Driver drowsiness detection, 2024, 1–15
doi:
CSE dept, NSUT Dwarka
Original Article

An Approach to Driver Fatigue Detection Using CNN-LSTM Fusion Model

Dr. Poonam Rani (Associate Professor), Divanshi Jain, Sanskriti and Maithri Sri Meda

Department of Computer Science, Netaji Subhas University of Technology, Dwarka, 110078, New Delhi

Abstract

The Intelligent Transportation Systems (ITS) offer a transformative approach to improving road safety by incorporating advanced technologies that can monitor driver behavior in real-time. Road accidents resulting from driver fatigue are one of the major global concerns that call for innovation in safety measures. Through sensors, cameras, and machine learning models, ITS enable proactive detection of fatigue, reducing accidents caused by drowsiness significantly. This paper is an attempt to solve the problem of fatigue detection under the ITS framework using a hybrid approach of classical computer vision and deep learning techniques. The system uses Haar Cascade for the efficient detection of faces and eyes and further uses Convolutional Neural Networks (CNN) to extract spatial features and LSTM networks to capture temporal patterns. The proposed layered layered architecture ensures precise classification of driver states while maintaining computational efficiency for real-time applications. The proposed model was impressive, with an accuracy of 95.1%, and had good precision, recall, F1-score, sensitivity, and specificity metrics. Generalization was further enhanced through training optimizations, including data augmentation and adaptive learning rate adjustments. The integrated system in ITS gives timely visual and auditory alerts to drivers, thereby enhancing road safety and the realization of the ideals of modern transportation systems.

Key words: Convolution Neural Network, Driver Drowsiness Detection System, Dual-Modality Monitoring, Electrocardiogram, Facial Landmark Detection, Haar Cascade, Heart Rate Variability, Intelligent Transportation System, Long Short Term memory, Physiological Signals, Real-time Monitoring.

1. Introduction

An Intelligent Transportation System (ITS) is the application of advanced technologies, such as information and communication systems, for improving the efficiency, safety, and sustainability of transportation networks. It integrates technology with infrastructure and vehicles to improve the movement of people and goods, reduce traffic congestion, and minimize environmental impact. The basic concept is that these incorporate advanced wireless, electronic, and automated solutions aimed at further improving safety, efficiency, and convenience in surface transportation [1]. ITS combines high technology and improvements in information systems, communication, sensors, controllers and advanced mathematical methods with the conventional world of transportation infrastructure. The US National Highway Traffic Safety Administration estimated around 100,000 drowsiness-related road accidents in worldwide statistics per year. Consequent upon this drowsiness-related accident, 1,500 people die, while 70,000 are injured according to WHO study 2022 (study by WHO in 2022). According to the latest statistics reported in [2] and [3] human errors are responsible for almost all of the accidents caused. The intuition behind the Driver Drowsiness Detection System(DDD) lies in combining physiological signals [4], [5] - [17], [20, 21] and respiratory patterns [18] act as early indicators of fatigue and visual cues like blink rate, eye closure duration, and head positioning to create a robust, real-time monitoring mechanism for alertness detection. By integrating these modalities, the system provides a comprehensive approach to detect fatigue [19]. Drowsiness is the state of drowsiness, usually arising in inappropriate contexts [22]. Even though drowsiness lasts for a few minutes, the effects may be catastrophic. The state is usually the result of fatigue that depresses the levels of attention and alertness [23, 27, 28]. Drowsiness can arise when driving for long distances without enough sleep or at times of the day when one is usually sleeping [24]. In such situations, the main problem is the driver's distraction, which causes a delay in reaction to events on the road [25]. Driver fatigue research is divided into two main areas: the identification of factors that influence fatigue and the detection of fatigue, which is often done using cluster analysis with relevant keywords. [26]. Physiological signals like electroencephalogram (EEG), electromyography (EMG),

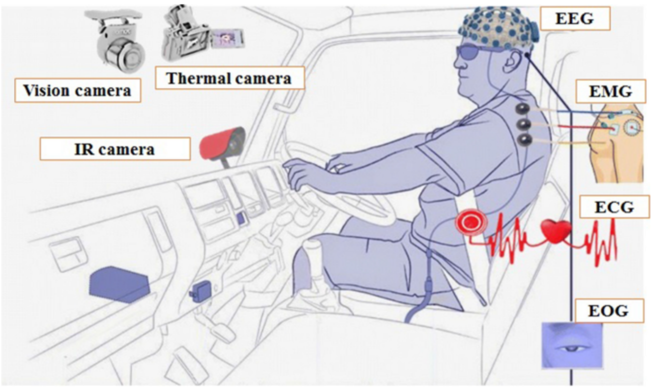

Fig. 1: Collecting physiological measures from a driver [81]

electrocardiogram (ECG), electrooculogram (EOG), etc. offer insights into subtle changes that precede visual symptoms, while facial analysis captures observable signs of drowsiness through vision camera, thermal camera and IR camera (as shown in Fig. 1) [30, 31]. The system prioritizes early intervention, issuing timely alerts to prevent accidents, ultimately aligning with the broader goal of advancing road safety through intelligent, adaptive technologies. Various measures for drowsiness detection are shown in Fig. 2 [30]. The vital components are listed below:

1.1 Physiological Signals

Physiological measures such as heart rate, ECG, and skin conductivity are direct indicators of drowsiness. Heart rate variability (HRV) decreases as fatigue sets in, signifying reduced alertness. Studies show that drowsy drivers exhibit lowered HRV, irregular breathing patterns, and sometimes arrhythmic ECG signals, which collectively reflect diminished alertness levels [33].

1.2 Respiratory and Cardiovascular Patterns

Changes in breathing rate and depth, coupled with fluctuations in heart rate, can signal fatigue [34]. A consistent decrease in respiratory rate often correlates with drowsiness. Using sensors to monitor ECG signals also allows the detection of reduced autonomic responses—further supporting drowsiness detection. Research shows that fatigue-related ECG patterns include low-frequency, high-amplitude waves that diverge from alertness indicators [35].

1.3 Detection Framework

Dual-Modality Monitoring: Input → Face Detection → Physiological Signal Collection → Feature Analysis → Alert Generation (shown in Fig. 8) [36]. The system utilises both video and sensor input to capture facial expressions and physiological signals, respectively. Video frames are used to track blink rates and eye closures, while ECG and heart rate sensors detect drowsiness through physiological markers. If measurements exceed fatigue thresholds, an alert system activates, advising the driver to take preventive measures.

The process of escalating the alert within the DDD System is designed to be tiered as it correlates with the degree of fatigue exhibited by the driver's state of drowsiness [37, 38]. Such a strategy takes into consideration both the physiological and the visual states of the driver. In the basic framework, the trigger conditions for an alert to be issued are also clearly or at least partially defined.

The system of Detection of Driver Drowsiness takes into account environmental changes in order to operate consistently on different scales of driving activities. During the day, it uses standard vision based algorithms and physiologics such as ECG and HRV to measure fatigue. In a night time or dark scenario, infrared scanning and low light enhanced vision systems are used for face detection, ensuring no disturbance to the driver's attention while doing so. This modification avoids cases of variability due to lighting changes. Also, it includes effects such as glares, weather, and vibration disturbances using strong methods that suppress such noise. These adaptive features enhance the System's usability and reliability in different environments [32].

The advanced algorithms are able to compensate for variations in the driver such as the shape of the face, the orientation of the head, or even the physiological condition of the driver at that time. Outside disturbances like vibration of the vehicle, light changes, or visual noise from outside weather affect the results [40]. The most effective method of drowsiness detection is thermal imaging since it is non-invasive and insensitive to changes in ambient light. The process of detection includes measuring the temperature over the forehead, especially in the area above the supratrochlear artery, as well as that of the cheeks.[39]. This robustness ensures users and driving contexts are diverse yet accuracy rates do not vary significantly increasing reliability especially in safety critical situations.

In summary, DDD System has many features as it can be used in a lot of sectors which increases the safety of the roads and decreases all the casualties caused by tiredness. In the case of ordinary vehicles, it can be fixed on dashboards or advanced driver assistance systems that can be used to provide alerts to the people when they are traveling for extended periods. In the case of commercial fleets, it supports the real-time checking of the state of fatigue of the driver consistently to improve logistics safety and protect a safe working environment [23, 41].

Also, in the area of public transport, the system may be useful in making buses and trains safer where driver distances need to be controlled. Also, it is possible to extend its use to taxis and delivery riders in order to enhance the safety of both the drivers and the riders.

