Order N^{o} : ../../..

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY KASDI MERBAH OF OUARGLA

Faculty of New Information Technologies and Communication

Department of Electronics and Telecommunications

ACADEMIC MASTER'S DEGREE

Field : **Sciences and Technology**

Branch : **Electronics**

Option : **Electronics of Embedded systems**

Presented by :

AOUINI Samar

CHAREF Khouloud

Theme

Development of driver safety system through the application of deep learning techniques for drowsiness detection

Publicly defended on 02/06/2025 before the jury composed of :

Dr. AZEDDINE benlamoudi	*MCA*	*University Kasdi Merbah - Ouargla*	*President*
Dr. KORICHI Maarouf	*MCA*	*University Kasdi Merbah - Ouargla*	*Examiner*
Dr. SAMAI Djamel	*Pr.*	*University Kasdi Merbah - Ouargla*	*Supervisor*
RAHMANI chahrazed	*PhD*	*University Kasdi Merbah - Ouargla*	*Co-Supervisor*

Dedication

To my beloved parents,
You alone deserve the first lines of my gratitude. You are the original story from which I learned patience and the true meaning of unconditional giving. I would not have reached this stage without your sincere prayers and hearts that believed in me even in my silence.

To the soul that illuminated the darkness of my days,
Your silent support and hidden encouragement left an impact on me beyond words. Though your name is unspoken, your presence is felt in every achievement.

To my colleague Charef Khouloud,
My partner in both the small and great details, who has always illuminated my days with her smile and sincerity. When work weighed heavily on me, your words eased the burden, and when the path seemed complex, you were the compass that brought me back to focus. Thank you for being more than a colleague ; you have been a true support throughout this journey.

And to myself,
To that resilient soul who endured more than expected, persisting despite fatigue, setbacks, and fear. Every time you were close to surrendering, you found a new reason to continue. This work is not just an academic achievement but a reflection of a long journey filled with challenges and small victories unseen by anyone but you. Be proud of how far you have come and always remember that you have deserved to arrive since the very beginning.

AOUINI SAMAR

Dedication

In the name of ALLAH, my Creator and the One who facilitates my affairs, the Protector of my journey to You belongs all praise and gratitude.

I dedicate this success first to myself, and then to everyone who supported me in completing this journey.May you always remain a pillar of strength in my life.

To the one who supported me unconditionally and gave without expecting anything in return,to the one who taught me that life is a struggle and that its weapon is knowledge and learning, and to the one who planted noble morals in my soul,my first supporter and my strength after Allah "my dear father".

To the one under whom Allah has placed Paradise, to the one whose prayers were the secret to my success, and whose kindness healed my wounds, my role model, my first teacher, and the companion of my days "my loving mother".

To those whom Allah strengthened me through, who have been my greatest support "my brothers and sisters".

To my colleague and companion in this project and throughout all the years of university,"Aouini Samar", who was my support in every step, we shared ambitions and challenges,you have always been a symbol of persistence and dedication,and a priceless partner in my university journey.

I do not forget the kindred spirits who walked this path with me, those who encouraged me to persevere and complete the journey,to lifelong friends, I am deeply grateful to you all.

KHOULOUD CHAREF

Acknowledgments

All praise is due to ALLAH, Lord of the Worlds, by whose grace and mercy this achievement was made possible. He granted us patience, strength, and determination to continue our academic journey despite all challenges. We extend our sincere gratitude to our esteemed supervisor, Prof. SAMAI Djamel, for his exceptional guidance and insightful directions. We also give special thanks to the distinguished Ms. CHAHRAZAD Rahmani, whose invaluable support was truly priceless. Her dedication, patience, and continuous encouragement played a vital role in overcoming obstacles, and her presence significantly contributed to enhancing the quality of this work and achieving its objectives.Additionally, we are grateful to Dr. AZEDDINE benlamoudi and Dr. KORICHI Maarouf, the jury members, for their readiness to review and assess our work.Our heartfelt thanks also go to our colleague, Mr. Aymen Benglia, for his ongoing cooperation and constructive motivation, which significantly supported and inspired us throughout this endeavor.

To all who stood by us and supported us, we offer our deepest appreciation and gratitude. You are all partners in this success and achievement.

