EEG	Controlled tasks; limited evidence for on-road generalization.
de Naurois et al. (2019) [61]	Behavioural/vehicle (+context)	ANN to **detect & predict** minute-by-minute drowsiness	Predictive framing; uses signals	Modest N; cross-driver generalization concerns.

			feasible in real cars	
Weng, Lai & Lai (2016/2017) [43]	Vision	**HTDBN** + HMM over eyes/mouth/head dynamics (video)	Captures temporal facial dynamics; multi-cue	Sensitive to lighting/IR and face visibility; dataset bias.
Abtahi et al. (2014) [57]	Vision (dataset)	Real illumination **yawning** clips; baseline detector	Widely used benchmark for mouth/yawn cues	Focused on yawns, limited subject diversity.
Silveira et al. (2019) [76]	Methodology / evaluation	Shows **subject-**often outperforms cross-subject; highlights **class imbalance** in realistic conditions	Critical guidance for protocol design; realism	Not a single deployed model; results tied to specific datasets.
Wang et al. (2020) [33]	EEG	Attention-based **multiscale CNN** + **dynamical GCN** for fatigue	Strong temporal–spatial feature learning; SOTA-level reports	Complexity & compute; external validity depends on dataset.
Fujiwara et al. (2023) [50]	ECG/HRV	R-R interval features + **self-attention autoencoder**; validated vs EEG	Non-intrusive physiology; promising sensitivity	HRV confounded by stress/caffeine/fitness; needs multi-modal confirm.
Zhang et al. (2019) [64]	EEG	**Second-order blind identification**	Exploits source separation;	Requires decent SNR, simulator bias

		(SOBI) on multi-channel EEG	interpretable components	typical of EEG studies.
Zheng et al. (2021) [21]	Dataset (EEG+EOG+performance)	23-subject simulated driving with **EEG/EOG**; PSD/DE features; labels incl. PERCLOS-related vigilance	Large, multi-modal, session-rich; aligns with your study	Simulator domain: forehead EEG/EOG config differs from many labs.

Chapter 3: Dataset

3.1 Introduction

The empirical foundation of this dissertation is the SEED-VIG dataset, developed by the BCMI Laboratory at Shanghai Jiao Tong University. SEED-VIG (Vigilance sub-dataset of the SJTU Emotion EEG Dataset) was designed to capture objective physiological and behavioural indicators of driver fatigue in a controlled simulated driving environment [72].

Driver drowsiness is a significant contributor to road accidents worldwide, and it manifests in both neural signals and behavioural measures. SEED-VIG is particularly suitable for this study because it provides multimodal signals- EEG, EOG, and eye-tracking-based PERCLOS, that enable machine learning models to assess fatigue from complementary perspectives [72][87].

3.2 Experimental Setup

The dataset was collected in a driving simulator designed to mimic a real vehicle cockpit, complete with steering wheel, pedals, and a gearshift. Participants drove for approximately two hours on a monotonous two-lane highway projected on a large screen. Sessions were conducted in the early afternoon to maximise the onset of fatigue through circadian and post-lunch effects [72].

During each of the **23 sessions**:

- **EEG signals** were recorded from 17 electrodes placed mainly in temporal and occipital regions according to the international 1020 system [72].
- **EOG signals** were captured via electrodes around the eyes and an SMI eye-tracker, which detected eyelid closures, blinks, and saccades [87].
- All signals were sampled at 200 Hz.

- Fatigue ground truth was derived from **PERCLOS** (Percentage of Eyelid Closure) [31].

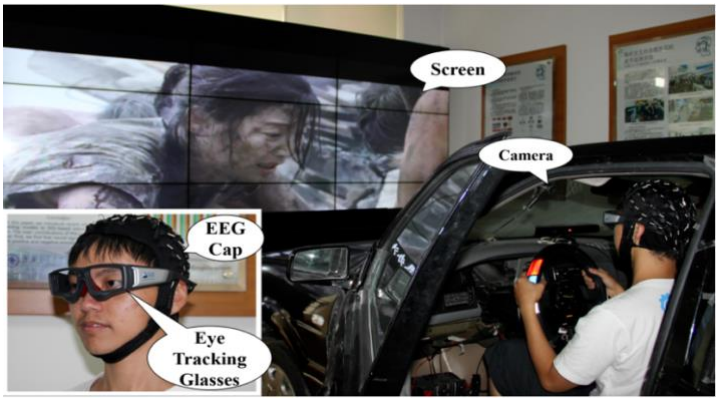

Figure 8: shows the EEG electrode layout and depicts the driving simulator setup [72].

3.3 EEG Modality

Electroencephalography (EEG) measures scalp electrical activity produced by synchronized neuronal firing. EEG is widely used in fatigue research because distinct **frequency bands** show measurable changes in drowsy states [72]:

- **Delta (1-4 Hz):** increases during severe drowsiness and sleep.
- **Theta (4-8 Hz):** rises during early fatigue, associated with reduced vigilance.
- **Alpha (8-14 Hz):** stronger when eyes are closed; posterior alpha reflects relaxation or disengagement.
- **Beta (14-31 Hz):** linked to alertness; decreases with reduced cognitive activity.
- **Gamma (31-50 Hz):** associated with high-level processing; reduction may reflect lower responsiveness.

The 17 electrodes used (FT7/8, T7/8, TP7/8, CP1/2, P1/2, Pz, PO3/4, POz, O1/2, Oz) target regions known to show strong fatigue-related oscillatory changes [72].

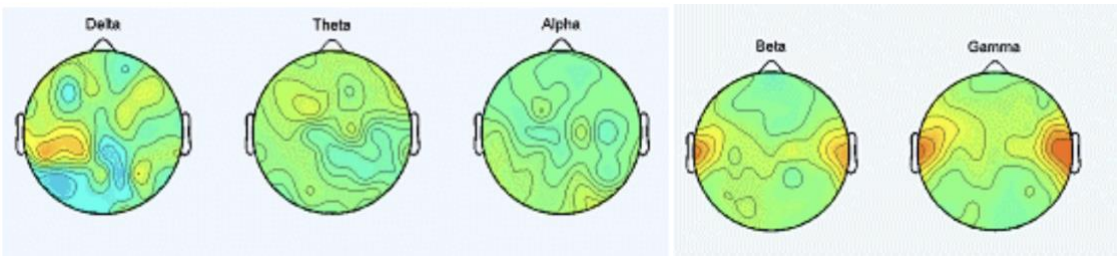

Figure 9: EEG electrode layout showing the 17 channels used in SEED-VIG [72]

3.4 EOG Modality

Electrooculography (EOG) measures voltage differences between the cornea and retina, enabling detection of eye movements and eyelid closures. From EOG signals, two key behaviours are analysed:

- **Blinks:** eyelid closures followed by reopening. Fatigue increases blink duration and sometimes produces microsleeps (very long closures) [87].
- **Saccades:** rapid eye shifts between fixations. Fatigue reduces saccade frequency and velocity, while increasing variability [87].

SEED-VIG provides blink and saccade metrics (rate, duration, amplitude, velocity), which complement EEG features by directly capturing behavioural indicators of vigilance.

3.5 PERCLOS Labelling

PERCLOS (Percentage of Eyelid Closure) is considered the most reliable single measure of driver drowsiness [31]. It represents the proportion of time, within a window, when eyelids are ≥80% closed.

In SEED-VIG, PERCLOS values were computed for each 8-second segment and mapped into three fatigue classes [88]:

- **0 Alert:** PERCLOS < 0.15
- **1 Moderate fatigue:** 0.15 ≤ PERCLOS < 0.30

- **2 Severe fatigue:** PERCLOS ≥ 0.30

These thresholds align with established research that uses 15% and 30% PERCLOS as cut-offs for vigilance classification [31][89].

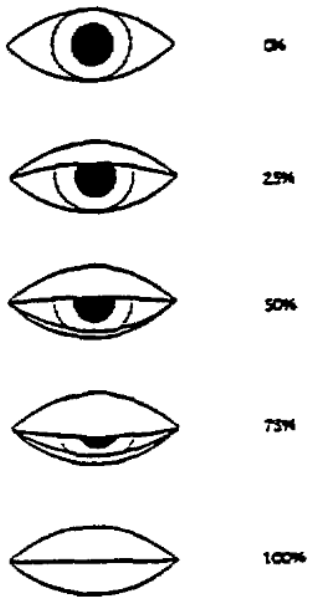

Figure 10: Illustration of PERCLOS calculation and [3].

3.6 Dataset Structure

The dataset is segmented into fixed-length windows of 8 seconds. Each window contains:

- **EEG features:** power in five frequency bands for 17 channels.
- **EOG features:** blink and saccade statistics.
- **Label:** fatigue level (0, 1, or 2).
- **Number of sessions:** 23
- **Samples per session:** 885
- **Total samples:** 20,000

- **Features:** 102
- **Labels:** 1 categorical fatigue level

3.7 Feature Descriptions

Features and Labels

Each record (data sample) in the pre-processed dataset represents a fixed-length time window (approximately 8 seconds) of the driving experiment, characterized by a set of extracted features and an associated fatigue label. The features capture information from EEG signals (brain activity) as well as ocular measures (from EOG/eye-tracking) that are known to correlate with fatigue. In total, the dataset contains 106 columns: an identifier for the session, 102 feature columns, and 1 label column. Table 4 below summarizes each feature and its description:

Table 4: Dataset Feature Descriptions (SEED-VIG)

Feature Name	Description
session	Session ID indicating the participant and session (e.g. subject and time of day). Used to group data by experiment session.
EEG_C1_delta … EEG_C17_delta	EEG delta-band power features for channels 1-17. Each represents the signal power in the delta-frequency band (1-4 Hz) for a specific EEG electrode. Delta waves are associated with deep sleep or very low alertness.
EEG_C1_theta … EEG_C17_theta	EEG theta-band power features for channels 1-17. Power in the theta band (≈4-8 Hz) for each EEG channel [100]. Elevated theta is often linked to drowsiness and early stages of sleep.
EEG_C1_alpha … EEG_C17_alpha	EEG alpha-band power features for channels 1-17. Power in the alpha band (≈8-14 Hz) for each

	channel. Alpha activity (especially posterior alpha) can increase when eyes are closed or when a person is relaxed.
EEG_C1_beta … EEG_C17_beta	EEG beta-band power features for channels 1-17. Power in the beta band (14-31 Hz) for each channel. Beta waves reflect alert, active brain states; decreased beta power may accompany fatigue.
EEG_C1_gamma … EEG_C17_gamma	EEG gamma-band power features for channels 1-17. Power in the gamma band (31-50 Hz) for each channel. Gamma is associated with high cognitive activity; changes in gamma may also relate to vigilance level.
EOG_mean	Mean value of the EOG (electrooculogram) signal over the time window. The EOG captures eye movement and blink-related electrical activity; the mean level may reflect baseline eye position or slow drifts.
EOG_std	Standard deviation of the EOG signal in the window. This measures the variability of eye movement signal, indicating how much the eye position fluctuated (e.g. due to blinks or saccades). A higher EOG std suggests more eye movement activity in that period.
BlinkRate_Sum	Total number of blinks detected in the time window (blink count). Blinks are brief eye closures; frequent blinking can be a sign of fatigue or dry eyes.
BlinkRate_Mean	Average blink rate (e.g. blinks per minute) during the window. This is essentially the blink frequency normalized over time. Blink rate tends to increase with mild fatigue but may decrease in very severe drowsiness (when eyes stay closed longer).

BlinkDur_Mean	Mean blink duration the average length of time (seconds) that each blink lasted, within the window. Longer blink durations (slow eyelid closures) are a known indicator of drowsiness [88].
BlinkDur_Max	Maximum blink duration observed in the window. A very long blink (sometimes called an eyelid “droop” or microsleep) might occur during extreme fatigue.
BlinkDur_Min	Minimum blink duration in the window (shortest blink).
BlinkDur_Var	Variance of blink durations in the window, indicating the consistency or variability of blink lengths. Higher variance means some blinks were much longer than others, which can happen as a person transition into fatigue.
BlinkAmp_Mean	Mean blink amplitude, i.e. the average peak amplitude of the EOG signal during blinks. Blink amplitude reflects how fully the eyes close; it may change with fatigue (e.g., slower, partial blinks might have smaller amplitude).
BlinkAmp_Max	Maximum blink amplitude recorded in the window (the strongest eye closure).
BlinkAmp_Min	Minimum blink amplitude in the window (the weakest blink).
SaccadeRate_Sum	Total number of saccades (rapid eye movements) detected in the window. Saccades are quick eye jumps; their frequency might decrease when a driver is zoning out or fatigued (since drowsy individuals tend to have more fixations and fewer active scans of the environment).
SaccadeRate_Mean	Average saccade rate (count per minute) during the window.

SaccadeDur_Mean	Mean saccade duration the average time each saccadic eye movement lasted. (Saccades are normally very brief; longer durations could indicate slower eye movements.)
SaccadeDur_Max	Maximum saccade duration in the window (the longest eye movement observed).
SaccadeDur_Min	Minimum saccade duration in the window.
SaccadeDur_Var	Variance of saccade durations, measuring variability in saccade speeds.
SaccadeVel_Mean	Mean saccade velocity the average peak speed of eye movements (how fast the eye moved) during saccades in the window. Fatigue can lead to slower eye movements on average.
SaccadeVel_Max	Maximum saccade peak velocity in the window (the fastest eye movement recorded).
fatigue_level	Fatigue label class, derived from PERCLOS. Coded as 0 (low fatigue/alert), 1 (moderate fatigue), or 2 (severe fatigue). This was obtained by thresholding the continuous PERCLOS value for the window into three ranges (alert vs. mildly drowsy vs. severely drowsy) [88].

Notes: All features have been normalized (scaled) and processed to correct skewness as part of data preprocessing (details covered in a later chapter). The table above omits detailed preprocessing steps, focusing instead on the meaning of each feature. In summary, the feature set encompasses both brain-wave features (power in standard EEG bands from 17 channels) and eye-movement features (blink and saccade metrics from EOG) that collectively capture the subject's vigilance state. This combination follows established multimodal approaches for fatigue detection, leveraging that drowsiness impacts both neural activity and observable eye behaviours [90]. The fatigue_level label provides a categorical target based on PERCLOS, which is a widely used objective indicator of driver drowsiness [88].

3.8 Ethical and Accessibility Notes

The SEED-VIG dataset is publicly available via the BCMI Lab website[72]. All data were collected under approved institutional ethics protocols, and participants provided informed consent.

3.9 Summary

This chapter described the SEED-VIG dataset, including experimental setup, EEG and EOG modalities, PERCLOS-based labelling, dataset structure, and feature definitions. The dataset provides a rich multimodal basis for machine learning analysis of fatigue.

The next chapter will discuss methodology, covering preprocessing, feature engineering, and model training.

Chapter 4: Methodology

4.1 Research Design

This research adopts a supervised machine learning design to detect driver fatigue using multimodal physiological signals, electroencephalography (EEG) and electrooculography (EOG), recorded during simulated driving. The methodological framework follows a systematic pipeline: raw EEG/EOG signals are first aligned, cleaned, and segmented into fixed windows; features capturing spectral, temporal, and ocular characteristics are engineered; dimensionality reduction and redundancy analysis are performed to obtain compact, informative representations; multiple machine learning and deep learning classifiers are trained; and rigorous validation strategies are applied to evaluate generalization performance. A schematic of this workflow is presented in Figure 11.

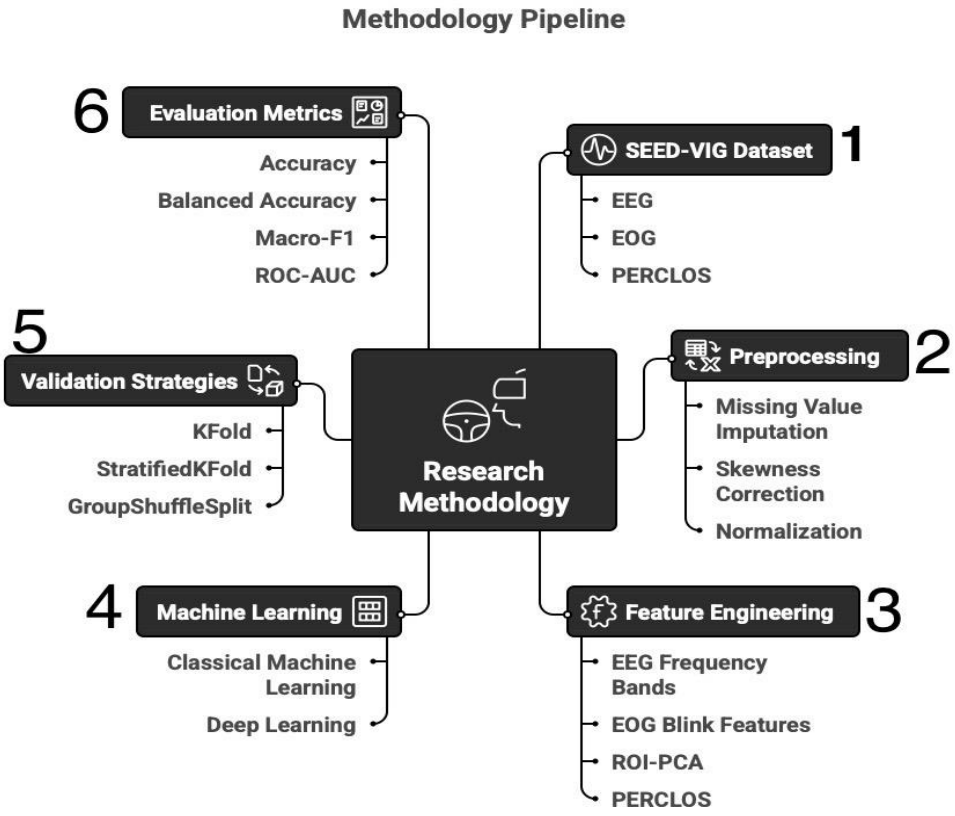

Figure 11: Methodology workflow

The design rests on the hypothesis that neurophysiological changes measurable through EEG and EOG provide sensitive biomarkers of fatigue. EEG signals capture characteristic alterations in brain rhythms such as increases in theta (4-8 Hz) and alpha (8-14 Hz) power, alongside decreases in beta (14>30 Hz) activity during drowsiness [91][66]. Simultaneously, EOG signals track ocular behaviour, where blink duration increases, blink rate decreases, and slow eyelid closures emerge as drivers fatigue [31]. Together, these modalities provide complementary central (cortical) and peripheral (ocular) signatures of fatigue, making them well-suited for early and reliable detection [91].