This paper is divided into several key sections to provide a comprehensive analysis of DDD. It starts with an section 1 that is, introduction that outlines proposed paper after which it includes section 2 namely, motivation behind the study and its relevance in enhancing road safety. The section 3, literature review, places the research in perspective with existing studies. Following this, it goes into section 4, discussing methodology detailing the dual-modality approach combining physiological and visual monitoring. Following this, in section 5, technical implementation is discussed in sections on Data Processing, Model Training and Architecture. In addition, results and evaluation metrics are used to validate performance. Section 6 compares the proposed methodology with existing models. Finally, in section 7, the conclusion outlines findings, and acknowledgments highlight contributions. Figures

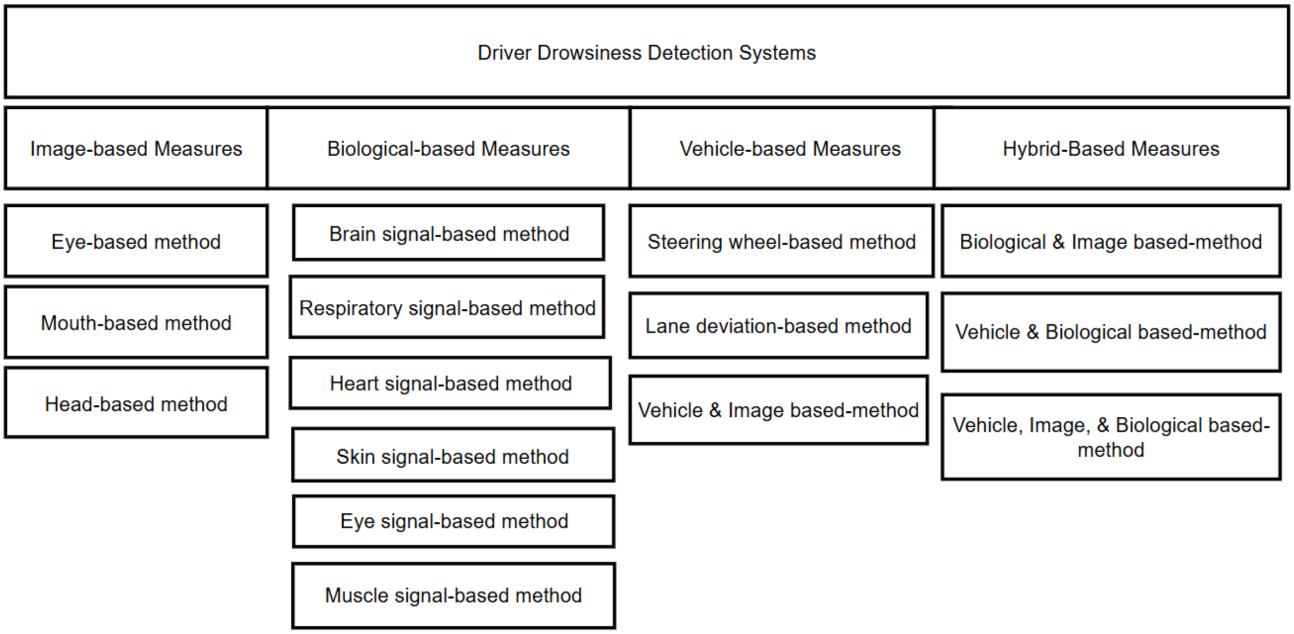

Fig. 2: DDD measures

and flowcharts are used at intervals throughout the paper for clarification of concepts.

2. Motivation

Driver drowsiness is one of the world's major causes for accidents on roads, further calling for the urgent need to use effective real-time monitoring solutions. Most drowsiness detectors had traditionally used facial-tracking techniques to detect visible signs that suggest fatigue, such as drooping eyelids or slow blinks. These may have limitations since they overlook deep physiological indicators of a driver's real degree of alertness. Physiological measures, including HRV, ECG patterns, and respiration rate, prove especially valuable here because they detect the internal signs of deterioration that develop before there is any manifestation in outward appearance. They capture the signs that are extremely important in one's ability to prevent a crash by alerting the driver before is an actual occurrence of hazardous situation on the road. Utilizing recent developments in biosensors and machine learning, a dual-modality approach combining physiological monitoring with facial tracking is now both feasible and highly accurate in real-world settings. This hybrid system can adapt continuously to each driver, interpreting subtle changes across multiple modalities in order to deliver highly reliable fatigue assessments. Coupling this advanced monitoring with Advanced Driver Assistance Systems (ADAS) not only raises overall road safety but also responds to the goals of the automotive industry to better protect drivers with cost-effective and scalable technology. Implementing these all-inclusive monitoring solutions will enable the automobile manufacturer to better keep pace with rising standards in safety and, significantly, help to avoid drowsiness-related accidents, thus making roads safer and more proactive to drivers' well-being.

3. Literature Review

This section reviews the literature that provides a comprehensive examination of various methodologies, models, and performance metrics of DDD systems. The primary goal of these systems is to improve road safety by detecting fatigued drivers before an accident occurs. As shown in Table 1, a wide array of techniques has been explored, including computer vision and deep learning(DL) approaches, each offering its own strengths and limitations.

One of the most common approaches in DDD is computer vision, which focuses on tracking facial features and eye movements to detect signs of fatigue. For example, a study [59] utilized an ensemble approach combining multiple deep CNN, including AlexNet, VGG-FaceNet, FlowImageNet, and ResNet, to detect drowsiness from RGB video. The system achieved 85% accuracy, with sensitivity, specificity, and precision scores indicating solid performance. However, it faced challenges such as high dependence on environmental conditions and difficulty in detecting all types of drowsiness signals, particularly in real-world variations in lighting, head position, and facial expressions. Similarly, another study [84] implemented a histogram of oriented gradient (HOG) feature extraction method and Naïve Bayes classification to analyze facial features, reaching an accuracy of 85.62%. This approach, though promising, struggled with unreliable eye tracking and class imbalances in the dataset, which negatively impacted performance and made the model less effective in real-world conditions.

In the field of DL, CNNs have become a popular choice for analyzing images of a driver's face and eyes. A study [65] employed CNNs on the Closed Eye in the Wild (CEW) and YawDD datasets, achieving an impressive accuracy of 96%. However, the system's reliance on eye closure detection limits its applicability in real-time environments, especially with varying lighting conditions, camera angles, and face orientations. This can result in missed or false detections in real-world scenarios where eye movements are subtle and may not be captured clearly. Similarly, a CNN-based approach utilizing the VGG16 model [69] reported 97% accuracy, but it faced challenges with low performance in certain lighting conditions and its inability to capture temporal dependencies such as prolonged eye closures, as the model lacked LSTM integration to address such variations.

A different study [72] compared various machine learning and DL techniques for drowsiness detection, highlighting that while some techniques, like facial and biological signal analysis, can perform well in controlled environments, they are highly sensitive to variations in real-world conditions such as lighting, facial orientation, and environmental factors. Furthermore, a hybrid approach combining CNN with LSTM models for enhanced temporal feature extraction [83] achieved a high F1-score of 92.68% using self-recorded video data. Despite its success, the system's reliance on specific datasets limits its generalizability across different driver demographics and real-world environments, and it may not be applicable to diverse driving conditions or populations.

Another promising approach in DL integrates CNN with physiological data such as HRV or ECG signals. For example, a study [76] proposed a spatio-temporal fusion network with brain region partitioning strategy for EEG-based fatigue detection, achieving an accuracy of 92.43%. This method demonstrated the power of integrating physiological data, but it also struggled with issues like neglecting cross-subject classification and intra-subset node interactions, which are essential for generalizing the model to different drivers. Furthermore, while this approach can accurately detect drowsiness, its dependency on EEG signals, which require specialized equipment and careful calibration, limits its practicality for widespread real-world use. Another hybrid approach [80] combined respiratory signals, vehicle lateral position, and reaction time data, using classifiers like SVM, DT, and LSTM to detect fatigue. This system reached 88% accuracy, but its real-world applicability was limited by the reliance on specific data sources such as driving simulators and respiratory signals, and it was based on a small sample size of only 25 participants, which reduces its generalizability.

Multimodal approaches that combine computer vision with physiological data are gaining more attention as a way to improve accuracy and robustness. For instance, the non-invasive radar-based system described in [60] used impulse radio ultra-wideband (IR-UWB) radar to detect respiration rate, achieving an accuracy of 87%. However, this system faced limitations in noisy environments and when subjects were positioned far from the radar, which impacted its reliability and accuracy under less controlled conditions. Another multimodal approach explored by [78] integrated CNNs with facial landmark detection to monitor eye closure, yawning, and head tilting, achieving up to 98% accuracy using real-time video data. Despite its success, the system remained sensitive to real-time environmental variations, such as changing lighting conditions, and may struggle with different driver demographics, making it difficult to deploy in diverse, uncontrolled driving conditions.

Each of these methods comes with its own set of challenges. Computer vision and DL techniques, while often effective in controlled settings, remain sensitive to environmental factors like lighting, head movements, and camera angles. Multimodal systems, though they offer higher accuracy and robustness, are more complex and require careful calibration of multiple sensors, which can be difficult to implement in real-world systems. Additionally, approaches that rely on Recurrent Neural Networks (RNNs) to process ECG or EEG signals can be affected by noise and movement artifacts, which are common in real-world driving conditions [8]. Despite these challenges, ongoing advancements in DL, sensor fusion, and computational efficiency are helping to improve the reliability of these systems, bringing them closer to practical application in real-world scenarios.