الملخص

يقدم هذا المشروع نظامًا لكشف نعاس السائقين اعتمادًا على تحليل صور الوجه باستخدام تقنيات التعلم العميق. الهدف هو التعرف تلقائيًا على حالة السائق (يقظ أو نعسان) من خلال مؤشرات سلوكية مثل إغلاق العين، التثاؤب، وتغير وضعية الرأس.
تم استخدام قاعدة بيانات UTA-RLDD التي تحتوي على صور واقعية لسائقين في حالات مختلفة. بعد معالجة الصور، قمنا بتطبيق التعلم بالنقل باستخدام أربعة نماذج جاهزة : ResNet50 ، MobileNet ، DenseNet201 ، و EfficientNetV2. حقق نموذج EfficientNetV2 أفضل أداء بدقة وصلت إلى 99.98% . وقد تم تقييم النظام باستخدام مقاييس أداء دقيقة مثل الدقة، الاسترجاع، F1-score و AUC .
يثبت هذا العمل أن الجمع بين الرؤية الحاسوبية والتعلم العميق يمكن أن يوفر حلاً فعالاً وعمليًا لاكتشاف النعاس والمساهمة في تقليل حوادث الطرق.

الكلمات المفتاحية : كشف النعاس، التعلم العميق، التعلم بالنقل، الشبكات العصبية، تصنيف الصور، النماذج المعاد استخدامها، التصنيف الثنائي، تحليل الوجه، أنظمة مساعدة السائق.

Abstract

This project presents a driver drowsiness detection system based on facial image analysis using deep learning. The goal is to automatically identify whether a driver is alert or drowsy by analyzing behavioral signs such as eye closure, yawning, and head position.We used the UTA-RLDD dataset, which contains real facial images of drivers in both drowsy and non-drowsy states. After preprocessing the images, we applied transfer learning with four pre-trained models : ResNet50, MobileNet, DenseNet201, and EfficientNetV2.The best results were achieved using EfficientNetV2, which reached an accuracy of 99.98%. Our system was evaluated using standard performance metrics such as precision, recall, F1-score, and AUC.
This work shows that combining computer vision with deep learning can provide an accurate and practical solution for detecting driver fatigue, helping reduce the risk of accidents.

Keywords : Drowsiness detection, deep learning, transfer learning, neural networks, image classification , reused models, binary classification, facial analysis, driver assistance systems.

Résumé

Ce projet propose un système de détection de la somnolence chez les conducteurs basé sur l'analyse des images faciales à l'aide des techniques d'apprentissage profond. L'objectif est d'identifier automatiquement si un conducteur est en état d'éveil ou de somnolence en analysant des signes comportementaux tels que la fermeture des yeux, le bâillement et les changements de position de la tête.Nous avons utilisé la base de données UTA-RLDD, qui contient des images faciales réelles de conducteurs dans différents états. Après un prétraitement des images, nous avons appliqué l'apprentissage par transfert (transfer learning) à l'aide de quatre modèles pré-entraînés : ResNet50, MobileNet, DenseNet201 et EfficientNetV2.Le meilleur résultat a été obtenu avec EfficientNetV2, atteignant une précision de 99,98 %. Le système a été évalué à l'aide de métriques standards telles que la précision, le rappel, le score F1 et l'AUC.

Ce travail montre que la combinaison de la vision par ordinateur et de l'apprentissage profond permet de développer une solution fiable et pratique pour détecter la somnolence au volant, contribuant ainsi à la sécurité routière.

Mots-clés : Détection de la somnolence, apprentissage profond, transfert d'apprentissage, réseaux neuronaux, classification d'images, modèles réutilisés, classification binaire, analyse faciale, systèmes d'assistance à la conduite.