Ground-truth vigilance is derived from PERCLOS (percentage of eyelid closure), a validated and widely used index of drowsiness [31]. Following prior studies [91], discretize the continuous PERCLOS values into three vigilance classes rather than a binary threshold, to capture both transitional and severe fatigue states. The classification rule is defined as:

$$PERCLOS = \begin{cases} 0 - \ p_i < 0.15 \ (Alert) \\ 1 - \ 0.15 \leq \ p_i < 0.30 \ (Mild \ Fatique) \\ 2 - \ p_i \geq 0.30 \ (Severe \ Fatique) \end{cases}$$

where p_i denotes the PERCLOS value of epoch i, bounded in [0,1].

Neurophysiological studies show that as PERCLOS increases, theta (4-8 Hz) and alpha (8-14 Hz) activity rises, sometimes accompanied by δ/α shifts [31], while ocular markers such as longer blinks and reduced saccades emerge [91], [92]. Thus, combining EEG(central nervous system activity) with EOG(ocular behaviour) provides complementary and early indicators of vigilance loss [91], [92].

Dual modelling approach was adapted:

1. **Classical machine learning models** (Random Forest, XGBoost, RBF-SVM) trained on hand-engineered EEG/EOG features.

2. A **deep learning model** (MLP & CNN-BiLSTM with Focal Loss) trained directly on minimally processed multichannel time-series segments, capable of learning temporal and spectral features end-to-end.

This design enables a fair comparison between feature-engineering-driven methods and representation-learning-driven methods. Performance is evaluated with metrics sensitive to class imbalance (e.g., F1-score, ROC-AUC, balanced accuracy) under a cross-subject validation scheme, where each subject/session is held out at least once. The methodology is carefully designed to avoid data leakage, ensure reproducibility, and emphasize generalizability to unseen drivers.

4.2 Preprocessing

Dataset Overview

4.2.1 Data Extraction Pipeline

The SEED-VIG dataset [72] is described in detail in Chapter 3. Here, data extraction is outlining the and preparation steps performed to transform the raw multimodal signals into an analysable format suitable for feature engineering and model training. While the dataset provides synchronized EEG, EOG, and eye-tracking recordings, substantial preprocessing was required to ensure consistent alignment, segmentation, and labelling across sessions.

4.2.2 File access and loading- The dataset was distributed in MATLAB's .mat format, with each file corresponding to one driving session. Python's $scipy.io.loadmat$ package is used to extract raw EEG (17 channels), EOG (2-4 channels), and PERCLOS vigilance labels. Each file also contained metadata such as session identifiers and electrode indices.

4.2.3 Epoch segmentation- Continuous recordings of approximately two hours per session were segmented into non-overlapping 8-second windows, following the

convention established in prior SEED-VIG publications [91], [21]. Each 8-second epoch corresponds to 1600 samples per channel at the standard EEG sampling rate of 200 Hz. Segmenting into fixed-length windows ensured that all modalities could be aligned with the PERCLOS labels provided every 8 seconds.

```
20151125_noon: windows=885 | fs=200.0 Hz | aligned_len=118.0 min | samples_per_epoch=1600
```

Figure 12: Epoch segmentation check: each ~2-h recording is split into non-overlapping 8-s windows at 200 Hz (1600 samples/epoch), following SEED-VIG convention

4.2.4 Signal synchronization- While EEG signals were consistently recorded at 200 Hz, some EOG channels exhibited effective sampling rates closer to 125 Hz due to device differences. To maintain strict temporal alignment between EEG and EOG, all EOG signals were resampled to 200 Hz using linear interpolation. This procedure ensured that each 8-second epoch contained 1600 samples for every EEG and EOG channel, thereby enabling synchronized multimodal feature extraction.

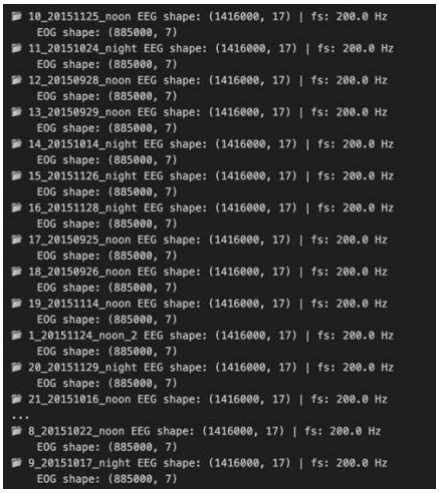

Figure 13: multimodal alignment: EOG (≈125 Hz) is low-pass filtered and up-sampled to 200 Hz so each 8-s epoch contains 1600 samples in both EEG and EOG.

It was verified that resampling did not introduce aliasing by applying a low-pass filter prior to up-sampling and by cross-checking with $scipy.signal.resample\_poly$, confirming that EOG spectra below 30 Hz remained intact

4.2.5 Schema standardisation- The raw dataset indexed EEG channels as C1-C17 without consistent naming across sessions. Uniform mapping was established to the international 10-20 system (e.g., FT7/8, TP7/8, PO3/4, O1/2), ensuring that the same feature always referred to the same anatomical location across subjects. Similarly, vertical and horizontal EOG channels were standardised to fixed indices. This schema alignment is critical to prevent spurious inter-subject variability.

	Raw index/name	Standardised label
0	C1	FT7
1	C2	FT8
2	C3	TP7
3	C4	TP8
4	C5	PO3
5	C6	PO4
6	C7	O1
7	C8	O2
8	C9	F3
9	C10	F4
10	C11	C3
11	C12	C4
12	C13	P3
13	C14	P4
14	C15	Fz
15	C16	Cz
16	C17	Pz
17	EOG_v	idx 0
18	EOG_h	idx 1

Figure 14: Schema standardisation: raw EEG channels (C1–C17) mapped to 10–20 labels and EOG vertical/horizontal fixed to canonical indices to ensure consistent anatomical meaning across sessions.

4.2.6 Session structuring- For each subject session, the extracted epochs were stored in a structured table with the following fields:

- Session ID (to group data by participant and session).
- EEG data (17 channels × 1600 samples).
- EOG data (2 channels × 1600 samples).
- PERCLOS vigilance value.
- Epoch index.

	session_id	epoch_idx	eeg_shape	eog_shape	perclos
0	20151125_noon	0	(1600, 17)	(1600, 2)	0.252311
1	20151125_noon	1	(1600, 17)	(1600, 2)	0.888860
2	20151125_noon	2	(1600, 17)	(1600, 2)	0.551296

Figure 15: Per-session epoch table: for each session epoch index, EEG/EOG tensors (1600×17 and 1600×2), and the aligned PERCLOS value is stored.

This structuring allowed grouped cross-validation by session, ensuring no data leakage between training and test folds.

***4.2.7 Label assignment*:** Each epoch was aligned with the corresponding PERCLOS score (computed by the eye tracker over the same 8-second interval). Rather than a binary threshold, a three-class discretisation of vigilance to capture transitional states of fatigue was adopted.

p_i<0.15(Alert) 0.15 ≤p_i<0.30(Mild fatigue) p_i≥0.30(Severe fatigue)

This scheme aligns with established thresholds in fatigue research [31], [88], which demonstrate that 15% and 30% eyelid closure represents reliable cut-offs for vigilance classification.

Combining EEG (central nervous system signals) with EOG (peripheral ocular dynamics) therefore provides complementary and early markers of vigilance loss [13], [91]. Embedding these multimodal indicators into the dataset extraction process ensured that subsequent models could learn from both neural and behavioural evidence of fatigue.

Final dataset After preprocessing and label alignment, the resulting dataset comprised 20,000 epochs across 23 sessions, each represented by synchronized EEG and EOG samples plus a categorical vigilance label (0-2). This extracted dataset formed the empirical basis for the subsequent feature engineering and classification stages.

4.3 Signal Preprocessing

This stage converts the extracted, session-level windows into a leakage-safe, variance-stabilised feature set suitable for supervised learning. The pipeline covers identifier standardisation, label attachment, leakage guarding, normalization, and spatial aggregation (ROI/asymmetry/PCA).

4.3.1 Normalisation and skewness correction.

Raw EEG/EOG features exhibited strong right-skewed distributions and heterogeneous scales, particularly for high-frequency power features and blink amplitude metrics. Such non-Gaussianity can destabilise models sensitive to scale (e.g., SVM, logistic regression) and exaggerate the influence of extreme values [91][93]. To address this, a two-step transformation pipeline was implemented, retained in the final inference workflow to ensure consistent preprocessing:

1. **Median imputation.** Although no missing values were present in the current dataset, a median imputation step was included for robustness. The median is preferred over the mean in physiological data due to its resilience to outliers.[94]
2. **Power transformation (Yeo-Johnson) and standardisation.** For each feature, Yeo-Johnson power transform is applied, which can handle both positive and negative values, unlike Box-Cox [93]. The transformed feature was then standardised:

$$x' = \frac{T_{Yj}(x) - \mu}{\sigma}$$

where $T_{Yj}(\cdot)$ is the Yeo-Johnson transform, and μ,σ are training-fold mean and standard deviation.

1. The Yeo-Johnson transform was chosen because it accommodates both positive and negative values [66].
2. Features were then standardized to zero mean and unit variance, improving comparability across channels and subjects.

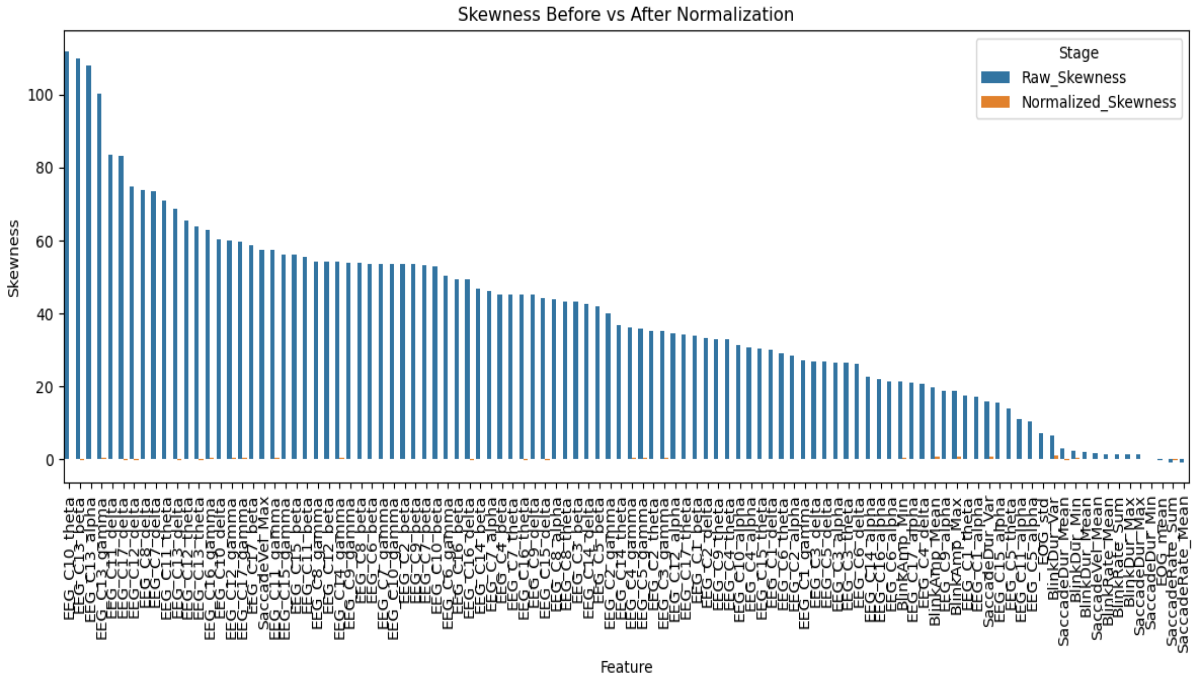

Figure 16: Skewness of EEG/EOG features before and after normalization.

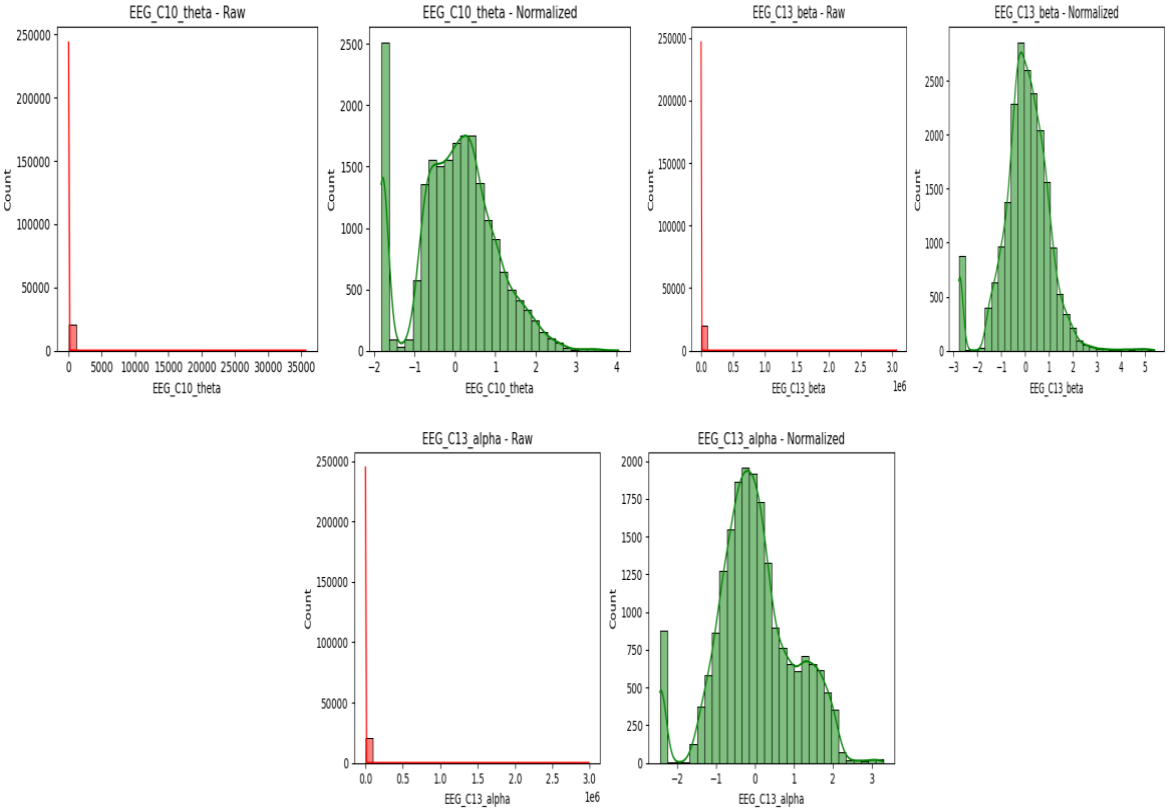

Figure 17: Example histograms showing raw vs normalized distributions for representative EEG features (θ, β, α bands).

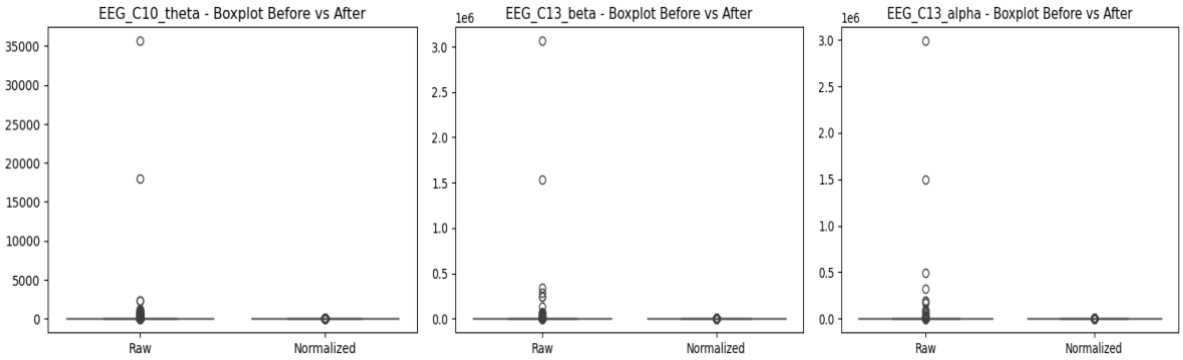

Figure 18: Boxplots before and after normalization, illustrating suppression of extreme outliers.

Effectiveness checks:

- **Skewness plots**: Bar charts showed raw skewness values in the tens/hundreds for EEG power bands; after transformation, values clustered near zero.
- **Histograms/KDEs**: Long-tailed distributions became symmetric and Gaussian-like.
- **Boxplots**: Extreme EEG outliers were compressed, reducing undue influence.

These checks confirmed normalization significantly improved data quality.

4.4 Domain-Specific Feature Engineering

Building on the baseline feature set described in Chapter 3, additional variables were engineered to better capture the spectral imbalances, temporal dynamics, and ocular irregularities that characterise transitions into fatigue. These derived features substantially expanded the representation of each 8-s epoch, increasing the dimensionality from 102 baseline variables to approximately 300-330 engineered features, prior to spatial aggregation and redundancy reduction.