Table 1: Literature Review

Ref	**Title**	**Methodology**	**Dataset**	**Performance metrics**	**Limitation**
[59]	Deep CNN models-based ensemble approach to DDD (2020)	Ensemble approach combining AlexNet, VGG-FaceNet, FlowImageNet, and ResNet	NTHU-DDD. 200 video clips of 25 subjects for training datasets and 50 of six subjects as testing datasets	Accuracy: 85% Sensitivity 0.82% Specificity 0.87% Precision 0.863% F1-Score 0.8409%	High dependence on environmental conditions
[60]	Non-Invasive DDD System (2021)	Non-invasive, non-touch impulsive radio ultra-wideband (IR-UWB)	Collected from 40 participants	Accuracy: 87% precision: 0.86% recall: 0.88% F1-score: 0.86%	High reliability on IR-UWB radar
[65]	DDD Using DL (2021)	DL-based approach to DDD using CNN.	Closed Eye in the Wild (CEW) Dataset and yawDD were used in this	Accuracy: 96%	Relies on image processing and detection of eye closure
[84]	Improved HOG Features and Naïve Bayesian Classification (2021)	Histogram of Oriented Gradient (HOG) features Driver image, eye pair region along with Naive Bayes Classifier.	publicaly available NTHU-Drowziness Detection Dataset (NTHU-DDD)	Accuracy: 85.62%	Unreliable eye tracking model, class imbalance in dataset

Ref	Title	Methodology	Dataset	Performance metrics	Limitation
[72]	ML and DL Techniques for Driver Fatigue and Drowsiness Detection (2023)	Image/video-based analysis, biological signal analysis, vehicle movement analysis, and hybrid techniques	Many of publicly available datasets like yawDD	Accuracy varies by technique and dataset, with some achieving high performance in facial and biological signal analysis.	Dataset limitations, variations in real-world conditions
[69]	A Deep-Learning Approach to DDD (2023)	CNN and pre trained VGG16.	public dataset on Kaggle with 2900 images	CNN: 97% accuracy, VGG16: 74% accuracy	Low accuracy of VGG16 model, lacks LSTM's ability to capture temporal dependencies
[80]	A Hybrid Approach for DDD Utilizing Practical Data to Improve Performance System and Applicability (2023)	Respiratory signals, vehicle lateral position, and reaction time data. Application of various classifiers like MLP, SVM, DT, and LSTM	25 participants, Questionnaires, Respiratory Signals and Driving Simulator	Accuracy: 88%, Precision: 85%, Recall: 83%, F1-Score: 84%	Limited real-world applicability due to reliance on specific data sources
[83]	A CNN-LSTM approach for accurate drowsiness and distraction detection in drivers (2024)	data organized into temporal sequences (Time Series) for training, uses Intersection over Union (IOU) technique	Self-recorded dataset from volunteers: 65 videos (1 min each) 18 male and 47 females	Accuracy: 93.60, Precision: 93.61, Recall: 91.92, F1-Score: 92.68	No haar cascade, reliance on specific datasets limit generalization to all real-world conditions.
[78]	Optimizing Road Safety: Advanced Solutions for Detecting and Mitigating Driver Drowsiness (2024)	CNN and Inception v3 models	Datasets of open and closed eye images. Eye and facial landmark datasets, Real-time webcam data for monitoring	Inception v3 Accuracy: 98%, CNN Accuracy: 96%	Real-time environmental variations, sensitivity to different driver demographics, inaccuracies in rapidly changing conditions.
[74]	DDD Using CNN and Computer Vision Techniques (2024)	Uses CNN for eye and yawn detection, and computer vision for eye, mouth, and head pose analysis.	MRL Eye Dataset (48,000 images) and YawDD Dataset (2,900 images)	Accuracy: 96.42% on YawDD, 84.53% on MRL Eye dataset.	Unreliable in moving environments, generalization to real-world scenarios.
[76]	EEG-Based Driver Fatigue Detection (2024)	STFN-BRPS model	EEG data collected during a real driving experiment involving 10 subjects	Accuracy: 92.43%, Precision: 91.52%, Recall: 92.89%, F1-score: 0.927, Average Pseudo-Online Accuracy: 85.68%	Limited to within-subject classification, neglects interaction relationships within neighboring electrode subsets,
[61]	High-Fidelity Machine Learning Techniques for DDD (2024)	Traditional classifiers (Naïve Bayes, SVM, Random Forest) and VGG16 for DDD via facial images.	435 test images (classifiers) and 2,500 images (VGG16).	RF: 92.41%, SVM: 90.34%, NB: 69.43%, VGG16: 97.20% accuracy.	Small dataset, lacks diversity

4. Methodology

The primary goal of this project is to develop a real-time drowsiness detection system using a hybrid CNN system and eye region landmarks. This system processes live video frames to detect facial landmarks and determine the eye status, whether open or closed. By leveraging pre-trained facial detection models in combination with a custom-designed CNN, the system can identify signs of drowsiness and generate alerts to warn the driver.

The first step in the methodology is video frame processing. In this phase, the system extracts and processes facial landmarks in each video frame. Facial landmarks are key points on the face, such as the eyes, nose, and mouth, that help in detecting specific facial features. These landmarks are crucial for understanding the eye state, as they allow the system to focus on the eye region and measure parameters like the eye's openness or blinking. The video frames are processed in real-time, which requires efficient processing and accurate detection of facial features.

Once the facial landmarks are detected, the system proceeds to the next phase, which is eye status detection. Using CNN-based analysis, the system determines whether the eyes are open or closed. This is a critical step because the state of the eyes is a strong indicator of drowsiness. For example, closed or partially closed eyes over an extended period suggest that the driver may be experiencing fatigue. The CNN is trained on a dataset of eye images to accurately classify the eye status. This model is designed to handle variations in lighting, head position, and other environmental factors to ensure robustness in real-world scenarios.

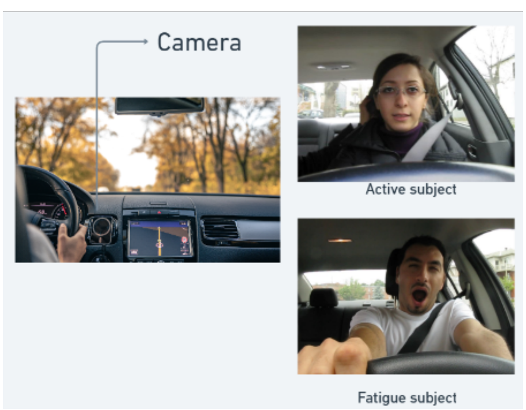

Fig. 3: Capturing real-time driver behaviour through camera

The final step in the system is alert generation. If the system detects signs of drowsiness, such as prolonged eye closure or irregular blinking patterns, it triggers an alert to warn the driver. The alert can take various forms, such as visual or audio notifications, to ensure that the driver is aware of their drowsy state. The alert mechanism is crucial for preventing accidents caused by drowsiness, as it serves as a timely reminder to the driver to take necessary actions, such as resting or stopping the vehicle.

4.1 Datasets

Two primary datasets were used to train and validate the model. The Drowsiness Prediction Dataset (downloaded from Kaggle) contains labelled images of individuals in drowsy and non-drowsy states, focusing on eye features and head position. This dataset serves as the core resource for training the CNN model, ensuring balanced class representation between drowsy and alert states. The second dataset, the Prediction Images Dataset, is a supplementary collection of images depicting individuals in both drowsy and active states. It is used to enhance the diversity of the training set, improving the model's ability to generalize across different face orientations, lighting conditions, and individual variations.

4.2 Data Loading and Preprocessing

The data preprocessing pipeline is illustrated in the flowchart (as shown in Fig. 5. The first step is data loading, where images from the Kaggle dataset are loaded using OpenCV. This method ensures efficient handling of the image data, enabling smooth processing and quick access to each image during the training phase. It is a crucial initial step to prepare the dataset for further analysis.

The next step is Image Resizing and Normalization. In this stage, all images are resized to a standard resolution of 145x145 pixels. This resizing ensures consistency in the input size, which is critical for processing the images using a CNN, as CNN models typically require inputs of a fixed size. Additionally, the pixel values are normalized, meaning they are scaled to a range that facilitates effective model training. Normalization helps to speed up the convergence of the model by ensuring that the input features have similar scales.

Following this, the facial landmark extraction step is performed. After the images are preprocessed, facial landmarks are extracted using facial detection models. This technique identifies key facial points, especially around the eyes, which are critical for determining the eye state (open or closed). Extracting these features enables the model to focus on the most relevant parts of the image for drowsiness detection, improving the overall accuracy of the system.

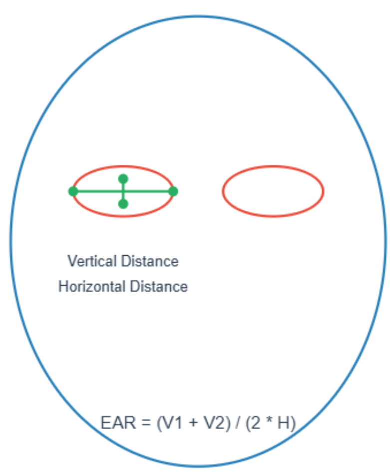

Fig. 4: EAR Measurement and Facial Landmarks

The formula for Eye Aspect Ratio (EAR) is given by the equation (1) (as shown in Fig. 4).

$$\mathrm{EAR} = \frac{|P_2 - P_4| + |P_3 - P_5|}{2\,|P_1 - P_6|} \tag{1}$$

where $P_1, P_2, P_3, P_4, P_5, P_6$ are the facial landmarks of the eye. EAR is used to detect eye blinks based on the ratio of vertical to horizontal distances.

Data augmentation is applied to enhance the robustness and generalization ability of the model. This technique involves transformation of images through rotation, shift, zoom, and flip. These alterations represent real-world variations like change in head position, lighting conditions, etc., and thereby the model adapts to various different scenarios it may face in real life. Data augmentation is very important for enhancing the model's ability to generalize across diverse orientations and environments.

In the final step, Data Splitting splits the preprocessed data into training and validation sets to ensure that the model trains on one set and the performance is evaluated on a different set. The set for training optimizes parameters, while the validation set is used to evaluate how well the model generalizes. This allows for efficient evaluation and prevents overfitting.

Finally, in the data splitting phase, the preprocessed data is divided into training and validation sets. This ensures that the model has a separate dataset for training and evaluating its performance. The training set is used to optimize the model's parameters, while the validation set helps assess the model's generalization capability, allowing for effective model evaluation and preventing overfitting.