Contents

List of Figures

List of Tables

Abbreviations

ADAS :	Advanced Driver-Assistance System
AI :	Artificial Intelligence
ANN :	Artificial Neural Network
AUC :	Area Under the Curve
CNN :	Convolution Neural Network
DDD:	Driver Drowsiness Detection
DL:	Deep Learning
ECG :	Electrocardiography
EEG :	Electroencephalography
EMG :	Electromyography
EOG :	Electrooculography
ML :	Machine Learning
ReLU :	Rectification Linear Unit
RNN :	Recurrent Neural Network
ROC :	Receiver Operating Characteristic
RT :	Response Time
SVM :	Support Vector Machine
SWM :	Steering Wheel Movement
TL :	Transfer Learning
UTA-RLDD :	University of Texas at Arlington - Real-Life Drowsiness Dataset

Chapter 1

General introduction

Road safety remains one of the most pressing public concerns worldwide, particularly as traffic accidents resulting from driver fatigue continue to claim lives and cause significant economic losses. Traditional safety measures, while helpful, often fall short of addressing the real-time cognitive state of drivers. In this context, the integration of intelligent systems into vehicles has emerged as a vital strategy for monitoring driver alertness and enhancing road safety.

With the rapid development of artificial intelligence (AI) and, more specifically, deep learning (DL), the ability to analyze complex patterns from high-dimensional data has greatly improved. Deep learning models, inspired by the structure and functioning of the human brain, have demonstrated exceptional performance in tasks such as image recognition, speech processing, and behavior analysis. These capabilities are particularly relevant for developing systems that can detect signs of driver drowsiness through facial analysis.

Biometric-based approaches, which rely on the unique behavioral characteristics of individuals, offer a promising direction for real-time driver monitoring. Facial image analysis, in particular, provides a non-invasive and widely accepted method for identifying subtle signs of fatigue, such as eye closure, yawning, or facial muscle relaxation.

In response to the challenges of limited labeled data and the need for accurate classification, this work employs a combination of advanced deep learning techniques and transfer learning strategies. Transfer learning enables the adaptation of pre-learned visual knowledge acquired from large-scale datasets to the specific task of drowsiness detection, thereby reducing the computational cost and training time while improving model accuracy. Additionally, a pseudo-labeling process is applied to leverage unlabeled data further and enhance the model's robustness.

This research focuses on developing a binary classification framework that distinguishes between alert and drowsy states based on visual cues extracted from facial images. The model is trained and evaluated on real-world data from the UTA-RLDD dataset, which is recognized as one of the most extensive and reliable datasets for early drowsiness detection, making the results realistic and applicable in real-life driving conditions.

This dissertation is structured in the following manner.

THE SECOND CHAPTER provides a comprehensive overview of drowsiness detection, including its fundamental concepts, key indicators, and various detection approaches. We also examine the architecture and essential components of a typical drowsiness detection system, as well as the processes involved in identifying and monitoring driver fatigue. The chapter concludes with a comparative analysis of traditional and modern methods for detecting drowsiness, highlighting their respective advantages, limitations, and areas of application.

THE THIRD CHAPTER reviews the fundamental concepts of deep learning and transfer learning, along with an explanation of convolutional neural networks and their role in facial image classification for drowsiness detection.

THE FOURTH CHAPTER provides system design, preprocessing, and training methods and discusses results, performance, and real-world data impact.

THE FIFTH CHAPTER concludes the study by summarizing the main findings and suggesting directions for future work in AI-driven driver monitoring systems.

Chapter 2

Comprehensive Overview of Driving Drowsiness Detection

2.1 Introduction

Understanding and addressing driver drowsiness is a critical step toward improving road safety and reducing fatigue-related accidents. Over the years, various methods have been developed to detect signs of drowsiness in drivers, ranging from traditional approaches based on physiological signals and behavioral observation to modern techniques that leverage advancements in machine learning and computer vision. Traditional methods, such as monitoring eye closure, yawning frequency, and heart rate variability, have laid the foundation for early detection systems but often face limitations in accuracy and adaptability. In contrast, modern methods utilize deep learning models, real-time video analysis, and sensor fusion to provide more reliable and automated solutions. The integration of these cutting-edge technologies enables systems to recognize complex patterns of drowsiness with minimal human intervention, significantly enhancing detection performance. This chapter provides a comprehensive overview of both traditional and modern approaches, highlighting their respective strengths and limitations. By comparing these methods, we aim to understand the evolution of drowsiness detection technologies and emphasize the importance of continuous innovation in ensuring safer driving environments.

2.2 Factors Contributing to Driving Drowsiness

Drowsiness can greatly impact focus and alertness, especially while driving. While lack of sleep is a primary cause, other factors also contribute.