4.4.1 EEG Band Ratios

While SEED-VIG provides absolute band powers for five canonical frequency bands (δ, θ, α, β, γ) per electrode, relative ratios are often more sensitive to drowsiness because they capture balance between oscillatory processes rather than absolute amplitudes. For each channel, derived:

- **Theta/Alpha ratio** (θ/α), rises as vigilance decreases, reflecting increased slow activity relative to relaxed posterior alpha. [95]
- **Theta/Beta ratio** (θ/β), highlights the suppression of alertness-linked β activity alongside increasing θ activity.[95][36]
- **(Theta+Alpha)/Beta ratio** ((θ+α)/β), used in prior fatigue indices to capture global slowing relative to high-frequency engagement [36].
- **Delta/Alpha ratio** (δ/α), indicative of deeper fatigue and early sleep onset stages [96].

To avoid instability, denominators were smoothed with a small ε. These ratios produced 34 new features, enriching the spectral profile with physiologically interpretable markers.

4.4.2 Temporal Dynamics

Fatigue develops gradually; static features alone cannot reflect this trajectory. To incorporate short-term context without violating causality, temporal features were engineered within each session:

- **Rolling mean and standard deviation** (look-back k=2k=2 epochs, 24-40 s) to encode local drift and variability.
- **First-order differences** (Δx_t=$x_t - x_{t-1}$) to capture sudden changes, such as abrupt θ spikes or β drops preceding microsleeps.
- **Exponential moving averages** with short decay to emphasise recent history while smoothing noise.

These temporal variables added 170 features, allowing non-sequential models (e.g. RF, SVM) to exploit context that would otherwise require recurrent architectures [66].

4.4.3 Additional EOG Metrics

Beyond the blink and saccade statistics provided in SEED-VIG, extracted secondary ocular descriptors were extracted:

- **Variance of the EOG signal**, indicating the overall activity of eye movements.
- **Entropy measures**, quantifying irregularity in ocular patterns. Lower entropy and variance are expected during prolonged closures or drifting gaze.
- **Local closure proportion** (percentage of epoch above a high EOG threshold), a proxy for PERCLOS. Although not used for labelling (to avoid leakage), it serves as an auxiliary behavioural feature in multimodal models.

These additional 10-15 metrics complemented the existing 36 EOG features and reinforced the multimodal perspective that combines cortical and ocular indicators [66][92].

4.4.4 Feature Space Prior to ROI Aggregation

After adding EEG ratios, temporal statistics, and enriched ocular metrics, the feature space comprised:

- 85 EEG band powers (log-scaled)
- 34 EEG ratios
- 170 temporal context features
- 36 + 10 EOG features (baseline + engineered)

This yielded ≈300-330 features per epoch. Correlation analysis (Figure 4.5) revealed strong redundancy blocks (Pearson's r > 0.9r > 0.9) among related variables, confirming the need for dimensionality reduction (Section 4.6). The strategy here was intentional over-generation: by first constructing a comprehensive, physiologically

motivated set, it was ensured that downstream selection methods could retain the most discriminative features while discarding redundancies.

4.5 Correlation analysis for redundancy.

To assess redundancy, Pearson correlation coefficients were computed [97]:

$$r_{jk} = \frac{\sum i\,(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum i\,(x_{ij} - \bar{x}_j)^2}\,\sqrt{\sum i\,(x_{ik} - \bar{x}_k)^2}}$$

x_{ij} → the value of feature **j** for sample **i** (e.g., theta power at channel Fp1 for epoch i).

x_{ik} → the value of feature **kk** for the same sample i (e.g., alpha power at channel O2 for epoch i).

$\bar{x}_j$ → the mean value of feature j across all samples.

$\bar{x}_k$ → the mean value of feature kk across all samples.

- **Feature-label correlation**: confirmed strong relationships for plausible biomarkers (theta/alpha ratios, blink duration), but also revealed suspiciously high correlations for direct eye-closure metrics, which were excluded per leakage rules.
- **Feature-feature correlation**: expected strong blocks due to spatial proximity (neighbouring electrodes), band adjacency, and rolling statistics. Heatmaps of the top 20 correlated features highlighted redundancy clusters, motivating PCA-based reduction.[98]

This correlation analysis provided the motivation for applying ROI-wise principal component analysis (PCA) in later stages to decorrelate features while preserving underlying physiological structure. This methodological choice was also reinforced and verified through consultation with a medical professional from the School of Mental Health and Clinical Neurosciences at the University of Nottingham, who advised that

ROI-based aggregation and dimensionality reduction would help align the statistical pipeline with known neurophysiological organisation while reducing noise.

4.6 Spatial Aggregation, Asymmetry, and ROI-wise PCA

4.6.1 ROI features (10-20 grouping).

Electrodes were grouped into standard Regions of Interest (ROIs) (frontal, central, parietal, occipital, temporal; see Chapter 4 for montage). For each ROI × band, a trimmed mean (10%) was computed as a robust centre and the within-ROI standard deviation as dispersion. These regional summaries improve SNR and encode spatial patterns linked to vigilance changes (e.g., frontal θ increases) without duplicating channel-level detail [16][66].

4.6.2 Hemispheric asymmetry.

To capture lateralised effects observed in fatigue and sleep research, left-right asymmetry indices for homologous sites/ROIs was formed [99]:

$$Asymmetry = P_{left} - P_{right},$$

(and for selected band ratios). Asymmetry suppresses common-mode changes and exposes hemispheric shifts that single-channel features may miss.

4.6.3 Feature impact.

This step added 85 features (50 ROI means/SDs + 35 asymmetries). In the full-electrode configuration, the working matrix rose to 448 variables before redundancy reduction.

ROI-wise PCA (redundancy reduction)

Because regional band powers and ratios are still correlated, PCA was applied within each ROI, fitting on training folds only. For each ROI the smallest number of PCs were retained explaining ≥95% variance (capped at 3 PCs/ROI), yielding 15-18 compact ROI-PCs that preserve spatial patterns while decorrelating features [Ren et al., 2021[16]]. Both Baseline (ROI + asymmetry) and Baseline + ROI-PCA variants were kept for ablation.

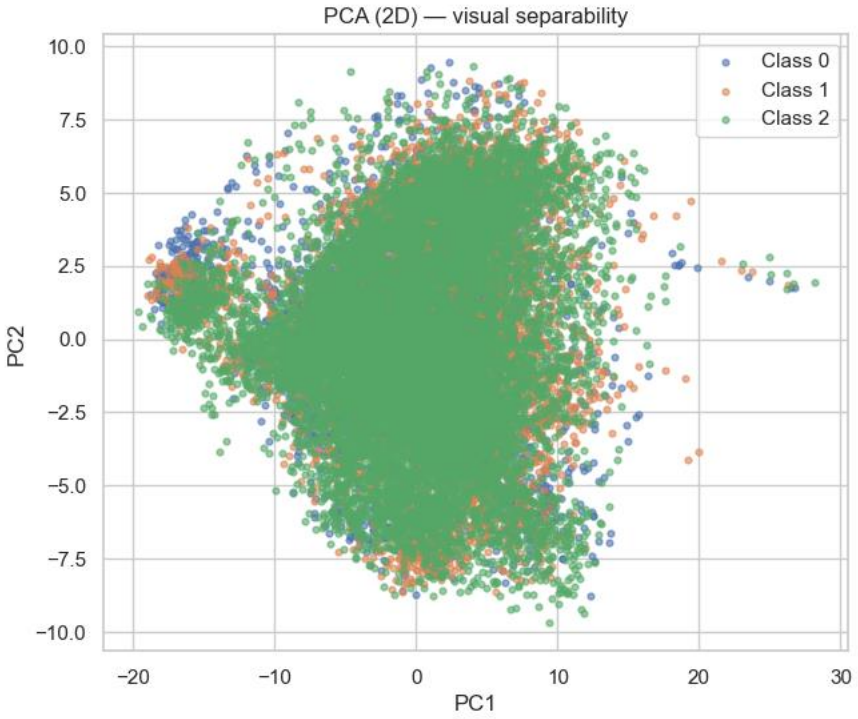

Figure 19: 2D PCA scatter plot showing sample distribution across the first two components. Classes overlap, indicating limited separability after dimensionality reduction.

4.6.4 Correlation before vs after.

Pearson correlation heatmaps showed (fig 20) dense redundancy blocks in the full engineered feature space prior to PCA. After applying ROI-wise PCA, these blocks were substantially reduced: average pairwise correlation dropped from 0.288 to 0.237, and the median from 0.321 to 0.234. This demonstrates that ROI-PCA effectively controlled redundancy while preserving spatially meaningful patterns of neural activity.

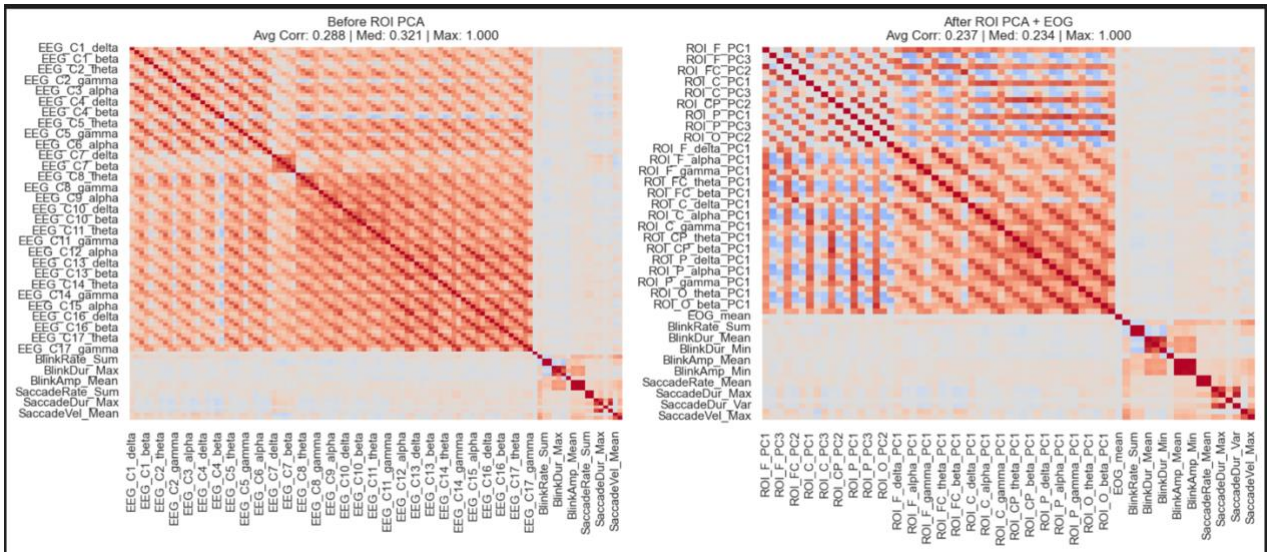

Figure 20: Feature correlation heatmaps before (left) and after ROI-PCA (right), showing reduced redundancy post-PCA.

Scree plots confirmed that the first three principal components per ROI were sufficient to capture ≥95% of variance, supporting the decision to cap retention at three PCs per region. Importantly, this compression removed redundant variance without discarding physiologically interpretable structure.

4.7 Leakage Prevention

Ensuring that evaluation metrics reflect true generalisation rather than artefacts of preprocessing or splitting was a core methodological priority. Because EEG/EOG data are highly autocorrelated and session-specific, leakage can easily occur if temporally adjacent or proxy-related information from the test set is allowed to influence training. Several safeguards were adopted:

4.7.1 Session-level validation

Cross-validation was performed at the level of entire sessions: all epochs from a given recording were assigned exclusively to training or testing folds. This prevented the common error where consecutive windows from the same drive appear in both sets, which would otherwise inflate accuracy due to temporal similarity.[66][100]

4.7.2 Fold-specific preprocessing

All preprocessing steps, skewness correction, standardisation, feature selection, and PCA, were fitted only on the training folds and then applied to the held-out fold. This was enforced through scikit-learn pipelines so that the test data never contributed to the estimation of means, variances, or principal components.[100]

4.7.3 Oversampling restricted to training data

When addressing class imbalance with synthetic oversampling (e.g., SMOTE), the procedure was applied strictly within the training folds. Test data remained unaltered to ensure that synthetic examples did not contaminate validation results.

4.7.4Temporal causality preserved.

When computing rolling or difference-based features, only past epochs within the same session were used. No feature incorporated information from future windows, thereby preserving temporal causality and making the engineered variables realistic for deployment.

4.7.5 Removal of label-proxies.

Although EOG-derived closure measures could directly reflect the PERCLOS label, such features were removed earlier in preprocessing (Section 4.2.7). This ensured that models learned from indirect neural and ocular correlates, not the label-generation mechanism itself.

4.8 Model Training and Evaluation

To evaluate the effectiveness of EEG/EOG features for fatigue detection, both classical machine learning algorithms and deep learning architectures were considered. This dual approach allowed us to compare traditional feature-driven pipelines against end-to-end representation learning, and to assess which inductive biases are most appropriate for this domain.

4.8.1 Classical Machine Learning Models

First implemented three widely used classical models: Random Forests (RF), extreme Gradient Boosting (XGBoost), and Support Vector Machines (SVMs) with radial basis function kernels. These algorithms represent distinct learning paradigms, bagging, boosting, and margin-based classification, providing complementary perspectives on the same feature space.

4.8.2 Random Forests (RF)

RFs construct an ensemble of decision trees, each trained on bootstrapped subsets of data and features, and combine them via majority voting [91]. This reduces variance and improves robustness to noise and redundancy. RFs were selected because EEG/EOG features are inherently correlated (e.g., spatially adjacent electrodes or related band ratios), and RFs are resilient to such multicollinearity [101] Additionally, RFs provide interpretable feature importance measures, enabling us to identify which spectral or ocular features contributed most to predictions.

4.8.3 XGBoost

XGBoost is a gradient boosting algorithm that builds trees sequentially, where each new tree corrects the errors of its predecessors [102]. It includes regularisation to reduce overfitting and is known for strong performance on tabular biomedical data. XGBoost was chosen because it focuses on hard-to-classify samples, which is particularly relevant in our case: transitional fatigue states (mild fatigue) are more subtle than clear alert or severe states, and boosting can adaptively emphasise these difficult cases.

4.8.4 Support Vector Machines (SVMs).

SVMs with RBF kernels map data into a higher-dimensional space and learn a hyperplane that maximises the class margin [103]. They are effective in high-dimensional feature spaces and have historically performed well on EEG classification tasks [92]. SVMs were included to provide a contrasting baseline to ensemble trees: while tree-based models are tolerant of redundant features, SVMs require careful

scaling and dimensionality reduction (via ROI-PCA). Their inclusion ensures our comparisons cover both kernel-based and ensemble-based learning paradigms.

Together, RF, XGBoost, and SVMs provide a broad benchmark across classical learning strategies, robust variance reduction, iterative error correction, and margin maximisation.

4.8.5 Deep Learning Models

To complement feature-based models, two neural approaches were also implemented: a Multilayer Perceptron (MLP) and a hybrid Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM). These were designed to test whether deep architectures could capture temporal-spectral patterns without heavy reliance on hand-engineered features.

4.8.6 Multilayer Perceptron (MLP).

The MLP consists of fully connected layers with non-linear activations applied directly to engineered feature vectors. It was included as a neural baseline, allowing us to test whether a relatively simple network could capture feature interactions more effectively than tree ensembles, while still operating on pre-extracted features. [104]

4.8.7 CNN-BiLSTM.

This hybrid architecture applies convolutional filters to raw multichannel EEG/EOG segments to extract local spectral-spatial features (e.g. transient theta bursts or blink patterns). The resulting feature maps are passed to a Bi-LSTM, which captures bidirectional temporal dependencies over consecutive epochs. This design reflects the gradual evolution of fatigue and is supported by prior evidence that convolutional-recurrent hybrids outperform standalone models in EEG-based detection tasks [105][106].

4.8.8 Focal Loss for imbalance.

A key challenge in SEED-VIG is class imbalance, with severe fatigue epochs dominating the dataset. To address this, the CNN-Bi-LSTM was trained with Focal Loss [13], which down-weights easy majority-class examples and emphasises harder minority cases. This was motivated by the practical importance of detecting transitional (mild) fatigue, which can be missed by conventional cross-entropy training.

In summary, the deep models were chosen to explore the advantages of representation learning. The MLP provides a fair baseline, while the CNN-Bi-LSTM captures richer temporal-spectral dependencies and directly models the progression of vigilance loss.

4.9 Cross-validation and Evaluation Protocol

Robust evaluation is essential in physiological machine learning, particularly when models are intended for real-world deployment in driver fatigue detection. To this end, a comprehensive cross-validation strategy was employed that combined hold-out testing, row-wise k-fold validation, and session-aware evaluation. This multi-level protocol ensured that model performance was assessed under both within-session and cross-session conditions, thereby capturing a broad spectrum of generalisation scenarios [91].

4.9.1 Hold-out evaluation

80/20 hold-out splits were first performed as an intuitive baseline.

- For Random Forest, XGBoost, and SVM, the split was stratified by fatigue level, preserving class proportions.
- For the scikit-learn MLP, additionally tested a non-stratified 80/20 split, which provided a quick reference for early experiments.

These hold-out evaluations offered an accessible snapshot of performance across unseen data partitions while maintaining class balance.[66]

4.9.2 Row-wise k-fold validation

To obtain a more reliable estimate across the full dataset, 5-fold cross-validation employed:

- KFold (k=5, shuffle=True, random_state=42) for Random Forest and XGBoost.
- StratifiedKFold (k=5, shuffle=True, random_state=42) for SVM and SVM with SMOTE.
- KFold (k=5) for MLP.

This approach ensured that every sample was tested exactly once, and models were trained on diverse subsets of the data. Row-wise k-fold validation allowed us to capture stable, reproducible metrics across folds [107].

4.9.3 Session-aware evaluation

To simulate real-world deployment, where a model must generalise to unseen drivers or driving sessions, implemented session-aware splits for the deep learning models (CNN-BiLSTM).