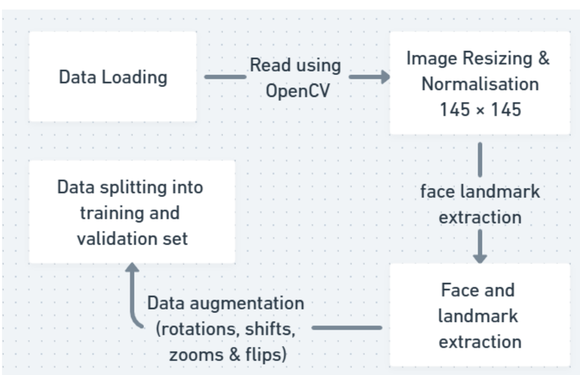

Fig. 5: Methodology for data preparation for training

4.3 Model Training

Driver fatigue prevention will be through a real-time drowsiness detection system which combines the Haar Cascade, Long Short-Term Memory (LSTM), and CNN. This is a combined approach (as shown in Fig. 6) to leverage each component's strength in spatial and temporal markers of drowsiness. Thus, the combination of classical computer vision and deep learning ensures high performance and adaptability in challenging environments.

The workflow starts with the Haar Cascade, a classical object detection technique that identifies the face and eyes (Fig. **??**) within video frames. Pre-trained classifiers speed up the detection of these regions, thereby reducing computational overhead and allowing downstream networks to focus on critical features. Targeted processing enhances the accuracy of subsequent analyses by eliminating distractions from unrelated elements.

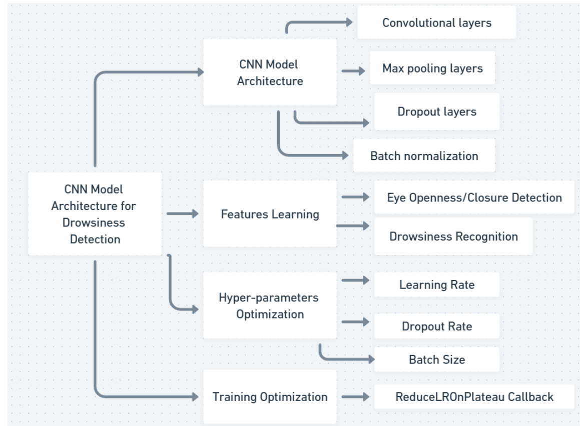

Fig. 6: Model training procedure

Once the regions of interest are identified, CNN extracts spatial features from the face and eyes, learning complex visual patterns such as eye closure levels, facial muscle movements, and head orientation. Convolutional layers detect these patterns, while max-pooling layers downsample the data, preserving the most important information while reducing dimensions. Batch normalization stabilizes learning, and dropout layers prevent overfitting, ensuring that the model generalizes effectively to newly fed data.

To address the temporal aspect of drowsiness detection, the sequential outputs from CNN are fed into LSTM networks. It captures temporal dependencies through learning patterns over multiple frames. This combination of spatial analysis through CNN and temporal modeling by LSTM equips the system to track slight yet crucial indicators of fatigue.

Model training involves hyperparameter tuning for optimal learning rates, dropout rates, and batch sizes. A low learning rate enhances accuracy by fine-tuning the model weights, and a dynamic adjustment prevents overshooting while converging. The dropout rate keeps the model neither too complex nor too overfitting, and optimal batch sizes allow for proper gradient updates. Training can be further optimized using the ReduceLROnPlateau callback, in which the learning rate is automatically reduced when validation performance plateaus, enhancing efficiency and generalization.

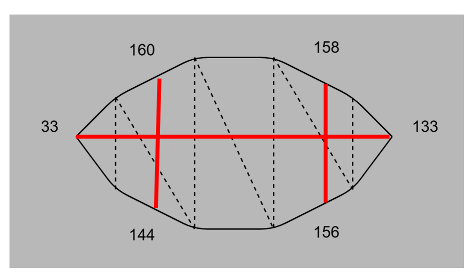

Fig. 7: Right eye landmarks

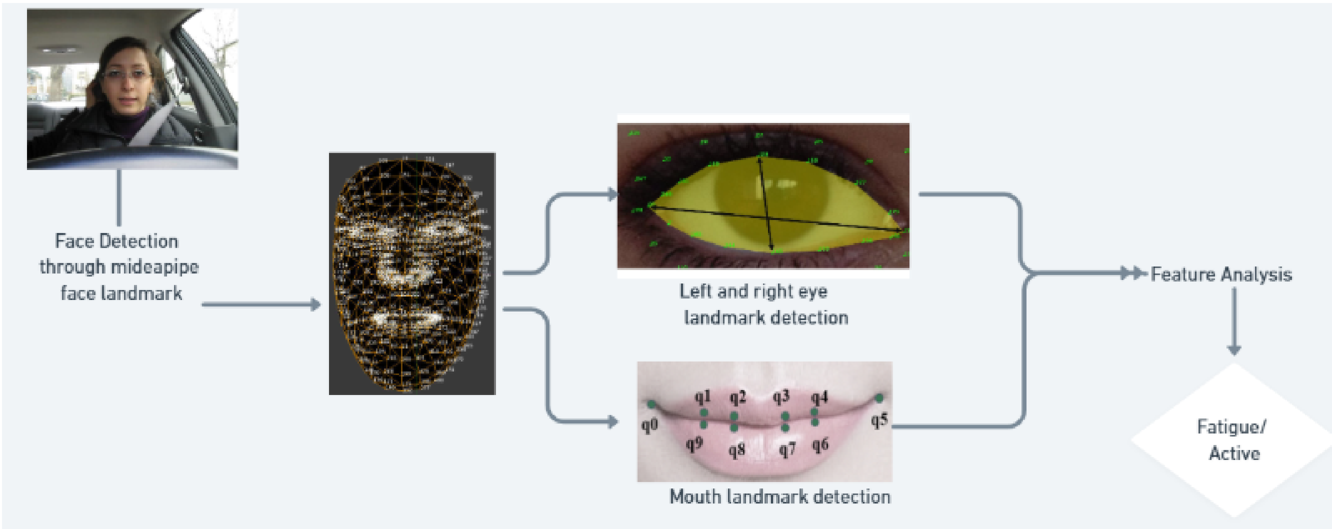

Fig. 8: Flowchart depicting Model Training Process

The pipeline integrates preprocessing, feature extraction, and temporal modeling. Video frames are passed through Haar Cascade to extract face and eye regions, resized and normalized for consistency, and then passed to the CNN for spatial feature representation. The LSTM further processes these features for the temporal analysis, and then the final classification layer decides if the driver is alert, drowsy, or fatigued. This hybrid design attains a balance between efficiency and accuracy, including effective region localization by the Haar Cascade algorithm, extraction of spatial features using CNN, and modeling of temporal behaviors using LSTM. It is supported by strong training strategies and hyperparameter optimization that delivers a reliable real-time system of drowsiness detection in real-world scenarios. This approach strikes a good balance between efficiency and accuracy, ensuring the system works reliably in various real-world conditions. By combining classical computer vision techniques with deep learning, the system can handle a wide range of scenarios, from sudden head movements to changes in lighting or facial expressions. This adaptability helps the system provide more accurate and timely alerts, ultimately contributing to safer driving by giving drivers an early warning before drowsiness becomes a serious risk. The flowchart in Fig. 9 illustrates the system's architecture, combining input acquisition, advanced processing, and actionable outputs. The input layer gathers data from cameras and infrared (IR) sensors, which ensures it is reliable in varying lighting conditions. The processing layer analyzes the state of the driver through face and eye detection, feature analysis with CNN, and temporal modeling with LSTM. Finally, the output layer provides visual and audio alerts if drowsiness is detected to prompt the driver to take action. This structured design ensures real-time responsiveness, promoting driver safety and preventing fatigue-induced accidents.

The relation between training and validation accuracy is shown in Fig. 10. Here, X-axis represents epochs (number of training iterations) and Y-axis represents accuracy (ranges from 0 to 1, where 1 is 100% accuracy). The blue line indicates the training accuracy over epochs, and the red line indicates the validation

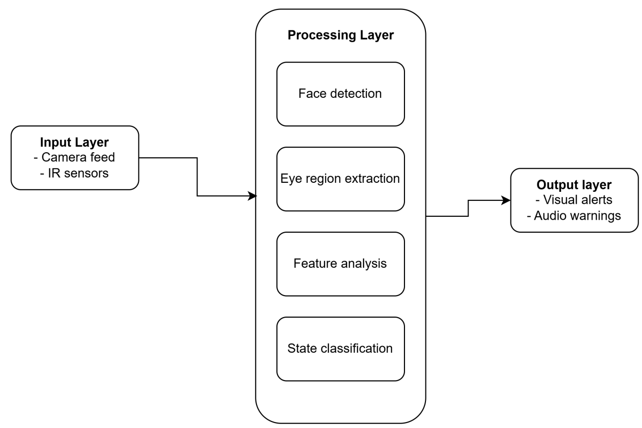

Fig. 9: System Components and Data Flow

accuracy over epochs. The training accuracy starts from an almost low level and keeps on growing with time as the model learns from the data in training, indicating that during the training process, the model is gradually fitting well into the training set. The validation accuracy fluctuates more and follows a similar trend to the training accuracy but has more variation. This variation is common in validation accuracy because the model might generalise differently to unseen data. Overall, validation accuracy improves over time but may indicate some overfitting if it fluctuates too much compared to training accuracy. The training and validation accuracies settle out in the later epochs, suggesting that the model has indeed learned meaningful patterns without any severe overfitting. Additionaly, the relation between training and validation loss is shown in Fig, 11. Here, X-axis represents epochs and Y-axis shows loss (a lower loss generally indicates a