• Not getting enough sleep before a long drive can make you very tired and unsafe on the road.

• It's important to rest well before driving to stay alert and safe. Being awake for more than 24 hours can slow your brain and body, making it dangerous to drive.

• Drinking alcohol makes things worse. It makes you sleepy and slows down how fast you react and make decisions.

• Stress and anxiety can also make you feel tired, so it's important to relax and take care of yourself to stay focused while driving.

• Some medicines can also make you sleepy. Ask your doctor if your medicine might affect your driving. If not, avoid driving.

• Being sick or stressed all the time can make you feel worn out, making it also harder to focus while driving.

The dangers of drowsy driving are well documented. According to a study by the American Automobile Association Foundation for Traffic Safety, more than 320,000 drowsy driving accidents occur annually, resulting in approximately 6,400 fatal accidents [1]. other Research indicates that fatigue-related accidents account for around 20% of all road incidents, with this figure rising to 50% on certain roads. Fatigue remains a significant factor in numerous vehicle accidents, with recent statistics estimating that each year, around 1,200 deaths and 76,000 injuries are directly linked to fatigue-related crashes.[2] alarming figures highlight the critical importance of addressing drowsiness and fatigue to improve road safety and reduce preventable accidents.(Figure 2.1).

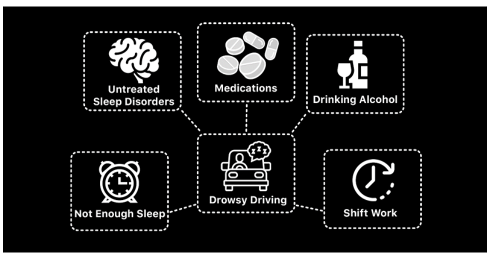


Figure 2.1 – Factors Causing Driving Drowsiness

2.3 Traditional Drowsiness Detection Techniques

Drowsiness detection methods are developed to help individuals recognize early signs of fatigue and take preventive actions before it poses a risk. Various approaches exist for detecting drowsiness, and many systems integrate multiple techniques to enhance accuracy and effectiveness.

2.3.1 Physiological Measures

Physiological measurements for detecting drowsiness involve monitoring a range of bodily functions and responses to assess an individual's level of alertness or fatigue. These methods are based on the principle that drowsiness is associated with specific physiological changes that can be detected and analyzed. Key physiological signals used in this context include electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), and electrocardiography (ECG). Although these techniques are often considered intrusive, they are widely recognized for their reliability and accuracy, as they provide direct insight into the individual's internal state. To improve the accuracy of drowsiness detection systems, these physiological indicators can be applied individually or in combination. Such methods are extensively used in clinical research, therapeutic settings, and in the development of technologies aimed at enhancing safety and performance in critical fields such as machine operation and driving.

2.3.1.1 Electroencephalography (EEG)

Electroencephalogram (EEG) signals are a powerful tool for real-time monitoring of neural activity with millisecond accuracy. EEG signals are categorized into frequency bands delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (>30 Hz) each associated with specific cognitive processes and brain states. researchers can analyze these signals to detect changes in brain activity. For practical applications, EEG monitoring often involves wearable devices, such as specialized headsets or caps equipped with electrodes placed on the scalp. These devices, like the Emotiv EPOC or NeuroSky MindWave, provide a noninvasive way to monitor brain activity in real time, making them suitable for use in environments such as driving, aviation, or industrial workplaces. However, EEG based systems have limitations, including susceptibility to noise from muscle movements, environmental interference, and the need for precise electrode placement, which can affect accuracy and user comfort.(Figure 2.2).