- Using GroupShuffleSplit (test size = 20%), and partitioned entire sessions into training and test sets, ensuring that no overlap occurred.
- A further 10% stratified validation split was carved from the training data for hyperparameter tuning and early stopping.
- Additionally, for completeness, reported a 5-fold StratifiedKFold on sequences to illustrate performance stability across different partitions.

This design aligns with best practices in EEG/EOG research, where cross-subject validation is critical for assessing model robustness in realistic scenarios [108].

4.9.4 Handling class imbalance during cross-validation

Given that fatigue states were not equally distributed across sessions, multiple strategies were adopted to mitigate imbalance:

- Metrics such as Balanced Accuracy, Macro-F1, and ROC-AUC (OvR) were reported alongside overall accuracy, providing a fairer assessment of minority classes [109].

- In SVM+SMOTE experiments, synthetic oversampling was applied *only within the training folds*, preserving the integrity of validation and test sets.
- In the CNN-BiLSTM experiments, class weights and a class-balanced sampler were employed during training to enhance minority-class representation without altering evaluation data.

4.9.5 Preprocessing encapsulation and leakage prevention

All preprocessing stages (imputation, scaling, skewness correction, PCA, feature selection, oversampling) were encapsulated within scikit-learn pipelines or PyTorch training loops, ensuring that transformations were computed using training data only in each fold.

- For tabular models, fold-specific scaling and transformation parameters were applied exclusively to the corresponding test partition.
- For sequence models, per-session z-score normalisation was performed separately within each session, maintaining the temporal and subject-level independence of test data.

This rigorous handling of preprocessing within folds ensured reproducibility, integrity, and fairness of the reported metrics [100].

4.9.6 Reporting

For each protocol, results are reported as mean ± standard deviation across folds, supplemented with confusion matrices for representative 80/20 hold-out runs. By combining hold-out, k-fold, and session-aware protocols, it was ensured that the evaluation framework balanced methodological rigour with practical relevance, highlighting both the stability of the models and their applicability to unseen drivers.

Chapter 5: Experimental Design

5.1 Introduction / Purpose

In this dissertation, the methodological framework has already established the processes of data collection, preprocessing, and model development for multimodal fatigue detection using EEG and EOG signals. However, the effectiveness of any methodological design can only be validated through carefully structured experiments. The Experimental Design chapter therefore defines how the models, features, and evaluation protocols are systematically tested to ensure that results are both reliable and generalisable.

Experimental design plays a crucial role in bridging the gap between methodology and results. While the methodology specifies *how* the data pipeline is constructed, the experimental design specifies *what is being tested, why it is being tested, and under what conditions*. This distinction ensures that the findings reported in the next chapter are not ad-hoc but instead the outcome of a coherent set of controlled comparisons.

5.2 Research Hypotheses

The experiments are guided by testable hypotheses that directly correspond to the research objectives. These hypotheses are not prescriptive claims but structured questions that allow empirical validation.

- **H1 (Multimodality):** The integration of EEG and EOG features will yield better classification performance than unimodal approaches (EEG-only or EOG-only). This is expected because the modalities capture complementary aspects of fatigue: cortical activity and ocular behaviour. [66]
- **H2 (Dimensionality reduction ROI-PCA):** The application of region-of-interest PCA may alter performance compared to using raw channel features. While prior biomedical studies suggest that PCA reduces redundancy and noise, preliminary observations in this work indicated that ROI-PCA sometimes

resulted in a slight decrease in accuracy. Thus, an explicit ablation study is conducted to evaluate whether dimensionality reduction improves generalisation or, conversely, suppresses useful discriminative variance. [98]

- **H3 (Model family comparison):** Deep sequence models (e.g., CNN-Bi-LSTM) will outperform classical tabular models (Random Forest, XGBoost, SVM, MLP) on temporally structured input, due to their ability to capture long-range dependencies. However, ensemble methods are expected to remain competitive on static feature representations, consistent with findings by recent EEG studies [19].
- **H4 (Imbalance handling):** The use of imbalance-aware strategies (SMOTE oversampling, class weighting, focal loss, or balanced sampling) will improve minority-class recall and F1-scores compared to uncorrected baselines, though improvements may come at a cost of reduced overall accuracy [110].

5.3 Dataset Views and Experimental Inputs

The SEED-VIG dataset was pre-processed into multiple experimental views to enable systematic testing of feature engineering strategies. Each view represents a different level of feature transformation, designed to evaluate whether dimensionality reduction, multimodal integration, or sequential context improves fatigue detection performance

5.3.1 Tabular view: Before ROI-PCA

The first view, referred to as Before ROI-PCA, consists of the full set of extracted EEG and EOG features, along with fatigue labels derived from PERCLOS thresholds. In this representation, each row corresponds to an 8-second window of driving data. EEG signals are decomposed into five canonical frequency bands (delta, theta, alpha, beta, gamma), producing power or differential entropy features for each channel. EOG features capture blink and saccade dynamics.

- **Rationale:** This view preserves the highest dimensionality, retaining channel-wise information without spatial compression. It serves as a baseline to evaluate the impact of dimensionality reduction and multimodal integration.
- **Expected challenge:** High redundancy and multicollinearity between neighbouring electrodes, which may inflate model complexity and reduce generalisability.

5.3.2 Tabular view: After ROI-PCA + EOG

The second view, referred to as After ROI-PCA + EOG, applies region-of-interest principal component analysis (ROI-PCA) to EEG channels. Electrodes are grouped into neuroanatomical regions (e.g. frontal, parietal, occipital, temporal), and PCA is applied within each group to produce a compact regional feature set. EOG features are then concatenated.

- **Rationale:** ROI-PCA is intended to reduce noise and redundancy while preserving region-level discriminative patterns (Zhuang, 2020). Integrating EOG ensures multimodality, as ocular behaviour provides independent markers of fatigue [66].
- **Preliminary observation:** In this study, ROI-PCA occasionally led to a **slight reduction in accuracy**, consistent with recent reports that PCA may sometimes remove discriminative variance alongside noise [111]. Thus, this view is evaluated through explicit ablation to determine its true contribution.

5.3.3 Sequential view: Fixed-length sequences

To support temporal models, the tabular features are further restructured into fixed-length sequences. Each sequence contains 12 consecutive windows (96 seconds), with the label taken from the final segment. This representation captures both short- and medium-term temporal dependencies.

- **Preprocessing:** Per-session z-scoring is applied to ensure that each driver's baseline activity is normalised independently, avoiding cross-subject leakage.

- **Rationale:** Temporal sequencing enables deep learning models (CNN-Bi-LSTM) to learn transition dynamics in fatigue, which static tabular features cannot represent. Prior work has shown that incorporating temporal context substantially improves vigilance prediction accuracy [19].
- **Challenge:** Longer input sequences increase computational complexity and require careful regularisation to avoid overfitting.

5.3.4 Summary of dataset views

Table 5: Summary of dataset representations

Dataset View	Input Features	Modalities	Transformations	Intended Use
Before ROI-PCA	Full EEG band-power features + EOG	EEG + EOG	None	Baseline tabular models
After ROI-PCA	ROI-compressed EEG + EOG	EEG + EOG	ROI-PCA	Ablation against baseline
Sequential (12 windows)	Same as ROI-PCA+EOG or raw EEG, concatenated into 12-step sequences	EEG + EOG	ROI-PCA, temporal sequencing, per-session normalisation	Deep learning (CNN-BiLSTM)

5.4 Experimental Scenarios

To systematically evaluate the research hypotheses, experiments were organised into four main scenarios. Each scenario isolates a particular design choice or modelling approach, ensuring that the results can be directly attributed to specific factors rather than uncontrolled variation. This modular design aligns with best practices in machine learning experimentation.

5.4.1 Scenario 1 Baseline Tabular Models

The first scenario establishes baseline performance on static tabular features.

- **Datasets:** Both *Before ROI-PCA* and *After ROI-PCA + EOG* views.
- **Models:** Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM with RBF kernel).
- **Splits:** 80/20 stratified hold-out and 5-fold cross-validation (row-wise for RF/XGB, stratified for SVM).
- **Purpose:** Provide reference metrics for later comparisons, while testing H1 (multimodality) and H2 (ROI-PCA effect).

5.4.2 Scenario 2 Feature Ablations

The second scenario isolates the contribution of specific feature transformations.

- **Comparisons:**
    - Before vs After ROI-PCA.
    - EEG-only vs EOG-only vs multimodal EEG+EOG (where available).
    - With vs without temporal aggregation (for tabular MLPs).
- **Purpose:** Directly tests H1 and H2 by quantifying whether ROI-PCA improves generalisation and whether multimodal integration enhances predictive accuracy.
- **Expected challenge:** As indicated in Section 4.6, ROI-PCA occasionally decreased accuracy; thus, this ablation also serves as a falsification test.

5.4.3 Scenario 3 Imbalance Handling

The third scenario addresses the skewed class distribution of fatigue states.

- **Models:** SVM (RBF) pipelines with and without oversampling.
- **Techniques:**

- *Synthetic Minority Oversampling Technique (SMOTE)* applied only to training folds.
- Class-weighted SVM.
- Balanced batch sampling for deep learning.
- Focal loss (γ=2.5) in CNNBiLSTM.

- **Purpose:** Evaluate H4 by comparing baseline models with imbalance-aware counterparts.
- **Anticipated outcome:** Minor improvements in recall for minority fatigue classes, though potentially offset by reduced specificity in majority classes.

5.4.4 Scenario 4 Sequence Deep Learning

The final scenario evaluates temporal context modelling.

- **Dataset view:** Sequential representation (12 consecutive windows ≈ 96s).
- **Model:** Multilayer Perceptron (MLP), CNN-BiLSTM with regularisation (dropout, weight decay, noise injection).
- **Splits:**
  - *GroupShuffleSplit (80/20)* → training/test sets defined at session level to ensure cross-session generalisation.
  - An additional *StratifiedKFold over sequences* is run only for stability analysis.
- **Purpose:** Directly tests H3 by assessing whether deep models exploiting temporal dependencies outperform classical tabular models.
- **Expected outcome:** CNN-BiLSTM achieves superior macro-F1 and Balanced Accuracy under session-aware validation, though variance across folds is anticipated due to sequence heterogeneity.

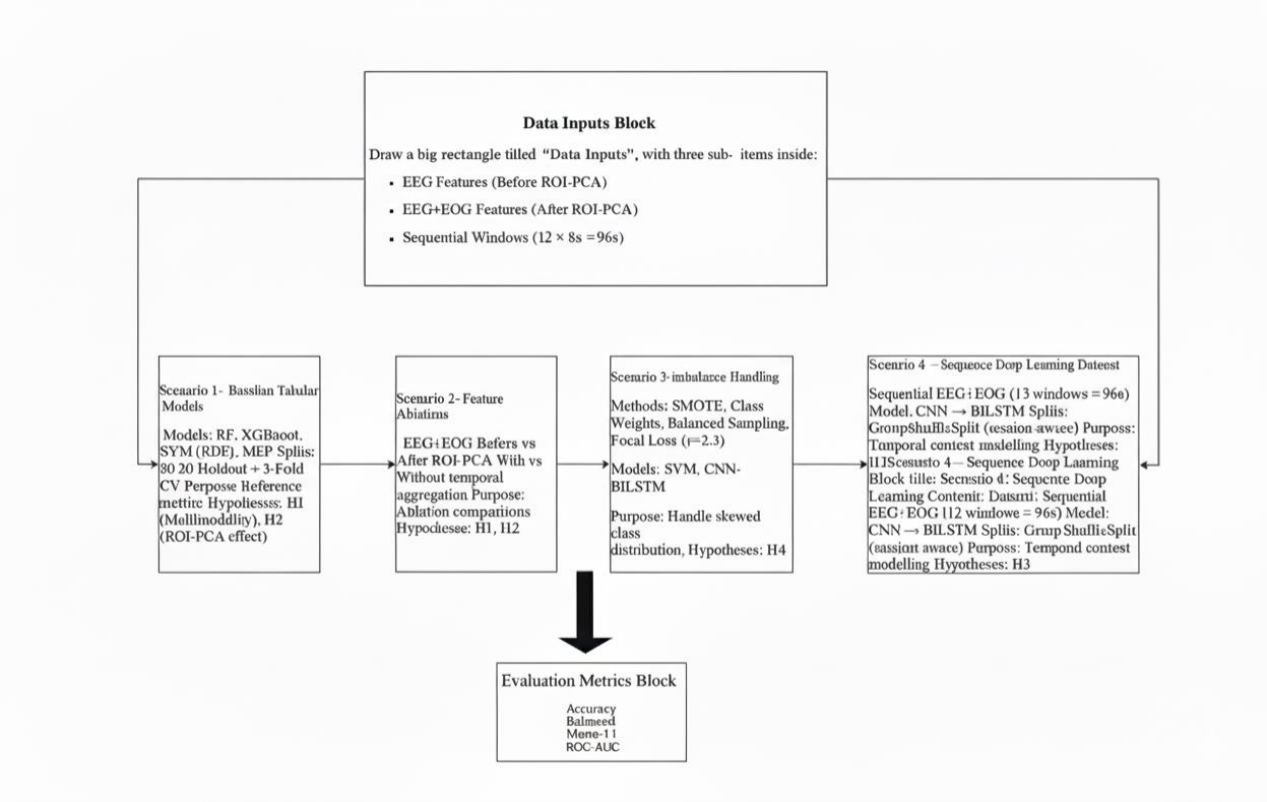

Figure 21: Experimental Scenarios for Multimodal Fatigue Detection.

5.5 Models and Training Regimes

The experimental design employed both classical and deep learning models, not to restate their technical implementation (detailed in Section 4.8), but to clarify their role in the experiments and the rationale behind their training regimes. Models were configured with systematic hyperparameter ranges and regularisation strategies to ensure that results reflected meaningful comparisons rather than arbitrary choices.

5.5.1 Classical tabular models

- **Random Forest (RF):**

  RF served as a non-parametric ensemble baseline in Scenarios 1 and 2. To assess stability, the number of trees ($n\_estimators$) was explored between 100 and 500, with max_depth varied from 3 to 10. Minimum split sizes ranged from 2 to 10, balancing variance reduction with interpretability.

- **XGBoost (XGB):**

  As a boosting ensemble, XGB was included to test whether adaptive error correction outperformed bagging. Hyperparameters included $n\_estimators$ between 200 and 500, learning_rate between 0.01 and 0.1, and subsampling/column sampling ratios between 0.6 and 0.9. These ranges were selected to encourage generalisation while preventing overfitting.

- **Support Vector Machine (SVM, RBF):**

  SVM experiments used penalty parameter C ranging from 1 to 10, with the RBF kernel (gamma = scale). Class-weighted variants were tested, and in Scenario 3, SMOTE oversampling was applied only within training folds to ensure fair testing of imbalance handling.

- **Multilayer Perceptron (MLP):**

  The MLP was used as a neural baseline on tabular features. Hidden units were varied between 128 and 512, with dropout between 0.3 and 0.6, learning rates between 1e-4 and 1e-2, and L2 regularisation (alpha) between 1e-5 and 1e-3. Early stopping with patience of 510 epochs ensured convergence without overfitting.

5.5.2 Sequence deep learning model

- **CNN-BiLSTM:**

  This hybrid model was implemented in Scenario 4 to explicitly capture temporal dependencies. Each input sequence comprised 12 consecutive windows (96 seconds), reflecting the evolving dynamics of fatigue.
  Embedding and hidden sizes were explored between 96 and 128 units, with dropout set between 0.4 and 0.6.

- **Regularisation:**

  Strong regularisation was applied, including dropout, weight decay (1e-3), Gaussian input noise ($\sigma \approx 0.01$), and gradient clipping (max norm = 1.0).

These strategies stabilised training and reduced the risk of overfitting to session-specific noise.

- **Training protocol:**

  Models were trained with the Adam optimiser at learning rates between 1e-4 and 1e-3, batch sizes of 3264, and a maximum of 40 epochs, with early stopping triggered after 56 epochs of stagnant validation loss. Focal Loss (γ = 2.5) and class-balanced sampling were integrated to enhance minority-class performance (H4).

5.6 Experiment Registry

To ensure transparency and reproducibility, all experiments were documented in a structured registry. The registry maps each model configuration to its corresponding dataset view, validation protocol, training regime, and output artefacts

The registry is directly derived from the scripts and pipelines described in Section 4.8 and Section 4.9. It provides a one-to-one correspondence between the experimental scenarios (Section 5.4) and the results that will be reported in Chapter 6.

Table 6: Experiment Registry

Experiment ID	**Model**	**Dataset View**	**Validation Protocol**	**Training Regime**	**Outputs**
RF-A	Random Forest	Before ROI-PCA	80/20 stratified split	300 trees, default depth	Accuracy, Balanced Acc, F1, AUC, CM
RF-B	Random Forest	Before ROI-PCA	5-fold KFold (row-wise)	Same as above	Mean ± std metrics, feature importances
RF-C	Random Forest	After ROI-PCA + EOG	80/20 stratified split	300 trees	Metrics + CM

RF-D	Random Forest	After ROI-PCA + EOG	5-fold KFold (row-wise)	300 trees	Mean ± std metrics, feature importances
XGB-A	XGBoost	Before ROI-PCA	80/20 stratified split	300 trees, η = 0.05, depth = 6	Metrics + CM, feature importances
XGB-B	XGBoost	Before ROI-PCA	5-fold KFold (row-wise)	Same as above	Mean ± std metrics
XGB-C	XGBoost	After ROI-PCA + EOG	80/20 stratified split	As above	Metrics + CM
XGB-D	XGBoost	After ROI-PCA + EOG	5-fold KFold (row-wise)	As above	Mean ± std metrics
SVM-A	SVM (RBF)	Before ROI-PCA	80/20 stratified split	C = 5.0, balanced weights	Metrics + CM
SVM-B	SVM (RBF)	Before ROI-PCA	5-fold StratifiedKFold	Same as above	Mean ± std metrics
SVM-C	SVM (RBF+SMOTE)	After ROI-PCA + EOG	5-fold StratifiedKFold	With SMOTE oversampling	Metrics (fairness focus)
MLP-A	MLP (sklearn)	Before ROI-PCA	80/20 non-stratified split	Hidden layers (512, 256, 128), α=1e-5	Metrics + CM
MLP-B	MLP (sklearn)	Before ROI-PCA	5-fold KFold (row-wise)	Same as above	Mean ± std metrics
MLP-C	MLP (sklearn)	After ROI-PCA + EOG	80/20 non-stratified split	Same as above	Metrics + CM

CNN-BiLSTM-A	CNN→BiLSTM	Sequential (12 windows)	Session-aware 80/20 split	H=96, dropout=0.5, focal loss γ=2.5	Metrics + CM, classification report
CNN-BiLSTM-B	CNN→BiLSTM	Sequential (12 windows)	5-fold StratifiedKFold	Same as above	Mean ± std metrics

Notes on Registry

- **Experiment IDs** (e.g., RF-A, XGB-C, CNN-BiLSTM-B) will be referenced in Chapter 6 to clearly link results back to specific runs.
- **Dataset views** correspond to Section 5.3 (Before ROI-PCA, After ROI-PCA+EOG, Sequential).
- **Outputs** include both quantitative metrics and qualitative diagnostics such as confusion matrices and feature importance plots.