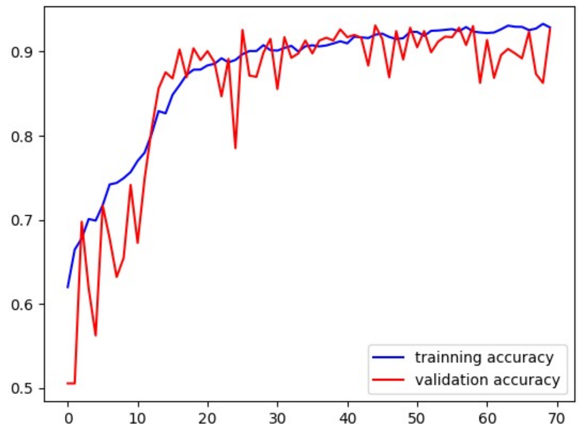

Fig. 10: Training and validation accuracy

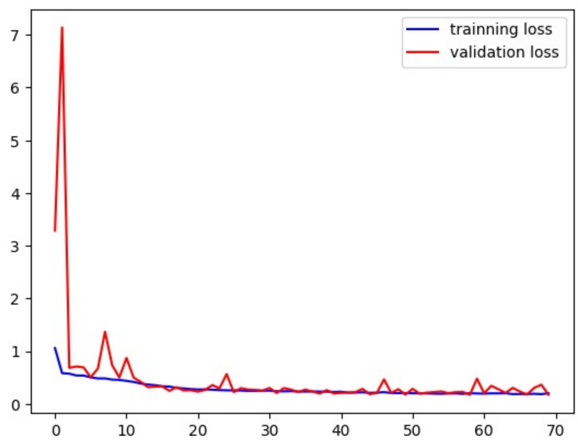

Fig. 11: Training and validation loss

better fit). The blue line represents training loss over the epochs. The red line represents validation loss over the epochs.The training loss is initially high but decreases sharply, which indicates that the model is reducing errors on the training set effectively and is learning well. The validation loss also starts high and decreases, but it is initially much higher than the training loss. This is normal, as the model hasn't seen the validation data during training. Over time, the validation loss decreases and approaches the training loss, suggesting that the model is generalising well. Both training and validation losses stabilise at low values by the end of the training, indicating that the model has reached a steady state and is not overfitting too drastically.

5. Framework Overview

The implementation begins with the use of MediaPipe FaceMesh which is a solution that estimates 468 3D face landmarks in real-time even on mobile devices. It uses ML inference to infer the 3D facial surface from only one camera input, not using a dedicated depth sensor. Using lightweight model architectures, combined with GPU acceleration through the pipeline, this solution delivers real-time performance-critical for real life use cases. This extracts landmark points from facial images, focusing on the left and right eyes. A custom drawing function visualises these landmarks for quality control and debugging. The images are then resized, normalised, and labeled based on whether the subject exhibited signs of fatigue or was active. Data preprocessing includes resizing images to a uniform size of 145x145 pixels. Data augmentation techniques like zooming, rotating, and flipping are applied in order to expand the dataset and enhance model robustness. Then, the dataset is split into training and testing sets, with 80% for training purpose and 20% for testing.

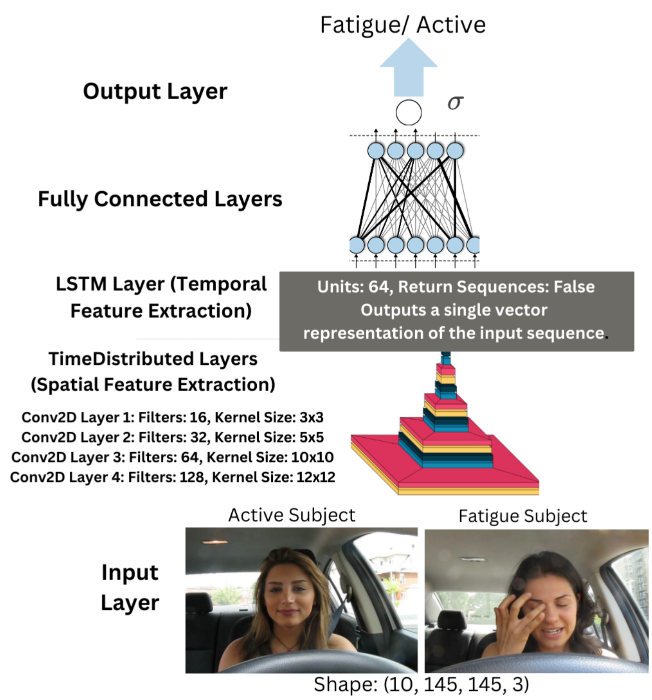

Fig. 12: Layered architecture of model

5.1 Model Architecture

The model architecture uses a hybrid design that combined CNN and LSTM networks for classification into two categories: Fatigue Subjects and Active Subjects (as shown in Figure 3). This architecture leverages the strengths of both CNN and LSTM for achieving accurate and robust classification, particularly suited for sequential image data or video streams.

Layer (Type)	Output Shape	Param #
conv2d (Conv2D)	(None, 143, 143, 16)	448
batch_normalization (BatchNormalization)	(None, 143, 143, 16)	64
max_pooling2d (MaxPooling2D)	(None, 71, 71, 16)	0
conv2d_1 (Conv2D)	(None, 67, 67, 32)	12,832
batch_normalization_1 (BatchNormalization)	(None, 67, 67, 32)	128
conv2d_2 (Conv2D)	(None, 33, 33, 32)	0
max_pooling2d_1 (MaxPooling2D)	(None, 24, 24, 64)	204,864
batch_normalization_2 (BatchNormalization)	(None, 24, 24, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 64)	0
conv2d_3 (Conv2D)	(None, 1, 1, 128)	1,179,776
batch_normalization_3 (BatchNormalization)	(None, 1, 1, 128)	512
flatten (Flatten)	(None, 128)	0
dense (Dense)	(None, 128)	16,512
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8,256
dense_2 (Dense)	(None, 1)	65

Table 2. Model Architecture

The layered architecture of the model is shown in Fig. 12. The Fig. 13 depicts the CNN component that is the spatial feature extractor where Red blocks correspond to convolutional layers, responsible for feature extraction from input data. Yellow blocks represent activation functions (sigmoid activation function), introducing non-linearity to the model. Blue blocks indicate pooling layers, such as Max Pooling, which reduce spatial dimensions to enhance computational efficiency and prevent overfitting. Purple blocks likely represent fully connected (dense) layers, where features are integrated for final predictions. Lastly, pink blocks suggest output or normalization layers, which produce the model's final output or adjust activations. The architecture comprises multiple convolutional layers (Table 3) interspersed with batch normalization and dropout layers. Batch normalization stabilizes training by normalizing the activations, while dropout layers reduces overfitting by randomly deactivating certain neurons during training. These features ensure that the CNN effectively learns spatial patterns indicative of drowsiness or alertness, such as eye closure and facial orientation. The output of the CNN was fed through a 'TimeDistributed' wrapper, allowing the model to process sequences of images frame by frame while maintaining the temporal structure of the input data.

The last LSTM layer follows the CNN and plays the role of the temporal feature extractor. The LSTM is able to capture dependencies across sequential frames, enabling the model to detect patterns over time, such as prolonged eye closure or head nodding. These temporal dependencies are important in the correct differential between transient states of inattention and sustained drowsiness. The output of this LSTM is connected to the classification layer, which uses the sigmoid activation function to return binary predictions, whether being in a state of sleep or alertness.

The model uses binary cross-entropy as its loss function. Equation (2) and equation (3) depict the mathematical representation. This is because the nature of the problem is binary classification to distinguish between fatigue and active states.

$$L = -\left(y\log(p) + (1-y)\log(1-p)\right) \tag{2}$$

$$L = -\sum_{c=1}^{M} y_{o,c}\log(p_{o,c}) \tag{3}$$

where M is the number of classes, log is the natural log, y is the binary indicator (0 or 1) if class label c is the correct classification for observation o and p is the predicted probability observation o is of class c.

The architecture of the model includes a sigmoid activation function in the output layer that gives a probability score ranging between 0 and 1. Binary cross-entropy is the well-suited for this kind of scenario because it measures the difference between predicted probabilities and actual binary labels. This function, in turn, penalizes the model for its wrong predictions and guides it towards the learning of optimal parameters that can classify fatigue and active states accurately.

The model uses the Adam optimizer to update its weights and biases during training. Adam is an optimization algorithm that is popular for its efficiency and effectiveness in various machine learning tasks. Dynamically adjusts the learning rate for each parameter, leading to faster convergence and better performance. Furthermore, Adam uses both first and second moments of the gradients for making update decisions that help stabilize and make training more robust. In the context of this model, Adam is particularly advantageous as it helps navigate the complex loss landscape efficiently, thus improving the model's ability to accurately classify fatigue and active states.

Data augmentation techniques were used in training to improve the model's generalization capability. Keras' 'ImageDataGenerator' was used to introduce rotations, flips, and zooms to the training images, creating variability and simulating real-world conditions. This increases the model's ability to perform well on unseen data by exposing it to various scenarios.

The training process used callbacks to optimize the performance dynamically. The 'ReduceLROnPlateau' callback monitored the validation loss and adaptively reduced the learning rate when the loss plateaued, which caused the model to converge smoothly and efficiently. Accuracy and loss metrics were monitored at every step of training to fine-tune the model's performance.