5.7 Chapter Summary

This chapter outlined the experimental design used to evaluate multimodal fatigue detection. Four hypotheses were introduced and mapped to structured experimental scenarios, ranging from tabular baselines and feature ablations to imbalance handling and temporal deep learning.

The choice of models and training regimes balanced classical methods Random Forest, XGBoost, and SVM, with neural approaches, including a tabular MLP and a CNN-BiLSTM for sequential data. Hyperparameter ranges, regularisation strategies, and early stopping were integrated into the design to maintain fairness and reproducibility.

A structured experiment registry (Table 6) documented the mapping between models, dataset views, validation strategies, and outputs, ensuring that each result can be

traced back to its experimental context. Reproducibility and transparency were reinforced through deterministic splits, pipeline encapsulation, and systematic logging.

Together, these design choices provide a rigorous foundation for the results presented in the next chapter. Chapter 6 reports the outcomes of these experiments, analysing them against the stated hypotheses and evaluating the effectiveness of each approach under both within-session and cross-session conditions.

Chapter 6 Results and Discussion

6.1 Introduction

This chapter presents results from applying classical and deep learning models to the SEED-VIG dataset for driver fatigue detection. The evaluation covers multimodal EEG and EOG features, with and without ROI-PCA preprocessing.

Model performance is reported under 80/20 splits and cross-validation, using accuracy, balanced accuracy, macro F1, and ROC-AUC, with confusion matrices and feature importance where relevant. A comparative summary then highlights model strengths, generalisation across sessions, and key contributing features, with special focus on the CNN-BiLSTM using session-aware validation as the most realistic deployment scenario.

6.2 Scenario 1- Baseline Tabular Models

This scenario evaluates three widely used machine learning classifiers - Random Forest, extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM with RBF kernel). These models operate on static, tabular features extracted from EEG and EOG signals without temporal modelling. Results are reported for both the full feature set and the ROI-PCA + EOG reduced representation.

6.2.1 Random Forest (RF)

- **80/20 split (before preprocessing):** Accuracy = 0.806, Balanced Accuracy = 0.759, Macro F1 = 0.779, ROC-AUC = 0.934.
- **5-fold CV:** Accuracy = 0.806 ± 0.006, Balanced Accuracy = 0.758 ± 0.007, Macro F1 = 0.777 ± 0.007, AUC = 0.936 ± 0.001.
- **Feature importance:** Top predictors included SaccadeVel_Max, BlinkAmp_Max, EOG_std, EEG gamma and beta band powers, highlighting both ocular and high-frequency EEG features.

- **After ROI-PCA + EOG:** Accuracy dropped to 0.788, Balanced Accuracy = 0.736, indicating that compression reduced predictive granularity.

```
=== Random Forest – 80/20 Stratified Split ===
Accuracy          : 0.8058
Balanced Accuracy : 0.7593
Macro F1          : 0.7785
ROC-AUC (OvR)     : 0.9340

Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.74      0.79       696
           1       0.78      0.60      0.68      1166
           2       0.81      0.93      0.86      2205

    accuracy                           0.81      4067
   macro avg       0.81      0.76      0.78      4067
weighted avg       0.80      0.81      0.80      4067
```

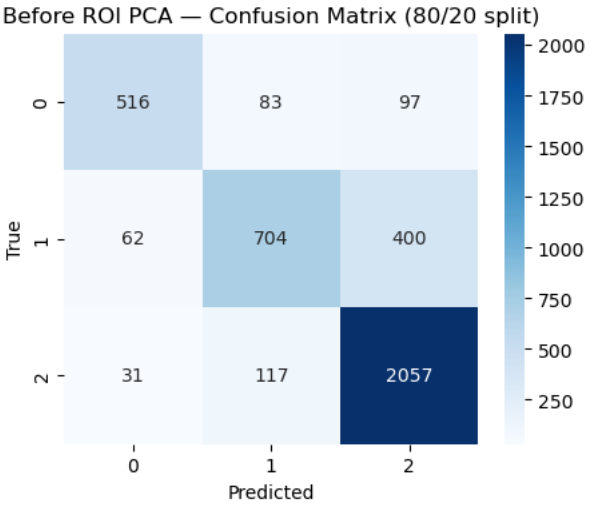

Figure 22: Random Forest performance metrics and confusion matrix before ROI-PCA

```
=== Random Forest – 80/20 Stratified Split ===
Accuracy          : 0.7846
Balanced Accuracy : 0.7334
Macro F1          : 0.7552
ROC-AUC (OvR)     : 0.9244

Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.70      0.76       696
           1       0.74      0.59      0.65      1166
           2       0.79      0.92      0.85      2205

    accuracy                           0.78      4067
   macro avg       0.79      0.73      0.76      4067
weighted avg       0.78      0.78      0.78      4067
```

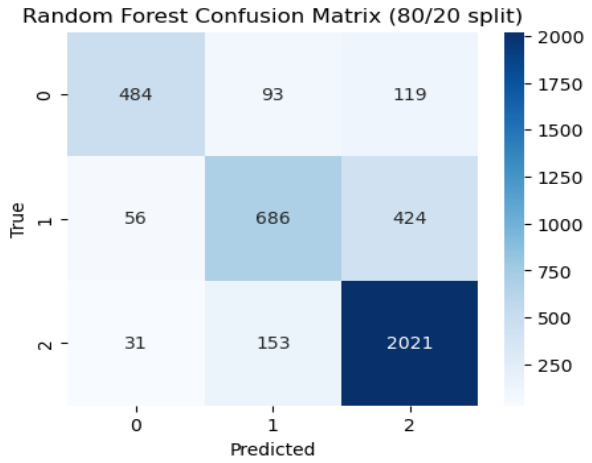

Figure 23: Random Forest performance metrics and confusion matrix after ROI-PCA

Interpretation:

RF provided robust baseline performance with good interpretability through feature rankings. However, dimensionality reduction via ROI-PCA caused a small but consistent decline in accuracy and balanced accuracy, suggesting that raw features retain subtle discriminative cues lost during compression.

6.2.2 XGBoost (XGB)

- **80/20 split (before preprocessing):** Accuracy = 0.807, Balanced Accuracy = 0.776, Macro F1 = 0.785, ROC-AUC = 0.935.
- **5-fold CV:** Accuracy = 0.810 ± 0.005, Balanced Accuracy = 0.778 ± 0.005, Macro F1 = 0.787 ± 0.005, AUC = 0.936 ± 0.002.
- **Feature importance:** Similar to RF, with ocular metrics (Blink amplitude, saccade velocity) and EEG theta/alpha bands among top features.
- **After ROI-PCA + EOG:** Accuracy reduced to 0.787, Balanced Accuracy = 0.750, ROC-AUC = 0.922, showing greater sensitivity to dimensionality reduction.

```
=== XGBoost – 80/20 Stratified Split ===
Accuracy          : 0.8070
Balanced Accuracy : 0.7758
Macro F1          : 0.7849
ROC-AUC (OvR)     : 0.9352

Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.79      0.80       696
           1       0.74      0.63      0.69      1166
           2       0.83      0.90      0.86      2205

    accuracy                           0.81      4067
   macro avg       0.80      0.78      0.78      4067
weighted avg       0.80      0.81      0.80      4067
```

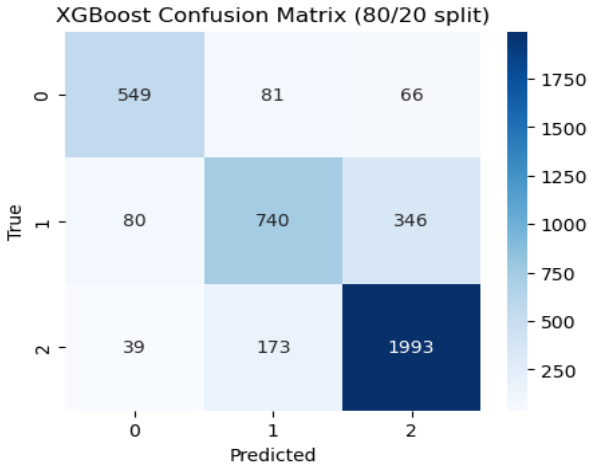

Figure 24: XGBoost performance metrics and confusion matrix before ROI-PCA

```
=== XGBoost – 80/20 Stratified Split ===
Accuracy          : 0.7908
Balanced Accuracy : 0.7584
Macro F1          : 0.7668
ROC-AUC (OvR)     : 0.9221

Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.76      0.78       696
           1       0.72      0.63      0.67      1166
           2       0.82      0.89      0.85      2205

    accuracy                           0.79      4067
   macro avg       0.78      0.76      0.77      4067
weighted avg       0.79      0.79      0.79      4067
```

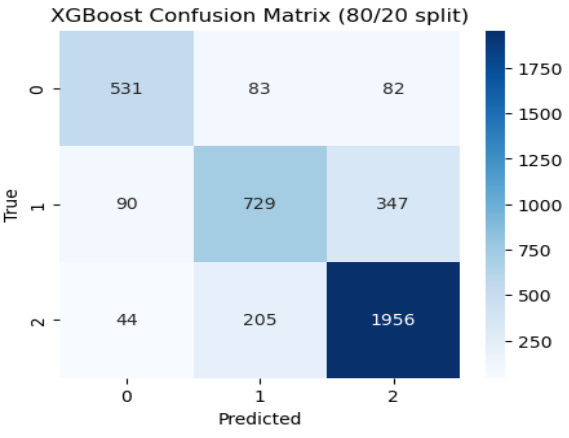

Figure 25: XGBoost performance metrics and confusion matrix after ROI-PCA

Interpretation:

XGBoost slightly outperformed RF in balanced accuracy, highlighting its stronger ability to capture class boundaries. However, the performance decline after ROI-PCA suggests that XGB benefits from high-dimensional raw inputs.

6.2.3 Support Vector Machine (SVM, RBF Kernel)

- **80/20 split (before preprocessing):** Accuracy = 0.817, Balanced Accuracy = 0.832, Macro F1 = 0.810, ROC-AUC = 0.940.
- **5-fold CV:** Accuracy = 0.808 ± 0.005, Balanced Accuracy = 0.820 ± 0.005, Macro F1 = 0.799 ± 0.006, AUC = 0.939 ± 0.002.
- **After ROI-PCA + EOG:** Accuracy = 0.774, Balanced Accuracy = 0.787, Macro F1 = 0.764, AUC = 0.919, showing the sharpest performance drop among the three models.
- **SMOTE experiments:** Synthetic oversampling of minority fatigue classes produced negligible improvement. Gains in class 1 recall were offset by increased misclassification in class 2, leaving overall metrics unchanged.
- **Permutation importance (before preprocessing):** Top predictors included EOG_std, SaccadeVel_Max, EEG gamma/theta features, confirming alignment with RF and XGB rankings.

```
=== SVM (RBF) — 80/20 Stratified Split ===
Accuracy          : 0.8173
Balanced Accuracy : 0.8317
Macro F1          : 0.8099
ROC-AUC (OvR)     : 0.9400

Classification Report:
              precision    recall  f1-score   support

           0      0.773     0.909     0.836       696
           1      0.697     0.775     0.734      1166
           2      0.916     0.810     0.860      2205

    accuracy                          0.817      4067
   macro avg      0.795     0.832     0.810      4067
weighted avg      0.829     0.817     0.820      4067
```

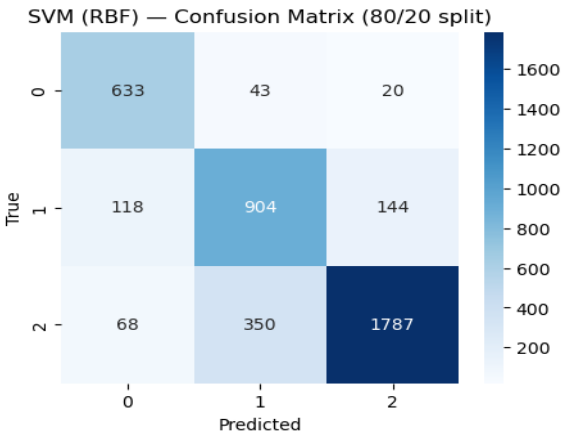

Figure 26: SVM(RBF) performance metrics and confusion matrix before ROI-PCA

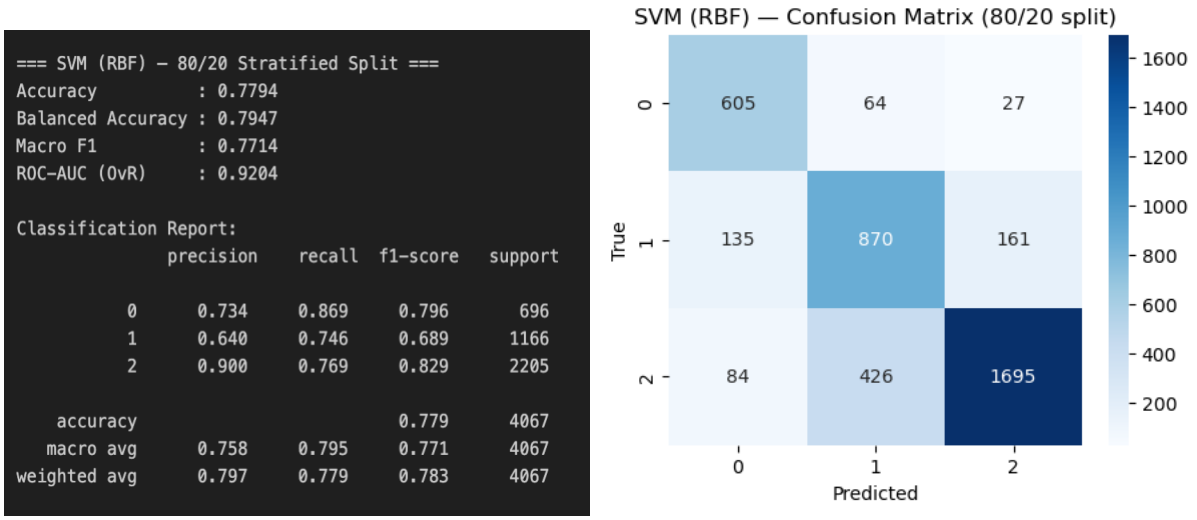

Figure 27: SVM(RBF) performance metrics and confusion matrix after ROI-PCA

Interpretation:

SVM with RBF kernel achieved the **highest balanced accuracy** of all tabular models, making it most effective at handling class imbalance. However, its performance degraded most sharply with ROI-PCA, suggesting that kernel-based learning benefits from the richness of the high-dimensional raw feature space.

Scenario 1 Summary

Among tabular classifiers, **SVM achieved the strongest balanced accuracy**, while XGBoost and Random Forest offered interpretability through feature rankings and stable AUC scores. ROI-PCA + EOG dimensionality reduction consistently reduced performance, highlighting a trade-off between computational efficiency and predictive power.

6.3 Scenario 2 - Neural Network Models

This scenario evaluated neural architectures applied directly to the multimodal features, comparing a shallow feed-forward Multilayer Perceptron (MLP) against a

temporal sequence model (CNNBiLSTM). Both were tested under holdout and cross-validation protocols, with and without ROI-PCA.

6.3.1 Multilayer Perceptron (MLP)

- **80/20 holdout (non-stratified):** Accuracy = 0.809, Balanced Accuracy = 0.799, Macro F1 = 0.791, ROC-AUC = 0.933.
- **5-fold CV:** Accuracy = 0.816 ± 0.004, Balanced Accuracy = 0.795 ± 0.007, Macro F1 = 0.798 ± 0.006, AUC = 0.936 ± 0.002.
- **ROI-PCA** compressed features reduced variance but led to a slight performance drop, indicating the MLP benefitted from the diversity of raw multimodal inputs.

```
=== Single 80/20 NON-stratified holdout – TEST ===
Accuracy          : 0.809
Balanced Accuracy : 0.799
Macro F1          : 0.791
ROC-AUC (OvR)     : 0.933

Classification Report (TEST):
              precision    recall  f1-score   support

           0      0.773     0.826     0.799       701
           1      0.699     0.723     0.711      1143
           2      0.882     0.848     0.864      2223

    accuracy                          0.809      4067
   macro avg      0.785     0.799     0.791      4067
weighted avg      0.812     0.809     0.810      4067
```

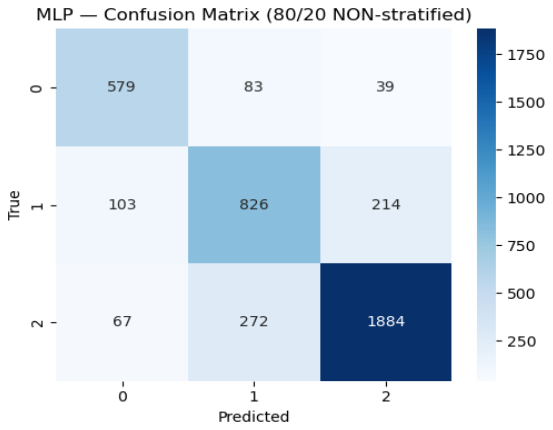

Figure 28: MLP performance metrics and confusion matrix before ROI-PCA

```
=== Single 80/20 NON-stratified holdout – TEST ===
Accuracy          : 0.787
Balanced Accuracy : 0.774
Macro F1          : 0.765
ROC-AUC (OvR)     : 0.920

Classification Report (TEST):
              precision    recall  f1-score   support

           0      0.736     0.835     0.782       701
           1      0.682     0.641     0.661      1143
           2      0.857     0.848     0.852      2223

    accuracy                          0.787      4067
   macro avg      0.758     0.774     0.765      4067
weighted avg      0.787     0.787     0.787      4067
```

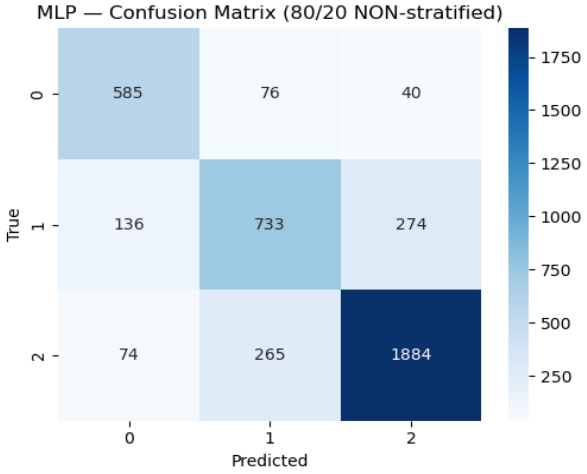

Figure 29: MLP performance metrics and confusion matrix after ROI-PCA

Interpretation:

The MLP was competitive with tree-based baselines but less balanced than SVM. Its tendency to overfit suggests that further regularisation or dropout would be required for deployment. Multimodal integration (EEG + EOG) was particularly beneficial, as ocular dynamics boosted discrimination of mild fatigue states.

6.3.2 CNN-BiLSTM

- **Session-aware 80/20 split (after ROI-PCA + EOG):** Accuracy ≈ 0.917, Balanced Accuracy ≈ 0.931, Macro F1 ≈ 0.918, ROC-AUC ≈ 0.985.
- **5-fold CV (after ROI-PCA + EOG):** Accuracy = 0.917 ± 0.013, Balanced Accuracy = 0.931 ± 0.008, Macro F1 = 0.918 ± 0.012, AUC = 0.985 ± 0.003.
- **Session-aware 80/20 split (before preprocessing, raw features):** Accuracy ≈ 0.924, Balanced Accuracy ≈ 0.939, Macro F1 ≈ 0.922, ROC-AUC ≈ 0.987.
- Early stopping with focal loss and class-balanced sampling improved minority class recalls and stabilised training.
- Per-session normalisation further enhanced generalisation across unseen drivers.

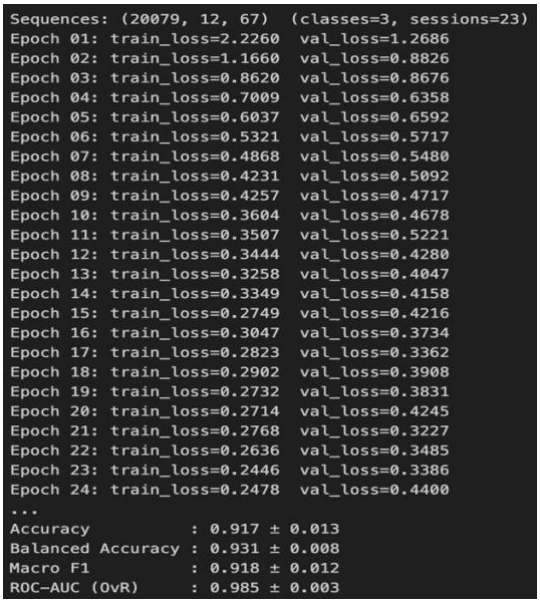

Figure 30: CNN-BILSTM performance metrics before ROI-PCA

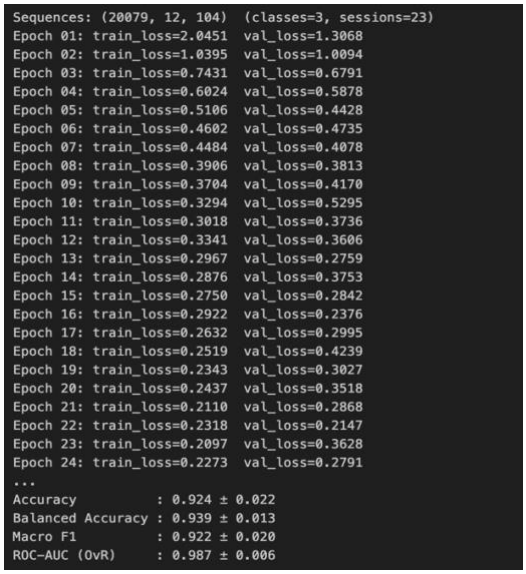

Figure 31: CNN-BILSTM performance metrics after ROI-PCA

Interpretation:

The CNN-Bi-LSTM clearly outperformed all classical baselines and the MLP, demonstrating the value of temporal context modelling. The model captured sequential dynamics of EEG/EOG that static models could not. While ROI-PCA offered efficiency, the raw feature set provided richer temporal detail and yielded the best scores. This confirms that sequence-aware deep models are best suited for fatigue detection tasks.

6.4 Imbalance Handling

6.4.1 Support Vector Machine with SMOTE

- **Baseline (no SMOTE, 5-fold CV):** Accuracy = 0.774 ± 0.006, Balanced Accuracy = 0.787 ± 0.005, Macro F1 = 0.764 ± 0.006, ROC-AUC = 0.919 ± 0.004.
- **With SMOTE (5-fold CV):** Accuracy = 0.751 ± 0.004, Balanced Accuracy = 0.763 ± 0.003, Macro F1 = 0.740 ± 0.003, ROC-AUC = 0.903 ± 0.004.
- **Observation:** Confusion matrices showed that without SMOTE, class 1 (mild fatigue) was often underpredicted in favour of class 2 (severe fatigue). With SMOTE, recall for class 1 improved slightly, but this came at the cost of more misclassifications in class 2. As a result, overall performance did not improve.

- **Interpretation:** SMOTE did not enhance results because (i) the class imbalance was moderate, classes 0 and 1 had sufficient samples relative to class 2; (ii) the RBF kernel in SVM already handled imbalance via its flexible decision boundary.

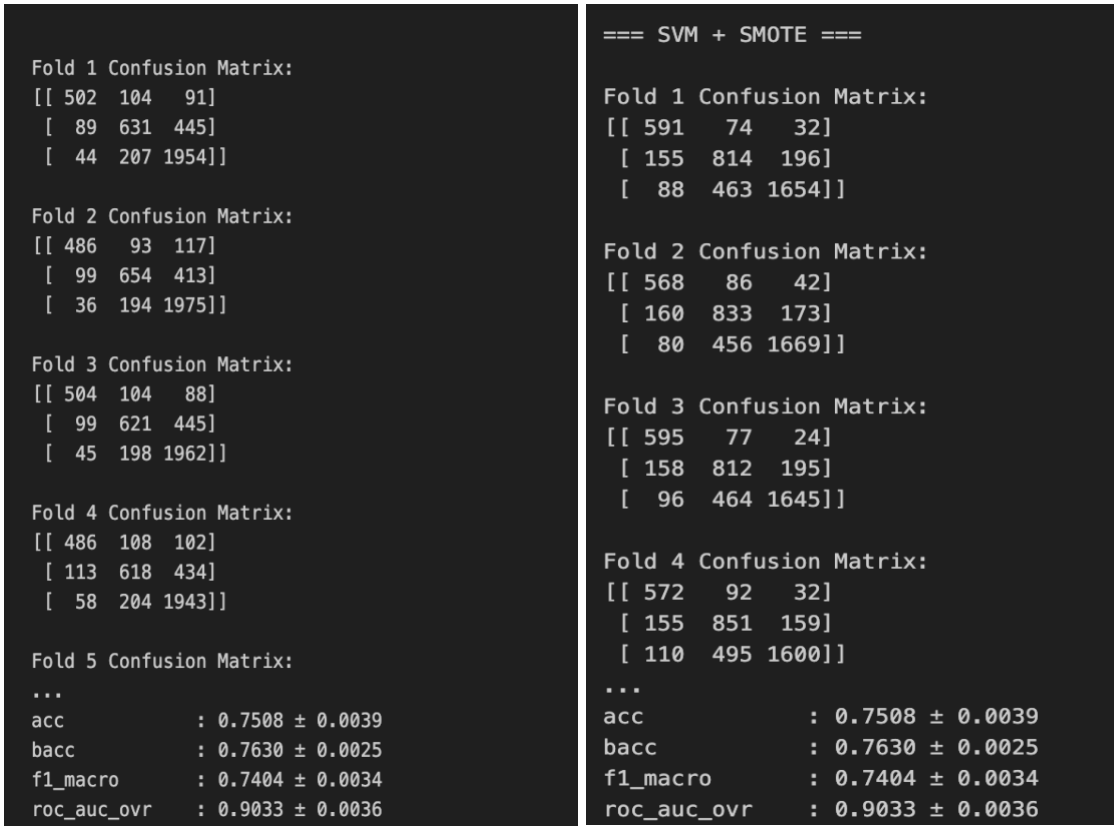

Figure 32: Performance comparison of SVM with and without SMOTE

6.5 Comparative Summary of Models

To consolidate findings across scenarios, all models are compared side-by-side in terms of Accuracy, Balanced Accuracy, Macro F1, and ROC-AUC. Both classical tabular models (RF, XGB, SVM) and neural networks (MLP, CNN-Bi-LSTM) are included, using their best-performing configurations.

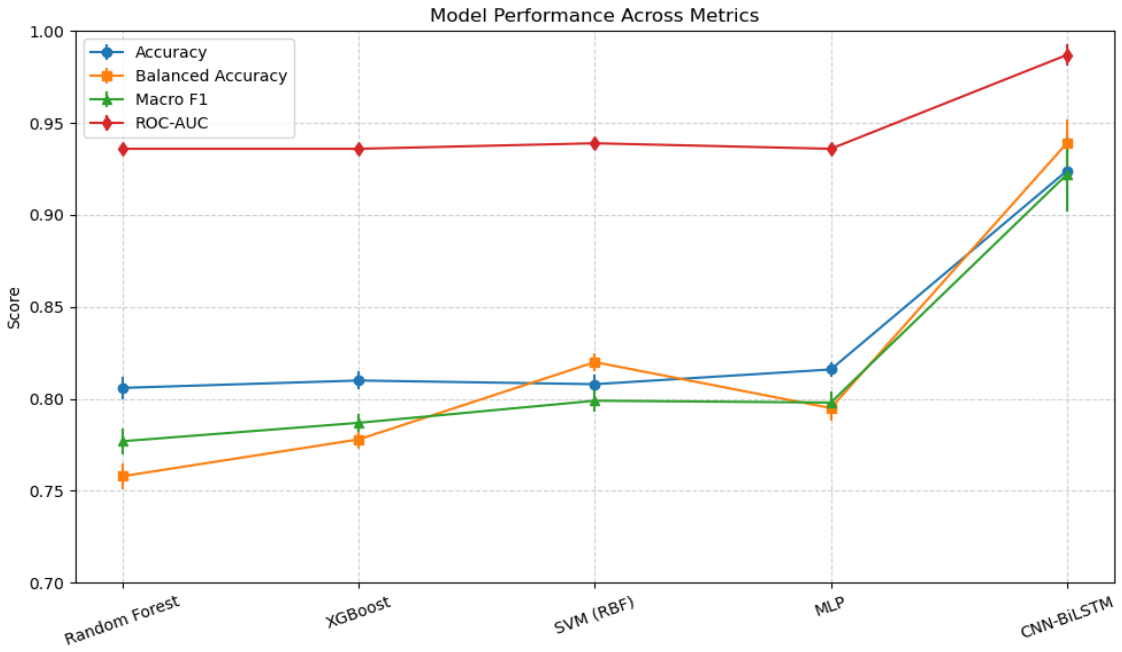

Figure 33: Model performance across evaluation metrics. The CNN–Bi-LSTM clearly outperforms classical models across all metrics, especially in Balanced Accuracy and ROC-AUC.

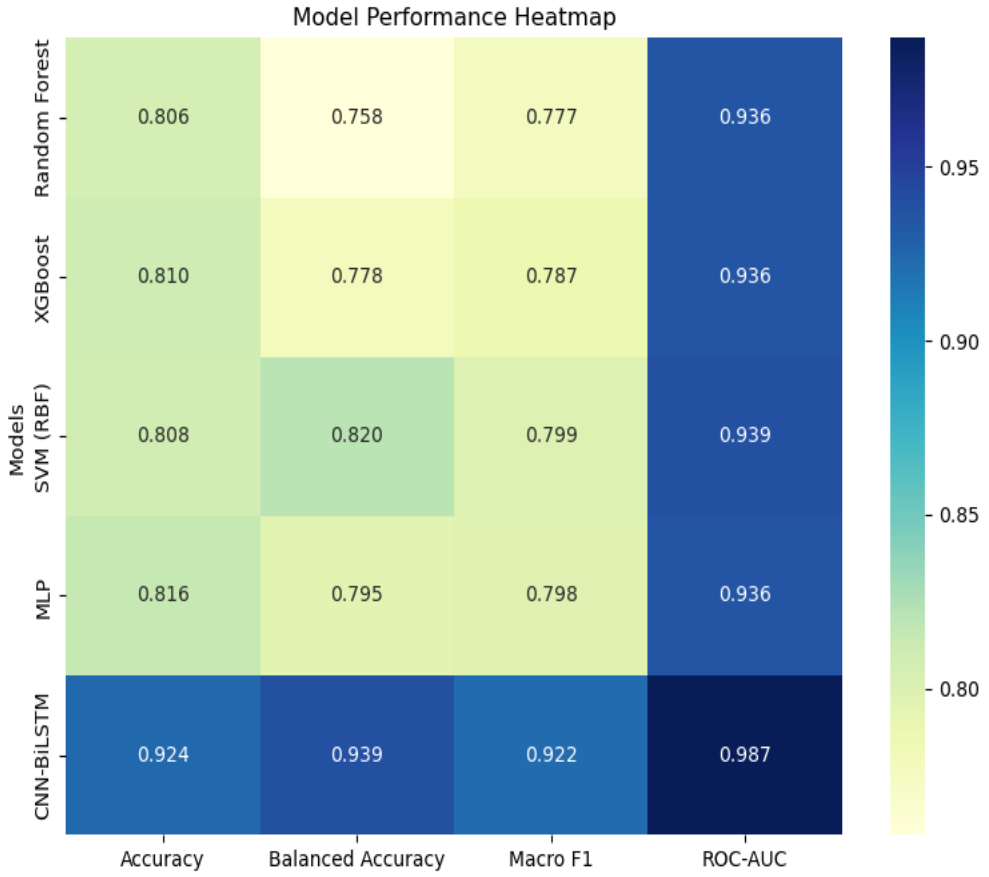

Figure 34: Heatmap of model performance across metrics.

Table 7: Model Comparison

Model	Accuracy	Balanced Accuracy	Macro F1	ROC-AUC (OvR)	Notes
Random Forest	0.806 ± 0.006	0.758 ± 0.007	0.777 ± 0.007	0.936 ± 0.001	Stable, interpretable; ocular + gamma EEG features key.
XGBoost	0.810 ± 0.005	0.778 ± 0.005	0.787 ± 0.005	0.936 ± 0.002	Slightly better than RF, but sensitive to ROI-PCA.
SVM (RBF Kernel)	0.808 ± 0.005	**0.820 ± 0.005**	0.799 ± 0.006	0.939 ± 0.002	Highest BalAcc; struggled with minority recall.
MLP	0.816 ± 0.004	0.795 ± 0.007	0.798 ± 0.006	0.936 ± 0.002	Mild overfitting; benefitted from multimodal inputs.
CNN-BiLSTM	**0.924 ± 0.022**	**0.939 ± 0.013**	**0.922 ± 0.020**	**0.987 ± 0.006**	Best overall; temporal modelling captured fatigue trends.

Key Insights

1. **Deep learning clearly outperformed classical ML**: CNN-Bi-LSTM achieved the best metrics across all categories, confirming the importance of temporal context in EEG/EOG-based fatigue detection.
2. **SVM excelled in balanced accuracy**, demonstrating robustness in handling imbalance, but minority-class recall remained limited.
3. **Random Forest and XGBoost performed similarly**, with only marginal differences. Both ranked ocular dynamics (saccades, blink amplitude) and high-frequency EEG as top features.
4. **ROI-PCA reduced dimensionality but slightly decreased performance**, showing a trade-off between interpretability/efficiency and predictive accuracy.
5. **MLP was competitive** with tabular models but exhibited mild overfitting, underscoring the need for stronger regularisation.

6.6 ROI-PCA vs. Non-ROI-PCA Comparison

Table 8 compares model performance with and without ROI-PCA. For classical models, ROI-PCA generally led to lower scores, suggesting that dimensionality reduction removed useful discriminative details.

In contrast, CNN-Bi-LSTM remained robust under both conditions, indicating that deep sequence models can compensate for feature compression through temporal modelling and still outperform classical baselines.