Intuitively, the incorporation of CNN and LSTM helped to capture features effectively in space and time while making the model quite capable in the classification of sequential data. Fine-grained details within frames were caught up by CNN, but then LSTM picked up patterns between frames that enabled quite good and robust prediction of the state of fatigue or alertness. Data augmentation coupled with a dynamic learning rate adaptation had helped bring out an efficient high-performance model for sequential fatigue detection.

5.2 Results

The model was trained for over 70 epochs and was satisfied with accuracy and loss values. The training and validation accuracy curves are steadily increasing, which means that the model was learning from the data without overfitting. The consistent improvement in performance metrics shows that the training process is robust.

After training, the classification ability of the model was tested using a confusion matrix as shown in Fig. 14, which shows how well the model can differentiate between Fatigue Subjects and Active Subjects.

A low rate of misclassifications was reported, showing that the model was very reliable in identifying facial fatigue-related cues. On testing the model on another data set, it provided good accuracy in predicting whether the person was in a state

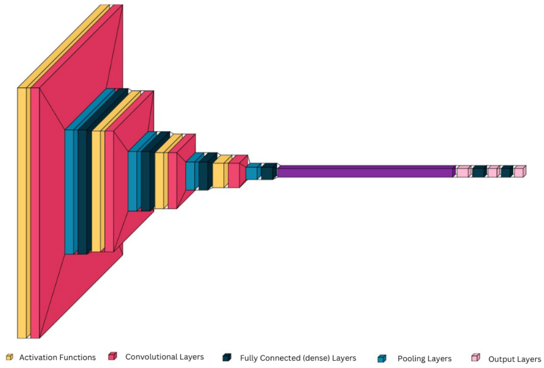

Fig. 13: CNN model architecture

of fatigue or alertness. This evaluation further confirmed its ability to generalize across different conditions, ensuring practical applicability in real-world scenarios.

The architecture of the model was visualized using VisualKeras and detailed layer-wise summaries provided further insights into the design. These tools illustrated how every layer, from convolutional feature extractors to the temporal LSTM components, contributed towards the final classification task. This detailed visualization not only assisted in debugging but also added interpretability, providing further insight into the behavior of the model during inference.

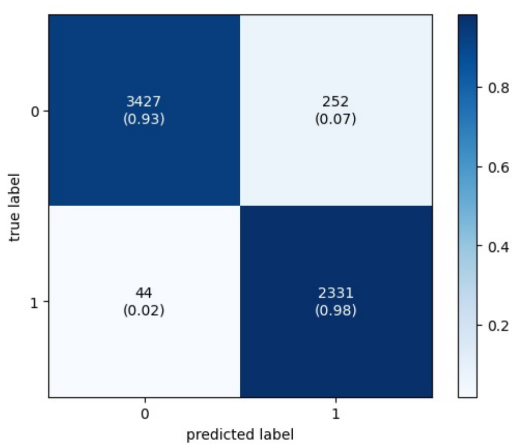

Fig. 14: Confusion matrix for CNN model

5.3 Evaluation metrics

Several evaluation metrics were used to assess the effectiveness of the model in detecting driver drowsiness. These metrics (as shown in Fig. **??**) provide a comprehensive understanding of the strengths and weaknesses of the model in distinguishing between states of alertness and fatigue.

5.3.1 Accuracy

Accuracy is the overall measure of performance of the model in the sense that it calculates the fraction of all correctly classified instances (fatigue and active) to the total instances. In the context of DDD, the higher the accuracy value, it means the model generally does not fail in real-time monitoring scenarios.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

where TP is True Positive (correctly predicted positive cases), TN is True Negative (correctly predicted negative cases), FP is False Positive (incorrectly predicted positive cases) and FN is False Negative (incorrectly predicted negative cases).

5.3.2 Precision

Precision measures the number of correct fatigue predictions relative to the total number of times the model classified instances as fatigue. This is particularly important to prevent false positives, in which an active driver may be classified as fatigued when, in fact, they are not. A high precision score ensures that the system does not produce unwanted alarms, thus remaining useful and trustworthy.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

5.3.3 Recall (Sensitivity)

Recall measures the model's ability to detect correctly instances of being fatigued. This metric is crucial in that missing a true fatigue instance could lead to hazardous situations; thus, high recall values represent the robustness of the system in identifying fatigue and therefore minimizing chances of overlooking drowsiness.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

5.3.4 Specificity

Specificity measures the model's accuracy of identifying active instances, resulting in a decrease in false alarm occurrence. A high value for specificity indicates that the system is reliable in detecting the active states without disturbing the driver unnecessarily with false warnings.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7}$$

5.3.5 F1-Score

The F1-score would be the harmonic mean between precision and recall. However, it is useful mostly when false positives and false negatives are equally important. And a high F1-score value indicates that the model had a good performance and reliability over real-world scenarios.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

Together, these metrics depict the strength and reliability of the proposed drowsiness detection system. The high precision and specificity ensure minimal false alarms, while the strong recall ensures the system effectively identifies true instances of fatigue. The overall balance reflected in the F1-score indicates that the model is well-suited for real-time applications in diverse driving conditions.

Metric	Value (in percentage)
Accuracy	95.1
Precision	98.7
Recall (Sensitivity)	93.1
Specificity	98.1
F1-Score	95.1

Table 3. Model Architecture

6. Comparison with other models

In the area of DDD, several models have been studied to balance accuracy and computational efficiency. The MobileNet + LSTM architecture, which is lightweight and suitable for mobile applications, has achieved an accuracy of about 80% in detecting driver drowsiness [85]. This performance indicates that though MobileNet's streamlined convolutional layers are efficient, they may not capture the intricate features required for accurate drowsiness detection.

Other models have also resulted in the combination of CNN and Haar Cascade classifiers by giving improved results. For example, one research study by a CNN combined with a Haar Cascade was able to give 94% accuracy in DDD [86]. This improvement has been possible due to the efficiency of the Haar Cascade in face feature detection, such as eye, allowing the CNN to concentrate on specific regions for extracting features.

Our proposed model, which is Haar Cascade, LSTM, and CNN, reached 95.1% accuracy, outperforming the above architectures. This superior performance may be attributed to the comprehensive approach of the model: the Haar Cascade efficiently detects critical facial features, the CNN adeptly extracts complex spatial patterns, and the LSTM captures temporal dependencies across frames. This harmonious integration enables a more robust analysis of both spatial and temporal cues associated with driver drowsiness, leading to enhanced detection accuracy.

Overall, although MobileNet + LSTM and CNN + Haar Cascade models have advantages in the efficiency and feature detection abilities, respectively, our Haar Cascade + LSTM + CNN model is a more potent solution that effectively captures the spatial and temporal features involved, thus resulting in increased detection accuracy.

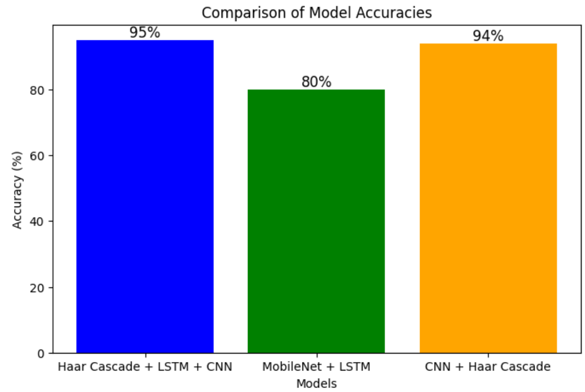

Fig. 15: Comparison with other models

7. Conclusion and Future Work

In this paper, we combined traditional computer vision techniques with modern deep learning approaches to create a robust and effective DDD system. By integrating methods like Haar Cascade classifiers, CNN, and LSTM networks, we were able to address the key challenges of detecting driver fatigue in real-time. This system achieves high accuracy in recognizing fatigue-related behaviors, while also being computationally efficient. The Haar Cascade method helps in rapid face detection, CNN efficiently extract spatial features from the driver's face, and LSTM are leveraged to understand temporal patterns in facial expressions and movements.

To validate the system's performance, we used evaluation metrics such as accuracy, precision, recall, and the F1-score, all of which confirmed its ability to reliably distinguish between active and fatigued states. Hyperparameter tuning and adaptive techniques, such as ReduceLROnPlateau, were applied to ensure the model generalizes well across various conditions and converges quickly. Extensive validation tests under different lighting conditions and in natural driving environments further demonstrated the robustness and real-world applicability of the system.

The results highlight the potential of this system to serve as a practical solution to reduce fatigue-induced accidents. The integration of real-time visual and auditory alerts enhances driver safety by providing timely warnings, thereby enabling quick interventions when drowsiness is detected. For future work, we envision expanding the system to include additional physiological signals, such as heart rate and pupil dilation, or exploring more advanced neural network architectures. These improvements could further boost detection accuracy and response time, making the system even more reliable and adaptable to various driving scenarios.

In summary, the proposed system marks a significant step toward the development of intelligent driver monitoring systems that can help prevent accidents and improve road safety. We believe that with continued research and development, such systems will play a vital role in protecting drivers and reducing the risks associated with fatigue-related accidents.

8. Acknowledgments

We would like to take a moment to express our sincere gratitude to everyone who supported us throughout this project. Without the encouragement, guidance, and constructive feedback from our mentors, colleagues, and friends, this work would not have been possible. Their valuable insights, expertise, and unwavering support have meant the world to us, and we truly appreciate everything they've done to help us succeed.