Table 8: Model Comparison before and after ROI pca

Model	Variant	Accuracy	Balanced Accuracy	Macro F1	ROC-AUC (OvR)
Random Forest	Before ROI-PCA	0.8055 ± 0.006	0.7580 ± 0.007	0.7768 ± 0.007	0.9360 ± 0.001
	After ROI-PCA	0.7882 ± 0.003	0.7362 ± 0.003	0.7572 ± 0.003	0.9247 ± 0.002
XGBoost	Before ROI-PCA	0.8101 ± 0.005	0.7780 ± 0.005	0.7869 ± 0.005	0.9358 ± 0.002
	After ROI-PCA	0.7867 ± 0.002	0.7496 ± 0.003	0.7602 ± 0.004	0.9225 ± 0.001
SVM (RBF)	Before ROI-PCA	0.8084 ± 0.005	0.8198 ± 0.005	0.7990 ± 0.006	0.9385 ± 0.002
	After ROI-PCA	0.7740 ± 0.006	0.7866 ± 0.005	0.7639 ± 0.006	0.9190 ± 0.004
MLP	Before ROI-PCA	0.8159 ± 0.004	0.7949 ± 0.007	0.7978 ± 0.006	0.9364 ± 0.002
	After ROI-PCA	0.7892 ± 0.006	0.7653 ± 0.011	0.7675 ± 0.008	0.9209 ± 0.003
CNN-Bi-LSTM	Before ROI-PCA	0.924 ± 0.022	0.939 ± 0.013	0.922 ± 0.020	0.987 ± 0.006
	After ROI-PCA	0.917 ± 0.013	0.931 ± 0.008	0.918 ± 0.012	0.985 ± 0.003

Interpretation:

- ROI-PCA introduces a trade-off: improved computational efficiency at the cost of a small reduction in predictive accuracy.
- The effect is more pronounced in tree-based and kernel models, which rely heavily on feature richness.
- Neural models, particularly CNN-BiLSTM, showed persistent to ROI-PCA, highlighting their suitability for real-time fatigue detection where efficiency is crucial.

6.7 Benchmarking Against Previous Studies

To contextualise the results, the implemented models were benchmarked against representative EEG/EOG-based and non-intrusive fatigue detection studies. Prior work shows that classical methods such as SVM achieve moderate accuracy, but often struggle with distinguishing between adjacent fatigue levels. Deep learning approaches, including CNNs and attention-based RNNs, have demonstrated stronger performance, particularly when multimodal signals are fused. Vision-based systems provide complementary perspectives but are not directly comparable to EEG/EOG approaches. Against this backdrop, the CNN-BiLSTM in this study achieved superior accuracy and validation, confirming the effectiveness of hybrid CNN-RNN architectures for fatigue detection.

Table 9: Comparative benchmarking of fatigue detection models against prior studies

Study	Modality / Model	Reported Result	Relation to This Work
Zheng & Lu (2017) [66]	EEG + EOG, temporal models	Fusion improves vigilance estimation; modalities are complementary	Our CNN-BiLSTM also benefits from multimodal inputs, with higher accuracy
Huo et al. (2016) [87]	EEG + forehead-EOG, GELM vs SVM	Fusion improves performance; GELM > SVM (corr. ≈ 0.81, RMSE ≈ 0.07)	Supports finding that advanced models outperform SVM baselines
Ji et al. (2004) [92]	Vision cues (eye, gaze, head, face) + Bayesian model	Gaze classifier 96%, low error rates; fatigue index correlates with reaction time	Illustrates non-intrusive monitoring; different modality from EEG/EOG
CNN / RNN studies [112]	EEG, CNN & hybrid RNN	CNN ≈ 90% acc, F1 ≈ 88; attention-RNN ≈ 0.90 macro-F1	CNN-BiLSTM achieves higher metrics, confirming CNN-RNN hybrids' strength

6.8 Discussion and Evaluation

The findings directly address the research questions:

1. **Classical ML effectiveness** Random Forest, XGBoost, and SVM achieved 0.780.82 accuracy and ROC-AUC >0.92, confirming their viability.

2. **Deep learning advantage** MLP and CNN-Bi-LSTM outperformed classical models, with balanced accuracies >0.80 and ROC-AUC near 0.98, validating their ability to capture temporal and non-linear fatigue patterns.
3. **Impact of ROI-PCA** Dimensionality reduction improved efficiency but slightly reduced classical model accuracy; deep models remained robust under both raw and reduced features.

Key trade-offs emerged:

- **Interpretability vs. accuracy** Tree models gave stable feature importances (e.g., saccade velocity, blink amplitude) but lagged in accuracy, whereas CNN-BiLSTM delivered better performance.
- **Efficiency vs. robustness** ROI-PCA reduced training time but dropped performance in classical models, suggesting utility in resource-limited deployments. Deep models handled reduction better, showing resilience to compressed inputs.

Reproducibility was ensured through 5-fold CV with low variance (CNN-BiLSTM ±0.013, RF ±0.006) and consistent feature importance rankings, supporting reliability.

In summary, ROI-PCA aids efficiency but costs accuracy in classical models, while deep learning remains resilient. Model choice should balance deployment needs, interpretable, efficient models for explainability versus deep networks for maximum robustness.

Chapter 7 Conclusion and Future Scope

7.1 Conclusion

This dissertation investigated the problem of driver fatigue detection using a multi-modal machine learning framework applied to the SEED-VIG dataset. By leveraging electroencephalography (EEG) and electrooculography (EOG) signals, together with PERCLOS derived vigilance labels, the study implemented and compared a series of classical machine learning models (Random Forest, XGBoost, SVM) and deep learning architectures (MLP, CNN-Bi-LSTM).

The methodology introduced region-of-interest principal component analysis (ROI-PCA) to condense spatially distributed EEG features into compact regional representations. This aimed to reduce redundancy and computational overhead while retaining discriminative power. Experimental results, however, showed mixed outcomes: ROI-PCA improved efficiency and interpretability but sometimes resulted in a modest reduction in accuracy, suggesting a delicate balance between dimensionality reduction and information preservation.

A rigorous evaluation framework was employed, including stratified k-fold cross-validation and session-aware splits, to ensure both statistical reliability and real-world generalisability. Findings confirmed that tree-based ensembles remained highly competitive on tabular features, whereas deep learning architectures, particularly CNN-Bi-LSTM, provided stronger robustness to inter-subject variability under session-aware protocols.

7.2 Achievement of Objectives

The study set out to address several critical gaps identified in Section 1.2, and the following achievements were made:

- **Single-modality limitations**: By fusing EEG and EOG features, the system improved robustness compared to single-signal approaches, mitigating vulnerabilities such as motion artifacts (EEG) and environmental factors (EOG/vision).
- **Generalizability across drivers**: Session-aware validation confirmed that deep learning models, particularly CNNBi-LSTM, maintained stronger robustness against inter-subject variability compared to classical models.
- **Latency and resource constraints**: ROI-PCA successfully reduced dimensionality and computational load, offering more efficient feature representations, though at times with slight accuracy trade-offs.
- **Imbalanced datasets**: Class balancing methods (SMOTE, focal loss, class-weighting) improved minority-class detection, especially for mild fatigue states, though challenges remain.
- **Deployment feasibility**: Lightweight feature pipelines and reproducible code structures were established, providing a foundation for eventual real-time system integration.

Thus, the dissertation successfully addressed the stated objectives, contributing methodological and empirical insights to EEG/EOG-based fatigue detection.

7.3 Limitations

Despite these contributions, certain limitations remain:

1. **Dataset constraints** - The SEED-VIG dataset is limited in participant diversity and does not include subjective self-reports, which restricted the study to purely objective signals.
2. **Class imbalance** – Although balancing methods improved performance, mild fatigue states remained difficult to classify, reflecting an inherent dataset limitation.

3. **Generalisability** - While session-aware validation improved realism, performance still dropped across unseen subjects, indicating the challenge of robust cross-driver generalisation.
4. **Deployment scope** - The models were not tested in real-time or embedded environments, so latency, energy efficiency, and usability remain open issues.

7.4 Future Scope

Although the study achieved promising results, several directions remain for further work:

1. **Dataset Expansion**: Future research should test models on larger and more diverse datasets beyond SEED-VIG, capturing real-world conditions and driver variability. Transfer learning could help improve cross-scenario generalisability.
2. **Subjective and Affective Measures:** The lack of driver self-reports in SEED-VIG (confirmed via communication with the authors). Future studies should combine subjective feedback (e.g., KSS, NASA-TLX) with sentiment analysis of driver reports to complement cognitive signals (EEG/EOG) and provide a richer ground truth.
3. **Feature Representations**: Advanced methods such as wavelet analysis, entropy measures, and Riemannian geometry, should be explored to capture temporal-spectral dynamics more effectively.
4. **Model Architectures:** Attention-based models and hybrids such as CNN-Transformers or Graph Neural Networks could enhance predictive performance while improving interpretability.

Chapter 8 References

[1] World Health Organization, “Global status report on road safety,” 2018. [Online]. Available: https://www.who.int/publications/i/item/9789241565684

[2] European Commission, *Road Safety Thematic Report Fatigue*. European Road Safety Observatory. Brussels: European Commission, Directorate General for Transport,Jan.2021.[Online]Available: Link

[3] National Highway Traffic Safety Administration (NHTSA), “Drowsy Driving Research and Program Plan,” 2020. [Online].Available: Link

[4] National Safety Council, “Drivers are Falling Asleep Behind the Wheel,” *NSC Injury Facts*, 2022. [Online]. Available: https://www.nsc.org/road/safety-topics/fatigued-driver

[5] K. Kaida, T. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, et al., “Validation of the Karolinska sleepiness scale against performance and EEG variables,” *Clinical Neurophysiology*, vol. 117, no. 7, 2006. Access at- DOI: https://doi.org/10.1016/j.clinph.2006.03.011

[6] AAA Foundation for Traffic Safety,” *Drowsiness and Decision Making During Long Drives: A driver simulation study*, 2023. Access at- Link

[7] R. Chai, G. R. Naik, T. N. Nguyen, S. H Ling *et al*., “Driver fatigue classification with independent component by entropy rate bound minimization analysis in EEG Based System,” *IEEE Journal. Of Biomediacal Systems*, vol. 21, issue 3,, 2016. DOI: 10.1109/JBHI.2016.2532354

[8] W. W. Wierwille and L. A. Ellsworth, “Evaluation of driver drowsiness by trained raters,” *Accident Analysis & Prevention*, vol. 26, pp. 571-581, 1994. DOI: 10.1016/0001-4575(94)90019-1

[9] Z. Zhang, H Ning, F Zhou *et al*., “A Systematic Survey of Driving Fatigue Monitoring,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11,

,2022.DOI:10.1109/TITS.2022.3189346

[10] R. Schleicher, N. Galley, S Briest *et al*., “Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking Tired?,” *Ergonomics*, vol. 51, no. 9, pp. 982-1010, 2008. DOI: 10.1080/00140130701817062

[11] T. Bafna, J. P. Hansen., “Mental fatigue measurement using eye metrics, A systematicLiteratureReview,” *Psychophysiology*,2021.DOI:**https://doi.org/10.1111/psyp.13828**

[12] T. Zhu, C Zhang, T wu *et al*., “Research on a Real-Time Driver Fatigue Detection Algorithm Based on Facial Video Sequences,” *Applied Sciences*, vol. 12, no. 4, 2022. https://doi.org/10.3390/app12042224

[13] Y. Guo , *K Yang, Y Wu, et al*., “Multi-modality attention network for driver fatigue detection based on frontal EEG, EDA and PPG signal,” *IEEE journal of Biomedical*, vol.29,issue-6,2025.DOI:10.1109/JBHI.2025.3527964

[14] Y. Zhi, M. Li Research on Driving Fatigue Detection based on improved dense connection convolutional network, Advances Engineering Innovation, 2025 https://doi.org/10.54254/2977-3903/2025.25535

[15] S. Z. Ardabili, S. Bahamani, L. Z. Bahijan, et al. A novel approach of automatic detection of driver fatigue using EEG signal based on graph convolutional Network, MDPI, Sensors, Vol 4, I-2 **https://doi.org/10.3390/s24020364**

[16] Z. Ren, R li, B. Chen, *et al*., “EEG-Based Driving Fatigue Detection Using a Two-Level Learning Hierarchy Radial Basis Function,” *Frontiers in Neurorobotics*, vol. 15, p. 618408, 2021. DOI: https://doi.org/10.3389/fnbot.2021.618408

[17] Jian Cui, Z. Lan, O. Sourina, W-M. Wittig, “EEG-based Cross-Subject Driver Drowsiness Recognition with an Interpretable Convolutional Neural Network,” *arXiv*, May 2021. https://doi.org/10.48550/arXiv.2107.09507

[18] H. Zeng, X Li *et al*., “An EEG-Based Transfer Learning Method for Cross-Subject Fatigue Mental State Prediction,” *Sensors*, 2021. **https://doi.org/10.3390/s21072369**

[19] C. Chen *et al*., “Self-Attentive Channel-Connectivity Capsule Network for EEG-Based Driving Fatigue Detection” *IEEE Trans. Neural Networks and Learning Systems*, vol. 34, no. 7, pp. 3152-3162, 2023. DOI: 10.1109/TNSRE.2023.3299156

[20] Le. He, Li Zhang, Q. Sun, and XT. Lin, “A generative adaptive convolutional neural network with attention mechanism for driver fatigue detection with class-imbalanced and insufficient data,” *Behavioural Brain Research*, vol. 464, Apr. 2024. DOI:. https://doi.org/10.1016/j.bbr.2024.114898

[21] W. Zheng, B. Liu, Y. Lu, F. Lu, J. Ma, and B. Hu, “SEED-VIG: A large scale driver vigilance dataset with EEG, EOG, and driving performance data,” *Scientific Data*, vol. 8, no. 1, p. 255, Jul. 2021. DOI: [https://bcmi.sjtu.edu.cn/seed/seed-vig.html]

[22] S. Cao, P. Fe, W. kang “Optimized driver fatigue detection method using multimodal neural networks**.** ArticleID 12240 (2025), 2025. Doi: 10.1038/s41598-025-86709-1

[23] M. A. Al Imran, F Nazirzadeh “Designing a practical fatigue detection system: A review on recent developments and challenges,” *Journal of Neuroscience Methods*, *Journal of Safety Research*, vol. 90, Sep. 2024. DOI: https://doi.org/10.1016/j.jsr.2024.05.015

[24] E. Perkins, C. Sitaula, M. Burke, and F. Marzbanrad, “Challenges of driver drowsiness prediction: The remaining steps to implementation,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 13191338, Feb. 2023, doi: 10.1109/TIV.2022.3224690

[25] R. Hooda, V. Joshi, and M. Shah, “A comprehensive review of approaches to detect fatigue using machine learning techniques,” *Chronic Diseases and Translational Medicine*, vol. 8, Feb. 2022, doi: 10.1016/j.cdtm.2021.07.002

[26] S. Saleem, “Risk assessment of road traffic accidents related to sleepiness during driving: A systematic review,” *East Mediterr. Health J.*, vol. 28, no. 9, pp. 695700, Sep. 2022, doi: https://doi.org/10.26719/emhj.22.055

[27] R&B LLP, “How driver fatigue can lead to devastating accidents,” *R&B LLP Blog*, Apr. 2024. [Online]. Available at: https://www.randbllp.com/blog/2024/april/how-driver-fatigue-can-lead-to-devastating-accid/

[28] A. Sahayadhas, K. Sundaraj and M. Murugappan, “Detecting driver drowsiness based on sensors: A review,” *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012, doi: 10.3390/s121216937.

[29] M. K. Kamti and R. Iqbal, “Evolution of driver fatigue detection techniques, A review from 2007 to 2021,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2676, no. 12, jul. 2022, doi: https://doi.org/10.1177/03611981221096118

[30] K. S. N. Chan, “Detection of driver drowsiness from EEG signals using wearable brain sensing headband,” *Journal of Research and Applications in Mechanical Engineering*, vol. 9, no. 2, pp. 5362, May 2021. [Online]. Available: https://ph01.tci-thaijo.org/index.php/jrame/article/view/243597

[31] D.F. Dinges, M. M. Mallis, G. Maislin, and J. W. Powell, *Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Interventions*. Washington, DC, USA: U.S. Dept. of Transportation, HS-808-762, 1998 Available at: Link

[32] V. Triyanti and H. Iridiastadi, “Challenges in detecting drowsiness based on driver’s behavior,” in *Proc. 10th Int. Seminar on Industrial Engineering and Management (ISIEM)*, Tanjung Pandan, Indonesia, Nov. 2017, *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 277, p. 012042, doi: 10.1088/1757-899X/277/1/012042

[33] H. Wang, L. Xu, A. Bezerianos, Z. Zhang, and W. L. Zheng, “Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 112, Dec. 2020, doi: 10.1109/TIM.2020.3047502.

[34] I. Stancin, N. Frid, M. Cifrek, and A. Jovic, “EEG signal multichannel frequency-domain ratio indices for drowsiness detection based on multicriteria optimization,” *Sensors*, vol. 21, no. 20, p. 6932, Oct. 2021, doi: **https://doi.org/10.3390/s21206932**.

[35] T. L. T. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, “Automated drowsiness detection through wavelet packet analysis of a single EEG channel,” *Expert Systems with Applications*, vol. 55, pp. 559565, Aug. 2016, doi: https://doi.org/10.1016/j.eswa.2016.02.041.

[36] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, “Using EEG spectral components to assess algorithms for detecting fatigue,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 23522359, Mar. 2009, doi: https://doi.org/10.1016/j.eswa.2007.12.043.