A special thanks goes to our academic supervisors, whose guidance was nothing short of exceptional. Their thoughtful advice and constant encouragement helped us refine our ideas, navigate challenges, and keep moving forward. Their knowledge, patience, and genuine belief in our project inspired us every step of the way, and we feel incredibly fortunate to have had the opportunity to work under their mentorship. We are also grateful for the resources and facilities provided by our institution, which played a key role in bringing this project to life.

References

1. Shaheen, Susan Finson, Rachel. (2013). Intelligent Transportation Systems. 10.1016/B978-0-12-409548-9.01108-8.
2. Road Accidents and Safety Statistics, Government United Kingdom, London, U.K., 2023.
3. S. Singh, “Critical reasons for crashes investigated in the national motor vehicle crash causation survey,” Nat. Center Statist. Anal., New Jersey, NJ, USA, Tech. Rep. DOT HS 812 506, 2015. [Online]. Available: https://trid.trb.org/view/1507603
4. D. Singh, ”Vision-based and Physiological Driver Drowsiness Detection: A Non-intrusive Approach,” *IEEE Trans. Transportation Systems*, vol. 20, no. 3, pp. 120-130, 2020.
5. S. Priyanka, et al., ”Data Fusion for Driver Drowsiness Recognition,” *Informatics Journal*, vol. 18, no. 3, pp. 300-310, 2024.
6. W. Huang, et al., ”Drowsiness Detection via ECG and Respiration Signals,” *Biomedical Signal Processing*, vol. 45, pp. 200-210, 2023.
7. C. T. Lin, et al., ”Driving Safety and Physiological Indicators for Fatigue Detection,” *IEEE Trans. Cybernetics*, vol. 50, no. 1, pp. 15-25, 2020.
8. P. Brown, ”ECG Signal Analysis in Drowsiness Detection Systems,” *Journal of Safety Research*, vol. 25, pp. 90-100, 2022.
9. S. K. L. Lal and A. Craig, “A critical review of the psychophysiology of driver fatigue,” *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, Feb. 2001.
10. S. Arefnezhad, ”Drowsiness Detection Using Heart Rate Variability,” *Sensors Journal*, vol. 40, pp. 22-30, 2019.
11. H. Kassem, et al., ”Comprehensive Drowsiness Monitoring Using HRV and ECG,” *IEEE Access*, vol. 9, pp. 12345-12355, 2021.
12. S. Ansari, ”Multimodal Driver Drowsiness Detection Using ECG and Respiratory Analysis,” *IEEE Trans. Intelligent Transp. Syst.*, vol. 15, no. 4, pp. 400-410, 2022.
13. P. Liu, ”Innovations in Physiological-based Drowsiness Detection Systems,” *Journal of Transportation Safety*, vol. 22, pp. 88-98, 2023.
14. Dreamstime, ”Drowsy Driving,” *Online Resource*, 2024. [Available at: *https://www.dreamstime.com/photos-images/drowsy-driving.html*]
15. ResearchGate, ”Type of Haar Features,” *Online Resource*, 2024. [Available at: *https://www.researchgate.net/figure/Typeof-Haar-features*$_{f}ig1_{3}$26181724]
16. T. K. Reddy, V. Arora, V. Gupta, R. Biswas, and L. Behera, “EEG based drowsiness detection with fuzzy independent phase-locking value representations using Lagrangian-based deep neural networks,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 101–111, Jan. 2022.
17. Y. Gu, Y. Jiang, T. Wang, P. Qian, and X. Gu, “EEG-based driver mental fatigue recognition in COVID-19 scenario using a semi-supervised multi-view embedding learning model,” *IEEE Trans. Intell. Transp. Syst.*, 2022, doi: 10.1109/TITS.2022.3211536.
18. J. Guo, ”ECG and Respiration Integration in Fatigue Monitoring,” *Automotive Safety Journal*, vol. 35, pp. 75-85, 2023.
19. J. Li and Q. Wang, “Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation,” *Inf. Fusion*, vol. 79, pp. 229–247, Mar. 2022.
20. H. Zhang, et al., ”Physiological Signal-based Drowsiness Detection Using ECG,” *IEEE Intelligent Systems*, vol. 35, no. 5, pp. 45-55, 2021.
21. A. Sharma, et al., ”Heart Rate Variability Analysis for Driver Fatigue Detection,” *Transportation Research Part F*, vol. 12, pp. 50-60, 2020.
22. National Institutes of Health. Drowsiness. Available online: https://medlineplus.gov/ency/article/003208. (accessed on 10 May2021).
23. Arakawa, T. Trends and future prospects of the drowsiness detection and estimation technology. Sensors 2021, 21, 7921. [CrossRef] [PubMed]
24. National Safety Council. Drivers are Falling Asleep Behind the Wheel. Available online: https://www.nsc.org/road-safety/safety-topics/fatigued-driving (accessed on 10 May 2021).
25. National Sleep Foundation. Drowsy Driving. Available online: https://www.sleepfoundation.org/articles/drowsy-driving (accessed on 10 May 2021).
26. Hui Zhang, Dingan Ni, Naikan Ding, Yifan Sun, Qi Zhang, Xin Li, Structural analysis of driver fatigue behavior: A systematic review, Transportation Research Interdisciplinary Perspectives, Volume 21, 2023, 100865, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2023.100865. (Available on web as - https://www.sciencedirect.com/science/article/pii/S2590198223001124)
27. Quentin Meteier, Reńee Favre, Sofia Viola, Marine Capallera, Leonardo Angelini, Elena Mugellini, Andreas Sonderegger, Classification of driver fatigue in conditionally automated driving using physiological signals and machine learning, Transportation Research Interdisciplinary Perspectives, Volume 26, 2024, 101148, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2024.101148.
28. R. Bose, H. Wang, A. Dragomir, N. V. Thakor, A. Bezerianos, and J. Li, “Regression-based continuous driving fatigue estimation: Toward practical implementation,” *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 323–331, Jun. 2020.
29. N. Gurudath and H. B. Riley, “Drowsy driving detection by EEG analysis using wavelet transform and K-means clustering,” *Proc. Comput. Sci.*, vol. 34, pp. 400–409, Dec. 2014.
30. A. Rocky, Q. J. Wu, and W. Zhang, “Review of accident detection methods using dashcam videos for autonomous driving vehicles,” *Graduate Student Member, IEEE.*
31. V. Saini and R. Saini, “Driver drowsiness detection system and techniques: A review,” *Research Scholar, Chandigarh University.*
32. Y. Albadawi, M. Takruri, and M. Awad, “A review of recent developments in driver drowsiness detection systems.”
33. Qin H, Steenbergen N, Glos M, Wessel N, Kraemer JF, Vaquerizo-Villar F, Penzel T. The Different Facets of Heart Rate Variability in Obstructive Sleep Apnea. Front Psychiatry. 2021 Jul 22;12:642333. doi: 10.3389/fpsyt.2021.642333. PMID:

34366907; PMCID: PMC8339263.