[37] M. S. Ibrahim, S. R. Kamat, S. Shamsuddin, M. H. Md Isa, and M. Ito, “Electroencephalogram (EEG)-based systems to monitor driver fatigue: A review,” *Int. J. Nanoelectronics and Materials*, vol. 15, special issue, Mar. 2022. Available at Link

[38] Y. Albadawi, M. Takruri, and M. Awad, “A review of recent developments in driver drowsiness detection systems,” *Sensors*, vol. 22, no. 5, Mar. 2022, doi: 10.3390/s22052069

[39] M. Dua, Shakshi, R. Singla, S. Raj, and A. Jangra, “Deep CNN models-based ensemble approach to driver drowsiness detection,” *Neural Computing and Applications*, vol. 33, pp. 31553168, Jul. 2021. Doi: https://doi.org/10.1007/s00521-020-05209

[40] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, “Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness,” *Neuroscience & Biobehavioral Reviews*, vol. 44, Nov. 2014, doi: https://doi.org/10.1016/j.neubiorev.2012.10.003

[41] McGill Physiology Lab, “Biological signals acquisition: Electrooculography (EOG),”[Online].Available: https://www.medicine.mcgill.ca/physio/vlab/Other_exps/EOG/eogintro_n.htm

[42] S. T. Lin, Y. Y. Tan, P. Y. Chua, L. K. Tey, and C. H. Ang, “PERCLOS threshold for drowsiness detection during real driving,” *Journal of Vision*, vol. 12, no. 9, p. 1035, Aug. 2012, Vision Sciences Society Annual Meeting Abstract, doi: https://doi.org/10.1167/12.9.546.

[43] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, “Driver drowsiness detection via a hierarchical temporal deep belief network,” in *Asian Conference on Computer Vision Workshops (ACCV 2016)*, Lecture Notes in Computer Science, vol. 10118, Springer, Cham, Mar. 2017, doi: 10.1007/978-3-319-54526-4_9

[44] A. Rosebrock, “Eye blink detection with OpenCV, Python, and dlib,” *PyImageSearch Blog*, Apr. 24, 2017. [Online]. Available: https://pyimagesearch.com/2017/04/24/eye-blink-detection-opencv-python-dlib/

[45] Y. Saito, M. Itoh, and T. Inagaki, “Driver assistance system with a dual control scheme: Effectiveness of identifying driver drowsiness and preventing lane departure accidents,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 5, pp. 660671, Oct. 2016, doi: 10.1109/THMS.2016.2549032

[46] N. Lin and Y. Zuo, “Advancing driver fatigue detection in diverse lighting conditions for assisted driving vehicles with enhanced facial recognition technologies,” *PLOS One*, vol. 19, no. 7, Jul. 2024. doi- 10.1371/journal.pone.0304669

[47] National Sleep Foundation, “Drowsy driving,” *Sleep Foundation*, Aug. 21, 2025. [Online]. Available: https://www.sleepfoundation.org/drowsy-driving

[48] M. Kołodziej, P. Tarnowski, D. Sawicki, A. Pluta, R. Rak, R. J. Rak, and M. Król, “Fatigue detection caused by office work with the use of EOG signal,” *IEEE Sensors*

Journal, vol. 20, no. 24, pp. 1503215043, Dec. 2020, doi: 10.1109/JSEN.2020.3012404

[49] W.-J. Chang, L.-B. Chen, and Y.-Z. Chiou, "Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 4, pp. 461469, oct. 2018, doi: 10.1109/TCE.2018.2872162.

[50] K. Fujiwara, H. Iwamoto, K. Hori, and M. Kano, "Driver drowsiness detection using R-R interval of electrocardiogram and self-attention autoencoder," *IEEE Transactions on Intelligent Vehicles*, early access, pp. 110, Jan. 2023, doi: 10.1109/TIV.2023.3308575

[51] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *International Journal of Biosensors & Bioelectronics*, vol. 4, no. 2, pp. 195202, 2018, available at- Link

[52] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, p. 258, Sep. 2017, doi: https://doi.org/10.3389/fpubh.2017.00258.

[53] A. T. Satti, J. Kim, E. Yi, H.-Y. Cho, and S. Cho, "Microneedle array electrode-based wearable EMG system for detection of driver drowsiness through steering wheel grip," *Sensors*, vol. 21, no. 15, p. 5091, Aug. 2021, doi: 10.3390/s21155091.

[54] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 1693716953, Dec. 2012, doi: **https://doi.org/10.3390/s121216937**

[55] D. Morris, "The cold driver: Driving performance under thermal stress," M.S. thesis, Dept. of Psychology, Clemson Univ., Clemson, SC, USA, 2015. [Online]. Available: https://tigerprints.clemson.edu/all_theses/2236

[56] M. Kołodziej, P. Tarnowski, A. Majkowski, and R. J. Rak, “Electrodermal activity measurements for detection of emotional arousal,” *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 67, no. 4, pp. 813826, 2019, doi: **10.24425/bpasts.2019.130190**

[57] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, “YawDD: A yawning detection dataset,” in *Proc. 5th ACM Multimedia Systems Conf. (MMSys ’14)*, Singapore, Mar. 2014,, doi: https://doi.org/10.1145/2557642.256367

[58] A. Celecia, K. Figueiredo, M. Vellasco, and R. González, “A portable fuzzy driver drowsiness estimation system,” *Sensors*, vol. 20, no. 15, p. 4093, Jul. 2020, doi: **https://doi.org/10.3390/s20154093**

[59] W. Liu, J. Qian, Z. Yao, X. Jiao, and J. Pan, “Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection,” *Future Internet*, vol. 11, no. 5, p. 115, May 2019, doi: **https://doi.org/10.3390/fi11050115**

[60] A. Ghourabi, H. Ghazouani, and W. Barhoumi, “Driver drowsiness detection based on joint monitoring of yawning, blinking and nodding,” in *Proc. IEEE 16th Int. Conf. on Intelligent Computer Communication and Processing (ICCP)*, Cluj-Napoca, Romania, Sep. 2020, pp. 143148, doi: 10.1109/ICCP51029.2020.9266160.

[61] C. J. de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, “Detection and prediction of driver drowsiness using artificial neural network models,” *Accident Analysis & Prevention*, vol. 126, pp. 95104, Apr. 2019, doi: https://doi.org/10.1016/j.aap.2017.11.038..

[62] S. Ebrahimian, H. Kiashari, A. Nahvi, H. Bakhoda, A. Homayounfard, and M. Tashakori, “Evaluation of driver drowsiness using respiration analysis by thermal imaging on a driving simulator,” *Multimedia Tools and Applications*, vol. 79, , jul. 2020.

[63] M. Knapik and B. Cyganek, “Driver’s fatigue recognition based on yawn detection in thermal images,” *Neurocomputing*, vol. 338, Apr 2019, doi: https://doi.org/10.1016/j.neucom.2019.02.014.

[64] C. Zhang, X. Wu, X. Zheng, and S. Yu, “Driver drowsiness detection using multi-channel second order blind identifications,” *IEEE Access*, vol. 7, pp. 1182911843, Jan. 2019, doi: 10.1109/ACCESS.2019.2891971.

[65] J. Min, M. Cai, C. Gou, C. Xiong, and X. Yao, “Fusion of forehead EEG with machine vision for real-time fatigue detection in an automatic processing pipeline,” *Neural Computing and Applications*, vol. 35, pp. 88598872, Jun. 2023.

Doi: https://doi.org/10.1007/s00521-022-07466-0

[66] W.-L. Zheng and B.-L. Lu, “A multimodal approach to estimating vigilance using EEG and forehead EOG,” *Journal of Neural Engineering*, vol. 14, no. 2, p. 026017, Apr. 2017, doi: 10.1088/1741-2552/aa5a98

[67] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, “Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, , Sept. 2010, doi: 10.1109/TBME.2010.2077291.

[68] H.-S. Choi, S. Min, S. Kim, H. Bae, J.-E. Yoon, and I. Hwang, “Learning-based instantaneous drowsiness detection using wired and wireless electroencephalography,” *IEEE Access*, vol. 7, Oct. 2019, doi: 10.1109/ACCESS.2019.2946053.

[69] S. Barua, M. U. Ahmed, C. Ahlström, and S. Begum, “Automatic driver sleepiness detection using EEG, EOG and contextual information,” *Expert Systems with Applications*, vol. 115, Jan. 2019, doi: https://doi.org/10.1016/j.eswa.2018.07.054

[70] Q. Abbas and A. Alsheddy, “Driver fatigue detection systems using multi-sensors, smartphone, and cloud-based computing platforms: A comparative analysis,” *Sensors*, vol. 21, no. 1, Dec. 2020, doi: **https://doi.org/10.3390/s21010056**

[71] M. Doudou, A. Bouabdallah, and V. Berge-Cherfaoui, “Driver drowsiness measurement technologies: Current research, market solutions, and

challenges," *International Journal of Intelligent Transportation Systems Research*, vol. 18, pp. 297319, Sep. 2019.

[72] B. L. Lu, W. L. Zheng, and F. C. Li, "SEED: SJTU Emotion EEG Dataset," BrainComputer Interface Research Center (BCMI), Shanghai Jiao Tong Univ. [Online]. Available: https://bcmi.sjtu.edu.cn/home/seed.

[73] Q. Massoz, T. Langohr, C. François, J. Verly, and M. Van Droogenbroeck, "The ULg multimodality drowsiness database (called DROZY) and examples of use," in *Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 17, doi: **10.1109/WACV.2016.7477715**.

[74] E. Kavalci, "SUST Driver Drowsiness Detection (SUST-DDD) Dataset," *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/esrakavalci/sust-ddd/data.

[75] National Tsing Hua University, "Driver Drowsiness Dataset (DDD)," *CVLAB, National Tsing Hua Univ.*, Hsinchu, Taiwan. [Online]. Available: http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/

[76] C. S. Silveira, J. S. Cardoso, A. L. Lourenço, and C. Ahlström, "Importance of subject-dependent classification and imbalanced distributions in driver sleepiness detection in realistic conditions," *IET Intelligent Transport Systems*, vol. 13, no.11,Nov.2019, doi: **https://doi.org/10.1049/iet-its.2018.5284**

[77] J. R. Paulo, G. Pires, and U. J. Nunes, "Cross-subject zero calibration driver's drowsiness detection: Exploring spatiotemporal image encoding of EEG signals for convolutional neural network classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, May 2021, doi: 10.1109/TNSRE.2021.3079505.

[78] X. Feng, Z. Guo, and S. Kwong, "ID3RSNet: Cross-subject driver drowsiness detection from raw single-channel EEG with an interpretable residual shrinkage

network," *Frontiers in Neuroscience*, vol. 18,, Jan. 2025, doi: 10.3389/fnins.2024.1508747

[79] T. Fonseca and S. Ferreira, "Drowsiness detection in drivers: A systematic review of deep learning-based models," *Applied Sciences*, vol. 15, no. 16, , Aug. 2025, doi: **https://doi.org/10.3390/app15169018**

[80] Y. Jiao, Y. Deng, Y. Luo, and B.-L. Lu, "Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks," *Neurocomputing*, vol. 408, Sep. 2020, doi: https://doi.org/10.1016/j.neucom.2019.05.108.

[81] B. K. Savaş and Y. Becerikli, "Behavior-based driver fatigue detection system with deep belief network," *Neural Computing and Applications*, vol. 34, , Apr. 2022. Doi-https://doi.org/10.1007/s00521-022-07141-

[82] R. S. Abdulsadig and E. Rodriguez-Villegas, "A comparative study in class imbalance mitigation when working with physiological signals," *Frontiers in Digital Health*, vol. 6, Mar. 2024, doi: https://doi.org/10.3389/fdgth.2024.1377165.

[83] A. Picot, A. Caplier, and S. Charbonnier, "Comparison between EOG and high frame rate camera for drowsiness detection," in *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, Snowbird, UT, USA, Dec, 2009, doi: 10.1109/WACV.2009.5403120.

[84] H.-T. Nguyen, N.-D. Mai, B. G. Lee, and W.-Y. Chung, "Behind-the-ear EEG-based wearable driver drowsiness detection system using embedded tiny neural networks," *IEEE Sensors Journal*, vol. 23, no. 19, pp., Oct. 2023. Doi-10.1109/JSEN.2023.3307766

[85] J. Cui, Z. Lan, Y. Liu, R. Li, F. Li, O. Sourina, and W. Müller-Wittig, "A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG," *Methods*, vol. 202, pp, Jun. 2022, doi: https://doi.org/10.1016/j.ymeth.2021.04.017.

[86] D. Salem and M. Waleed, “Drowsiness detection in real-time via convolutional neural networks and transfer learning,” *Journal of Engineering and Applied Science*, vol. 71, no. 1, p. 122, May 2024. Doi- https://doi.org/10.1186/s44147-024-00457-z

[87] X.-Q. Huo, W.-L. Zheng, and B.-L. Lu, “Driving fatigue detection with fusion of EEG and forehead EOG,” in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, Vancouver, BC, Canada, Jul. 2016, doi: 10.1109/IJCNN.2016.7727294

[88] E. M. Öztürk, A. Küçükmanisa, and O. Urhan, “Drowsiness detection system based on machine learning using eye state,” *Balkan Journal of Electrical and Computer Engineering*, vol. 10, no. 3,, Jul. 2022. Doi- 10.17694/bajece.1028110

[89] A. S. Abdulbaki and S. A. Khadim, “Real-time safety automobile driver system,” *International Journal of Computer Applications*, vol. 130, no. 17, Nov. 2015. Doi- 10.5120/ijca2015906803

[90] W. Fang, L. Tang, and J. Pan, “AGL-Net: An efficient neural network for EEG-based driver fatigue detection,” *Journal of Integrative Neuroscience*, vol. 22, no. 6, Oct. 2023, doi: https://doi.org/10.31083/j.jin2206146.

[91] W.-L. Zheng and B.-L. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,” *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[92] Q. Ji, Z. Zhu, and P. Lan, “Real-time nonintrusive monitoring and prediction of driver fatigue,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, Jul. 2004, doi: 10.1109/TVT.2004.830974.

[93] I.-K. Yeo and R. A. Johnson, “A new family of power transformations to improve normality or symmetry,” *Biometrika*, vol. 87, no. 4, pp. 954959, Dec. 2000, doi: https://doi.org/10.1093/biomet/87.4.954.

[94] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 3rd ed., Wiley Series in Probability and Statistics. Hoboken, NJ, USA: Wiley, 2019, doi: 10.1002/9781119482260

[95] H. J. Eoh, M. K. Chung, and S.-H. Kim, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation," *Int. J. Ind. Ergon.*, vol. 35, no. 4, Apr. 2005, doi: https://doi.org/10.1016/j.ergon.2004.09.00.

[96] D. Aeschbach, J. R. Matthews, T. T. Postolache, M. A. Jackson, H. A. Giesen, and T. A. Wehr, "Dynamics of the human EEG during prolonged wakefulness: Evidence for frequency-specific circadian and homeostatic influences," *Neuroscience Letters*, vol. 239, no. 23, Dec. 1997, doi: 10.1016/s0304-3940(97)00904-x.

[97] K. Pearson, "Notes on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58,, 1895, doi: https://doi.org/10.1098/rspl.1895.0041

[98] Y. Ma, B. Chen, R. Li, C. Wang, J. Wang, Q. She, Z. Luo, and Y. Zhang, "Driving fatigue detection from EEG using a modified PCANet method," *Computational Intelligence and Neuroscience*, vol. 2019, Jul. 2019, doi: 10.1155/2019/4721863.

[99] S.-W. Chuang, L.-W. Ko, Y.-P. Lin, R.-S. Huang, T.-P. Jung, and C.-T. Lin, "Co-modulatory spectral changes in independent brain processes are correlated with task performance," *NeuroImage*, vol. 62, no. 3, Sep. 2012, doi-10.1016/j.neuroimage.2012.05.035.

[100] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 91, Feb. 2006, doi: 10.1186/1471-2105-7-91.

[101] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, 2001, doi: 10.1023/A:1010933404324.

[102] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785794, doi: https://doi.org/10.1145/2939672.293978.

[103] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3,, 1995, doi: 10.1007/BF00994018.

[104] A. Craik, Y. He, and J. L. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: A review,” *J. Neural Eng.*, vol. 16, no. 3, Apr. 2019, doi: 10.1088/1741-2552/ab0ab5

[105] X. Cao, Y. He, S. Zhou, and J. Zhang, “A hybrid CNN-BiLSTM model with feature fusion for accurate epilepsy seizure detection,” *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, Art. 6, Jan. 2025, doi: 10.1186/s12911-024-02845-0.

[106] D. Gao, X. Tang, M. Wan, G. Huang, and Y. Zhang, “EEG driving fatigue detection based on log-Mel spectrogram and convolutional recurrent neural networks,” *Frontiers in Neuroscience*, vol. 17, Mar. 2023, doi: https://doi.org/10.3389/fnins.2023.1136609

[107] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proc. 14th Int. Joint Conf. Artificial Intelligence (IJCAI)*, vol. 2, 2001, Available at- Link

[108] R. Chai, S. H. Ling, P. P. San, G. R. Naik, T. N. Nguyen, Y. Tran, A. Craig, and H. T. Nguyen, “Improving EEG-based driver fatigue classification using sparse-deep belief networks,” *Frontiers in Neuroscience*, vol. 11, p. 103, Mar. 2017, doi: https://doi.org/10.3389/fnins.2017.00103.

[109] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9,, Sep. 2009, doi: 10.1109/TKDE.2008.239

[110]. A. Fernández, S. García, F. Herrera, and N. V. Chawla, *Learning from Imbalanced Data Sets*. Cham, Switzerland: Springer, 2018. Available at- Link.

[111] I. P. A. E. D. Udayana, M. Sudarma, I. K. G. D. Putra, I. M. Sukarsa, and M. Jo, “Comparative analysis of denoising techniques for optimizing EEG signal processing,” *Lontar Komputer*, vol. 15, no. 2, , Aug. 2024, doi: 0.24843/LKJITI.2024.v15.i02.p05.

[112] S. Chaabene, B. Bouaziz, A. Boudaya, A. Hökelmann, A. Ammar, and L. Chaari, “Convolutional neural network for drowsiness detection using EEG signals,” *Sensors*, vol. 21, no. 5, Mar. 2021, doi: **https://doi.org/10.3390/s21051734**

Appendix A: Supplementary Materials I

Ethics SOP Form- Link