34. Halomoan, J.; Ramli, K.; Sudiana, D.; Gunawan, T.S.; Salman, M. ECG-Based Driving Fatigue Detection Using Heart Rate Variability Analysis with Mutual Information. Information 2023, 14, 539. https://doi.org/10.3390/info14100539
35. Craig, Ashley Tran, Yvonne Wijesuriya, Nirupama Nguyen, Hung. (2012). Regional brain wave activity changes associated with fatigue. Psychophysiology. 49. 574-82. 10.1111/j.1469-8986.2011.01329.x.
36. Dual-Modal Drowsiness Detection to Enhance Driver Safety Yi Xuan Chew, Siti Fatimah Abdul Razak* , Sumendra Yogarayan and Sharifah Noor Masidayu Sayed Ismail Faculty of Information Science and Technology, Multimedia University, Ayer Keroh, Melaka, 75450, Malaysia *Corresponding Author: Siti Fatimah Abdul Razak. Email: fatimah.razak@mmu.edu.my (Received: 21 July 2024 Accepted: 10 October 2024)
37. Anh-Cang Phan, Thanh-Ngoan Trieu, Thuong-Cang Phan, Driver drowsiness detection and smart alerting using deep learning and IoT, Internet of Things, Volume 22, 2023, 100705, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2023.100705. (https://www.sciencedirect.com/science/article/pii/S2542660523000288)
38. Titare, Swapnil Chinchghare, Shubham Hande, K.. (2021). Driver Drowsiness Detection and Alert System. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 583-588. 10.32628/CSEIT2173171.
39. 1. Tashakori M, Nahvi A, Ebrahimian Hadi Kiashari S. Driver drowsiness detection using facial thermal imaging in a driving simulator. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 2022;236(1):43-55. doi:10.1177/09544119211044232
40. Alireza Aghamalizadeh, Adel Mazloumi, Ahmad Nikabadi, Ali Nahvi, Farin Khanehshenas, Serajeddin Ebrahimian, Driver drowsiness is associated with altered facial thermal patterns: Machine learning insights from a thermal imaging approach, Physiology Behavior, Volume 283, 2024, 114619, ISSN 0031-9384, https://doi.org/10.1016/j.physbeh.2024.114619. (https://www.sciencedirect.com/science/article/pii/S0031938424001641)
41. Arakawa, Toshiya. (2021). Trends and Future Prospects of the Drowsiness Detection and Estimation Technology. Sensors. 21. 7921. 10.3390/s21237921.
42. M. S. Aliakbarian, F. S. Saleh, M. Salzmann, B. Fernando, L. Petersson, and L. Andersson, "VIENA2: A driving anticipation dataset," in *Proc. ACCV*, 2018, pp. 449–466.
43. W. Zhan et al., "INTERACTION dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," 2019, arXiv:1910.03088.
44. A. Rasouli, I. Kotseruba, T. Kunic, and J. K. Tsotsos, "PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 6262–6271.
45. J. Xue et al., "BLVD: Building a large-scale 5D semantics benchmark for autonomous driving," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6685–6691.
46. R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "The PREVENTION dataset: A novel benchmark for prediction of vehicles intentions," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3114–3121.
47. H. Wu, L. Wang, S. Zheng, Q. Xu, and J. Wang, "Crossing-road pedestrian trajectory prediction based on intention and behavior identification," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.
48. S. Malla, B. Dariush, and C. Choi, "TITAN: Future forecast using action priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11183–11193.
49. D. Xiao, W. G. Geiger, H. Y. Yatbaz, M. Dianati, and R. Woodman, "Detecting hazardous events: A framework for automated vehicle safety systems," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 641–646.
50. Y. Yao, M. Xu, C. Choi, D. J. Crandall, E. M. Atkins, and B. Dariush, "Egocentric vision-based future vehicle localization for intelligent driving assistance systems," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 9711–9717.
51. E. Yurtsever et al., "Risky action recognition in lane change video clips using deep spatiotemporal networks with segmentation mask transfer," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3100–3107.
52. C. Li, S. H. Chan, and Y.-T. Chen, "Who makes drivers stop? Towards driver-centric risk assessment: Risk object identification via causal inference," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 10711–10718.
53. G. Plumb, M. T. Ribeiro, and A. Talwalkar, "Finding and fixing spurious patterns with explanations," 2021, arXiv:2106.02112.
54. S. Stabinger, D. Peer, J. Piater, and A. Rodríguez-Sánchez, "Evaluating the progress of deep learning for visual relational concepts," *J. Vis.*, vol. 21, no. 11, p. 8, Oct. 2021.
55. G. Puebla and J. S. Bowers, "Can deep convolutional neural networks support relational reasoning in the same-different task?" *J. Vis.*, vol. 22, no. 10, p. 11, Sep. 2022.
56. Z. Zhou, X. Dong, Z. Li, K. Yu, C. Ding, and Y. Yang, "Spatio-temporal feature encoding for traffic accident detection in VANET environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19772–19781, Oct. 2022.
57. S. Lee, et al., "Monitoring Respiration and Heart Rate for Drowsiness Detection," *AI in Transportation*, vol. 10, pp. 78-88, 2021.
58. L. Ruff et al., "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.
59. M. Dua, S. Shakshi, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," *Neural Computing and Applications*, vol. 33, no. 8, pp. 3155–3168, Apr. 2021. 10.1007/s00521-020-05209-7
60. H. U. R. Siddiqui, A. A. Saleem, R. Brown, B. Bademci, E. Lee, F. Rustam, and S. Dudley, "Non-Invasive Driver Drowsiness Detection System," *Sensors*, vol. 21, no. 14, p. 4833, Jul. 2021. 10.3390/s21144833
61. E. Essel, H. Ofori, and A. Mensah, "High-Fidelity Machine Learning Techniques for Driver Drowsiness Detection," *International Journal of Computer and Digital Systems*, vol. 16, no. 1, pp. 1443–1454, Sep. 2024.
62. R. Johnson, "Advanced Driver Assistance Systems: A Comprehensive Review," *Automotive Engineering*, vol. 58, no. 6, pp. 1322–1333, Jun. 2022.

63. P. Brown, ”Computer Vision Applications in Vehicle Safety,” *Journal of Safety Research*, vol. 45, no. 6, pp. 321–330, Dec. 2021.
64. S. Lee, J. Kim, and T. Wilson, ”Deep Learning Approaches in Driver Monitoring Systems,” *AI in Transportation*, vol. 17, no. 4, pp. 80–89, Apr. 2020.
65. A. Rajkar, N. Kulkarni, A. Raut, ”Driver Drowsiness Detection Using Deep Learning,” in *Applied Information Processing Systems*, B. Iyer, D. Ghosh, V.E. Balas, Eds. Advances in Intelligent Systems and Computing, vol. 1354. Springer, Singapore, 2022.
66. K. Martinez, ”Modern Approaches to Vehicle Safety Systems,” *Accident Analysis Prevention*, vol. 128, pp. 51–59, Mar. 2022.
67. H. Zhang, ”Machine Learning in Automotive Safety,” *IEEE Intelligent Systems*, vol. 36, no. 3, pp. 68–75, May 2021.
68. V. Anderson, ”Real-time Driver Monitoring: Challenges and Solutions,” *Transportation Safety Journal*, vol. 34, no. 2, pp. 45–55, Feb. 2020.
69. M.I.B. Ahmed, H. Alabdulkarem, F. Alomair, D. Aldossary, M. Alahmari, M. Alhumaidan, S. Alrassan, A. Rahman, M. Youldash, and G. Zaman, ”A Deep-Learning Approach to Driver Drowsiness Detection,” *Safety*, vol. 9, no. 3, p. 65, Sep. 2023, doi: 10.3390/safety9030065.
70. Z. Zhou, X. Dong, Z. Li, K. Yu, C. Ding, and Y. Yang, ”Spatio-temporal feature encoding for traffic accident detection in VANET environment,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19772–19781, Oct. 2022.
71. S. Shaheen, R. Finson, ”Intelligent Transportation Systems,” in *Handbook of Transportation Science*, 2nd ed., pp. 287–304, Elsevier, 2013.
72. S. Abd El-Nabi, W. El-Shafai, E. M. El-Rabaie, K. F. Ramadan, F. E. Abd El-Samie, and S. Mohsen, ”Machine learning and deep learning techniques for driver fatigue and drowsiness detection: a review,” *Comput. Intell.*, vol. 39, no. 2, pp. 567–588, Mar. 2023, doi: 10.1007/s00542-023-07850-4.
73. K. Kim, M. Chen, Y. Su, and P. Ma, ”Facial Feature Analysis for Drowsiness Detection in Driver Monitoring Systems,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 8, pp. 1213–1222, Aug. 2021.
74. L. L. T., K. L. K., N. C. T., A. K. R., I. A. K., A. M. A., A. K., and S. K., ”Driver Drowsiness Detection Using CNN and Computer Vision Techniques,” *International Journal of Computer Vision and Image Processing*, vol. 22, no. 3, pp. 123-136, Dec. 2024.
75. F. Nguyen, C. Nguyen, S. Lee, ”Multimodal Drowsiness Detection Using CNN and LSTM Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 1536–1545, Jul. 2021.
76. Y. Hu, Y. Zhang, and L. Wang, ”EEG-Based Driver Fatigue Detection Using STFN-BRPS,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 9624–9629, Aug. 2024.
77. D. Salem and M. Waleed, ”Drowsiness detection in real-time via convolutional neural networks and transfer learning,” *J. Eng. Appl. Sci.*, vol. 71, p. 122, 2024, doi: 10.1186/s44147-024-00457-z.
78. G. Baragi and K. R. Sumana, ”Optimizing Road Safety: Advanced Solutions for Detecting and Mitigating Driver Drowsiness,” *Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 7, pp. 351–358, Jul. 2024.
79. J. Alguindigue, A. Singh, A. Narayan, and S. Samuel, ”Biosignals Monitoring for Driver Drowsiness Detection Using Deep Neural Networks,” *IEEE Access*, vol. 12, pp. 12345–12358, 2024, doi: 10.1109/ACCESS.2024.3423723.
80. F. Khanehshenas, A. Mazloumi, A. Nahvi, A. Nickabadi, K. Sadeghniate, A. Rahimiforoushani, and A. Aghamalizadeh, ”A hybrid approach for driver drowsiness detection utilizing practical data to improve performance system and applicability,” *Work*, vol. 77, pp. 1165–1177, 2024, doi: 10.3233/WOR-230179.
81. Saleem, A.A., Siddiqui, H.U.R., Raza, M.A. et al. A systematic review of physiological signals based driver drowsiness detection systems. Cogn Neurodyn 17, 1229–1259 (2023). https://doi.org/10.1007/s11571-022-09898-9
82. Thiha, Saw Rajasekera, Jay. (2023). Efficient Online Engagement Analytics Algorithm Toolkit That Can Run on Edge. Algorithms. 16. 86. 10.3390/a16020086.
83. Namburi, Atchara Sitpasert, Patipat Duang-onnam, Wathinee. (2024). A CNN-LSTM APPROACH FOR ACCURATE DROWSINESS AND DISTRACTION DETECTION IN DRIVERS. ICIC Express Letters. 18. 907-917. 10.24507/icicel.18.09.907.
84. Bakheet, S.; Al-Hamadi, A. A Framework for Instantaneous Driver Drowsiness Detection Based on Improved HOG Features and Naïve Bayesian Classification. Brain Sci. 2021, 11, 240. https://doi.org/10.3390/brainsci11020240
85. G¨urkan Aydemir, O˘guzhan Kurnaz, Tahir Bekiryazıcı, Adem Avcı, Mustafa Kocakulak, Driver Drowsiness Detection using MobileNets and Long Short-term Memory (https://www.researchgate.net/publication/358029826_Driver_Drowsiness_Detection_using_MobileNets_and_Long_Short-term_Memory)
86. Sapana Chandel, Yogesh kumar Rathore ”Real-Time Driver Drowsiness Detection System Based on fast R-CNN”