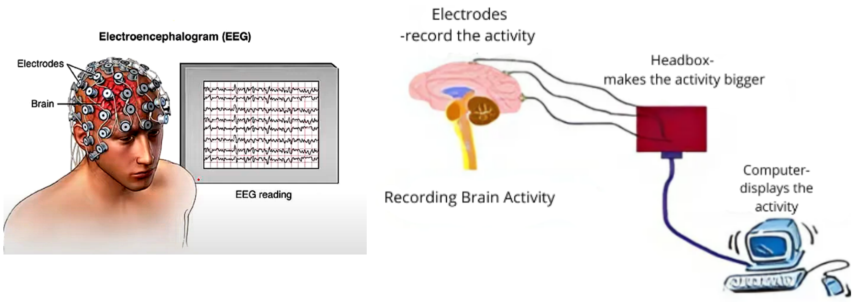


Figure 2.2 – EEG Signal Acquisition for Brain Activity Monitoring

2.3.1.2 Electrocardiography (ECG)

Electrocardiography (ECG) is a widely used, non-invasive technique for assessing the heart's electrical activity. It involves placing electrodes on specific areas of the skin, usually on the chest, arms, and legs, to capture and record the heart's electrical signals during contraction and relaxation. The resulting electrocardiogram (ECG waveform) comprises distinct components, including the P wave, QRS complex, and T wave, each representing different stages of the cardiac cycle. Electrocardiogram signals play a crucial role in detecting sleepiness by analyzing heart rate variability (HRV), which reflects the activity of the autonomic nervous system. The low frequency (LF) band indicates sympathetic influence, while the high frequency (HF) band represents parasympathetic activity. A high HRV indicates sleepiness due to parasympathetic dominance, while fatigue results from the balance between sleep (parasympathetic activation) and staying awake (sympathetic activation). Finally, stress states correspond with higher LF levels caused mainly by sympathetic activation while the subject is awake (Figure 2.3) [3]. However, challenges such as motion artifacts, skin contact variability, and individual heart rate differences can affect accuracy and reliability.

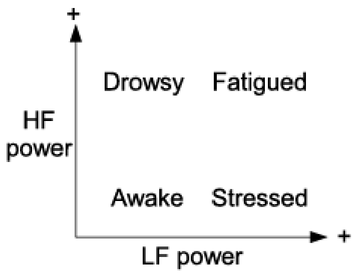


Figure 2.3 – Relationship Between LF/HF Power and Cognitive States

2.3.1.3 Electrooculography (EOG)

The skin surrounding the eyes can be used to measure the EOG, an electrical signal produced by the polarisation of the eye ball. Its size fluctuates according to how far the eyeball moves from its resting position.EOG features slow eye movements (SEM) and rapid eye movements (REM) [4],which happen when a person is awake , to estimate vigilance changes during a monotonous task . Ag/AgCI electrophysiology electrodes are placed around the eyes to obtain the EOG signal. For analysis, two bipolar EOG signal channels—the horizontal and vertical channels—are obtained. Whereas the vertical channel EOG shows vertical eye movements, the horizontal channel EOG shows horizontal eye movements.To measure vertical EOG, two disposable Ag/AgCI electrodes were positioned above and below the right eye. (Figure 2.4).

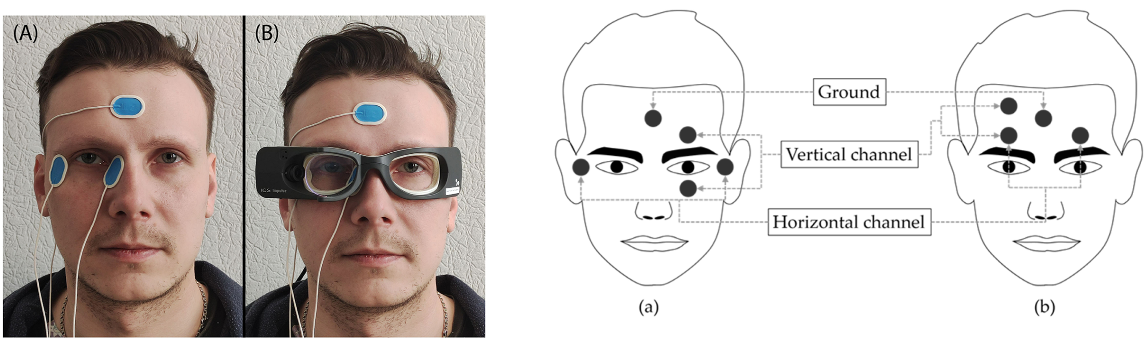


Figure 2.4 – EOG-Based Drowsiness Detection Using Electrodes and Wearable Glasses

2.3.1.4 Electromyography (EMG)

The diagnostic method for determining the electrical activity of skeletal muscles is electromyography (EMG). EMG measures electrical signals associated with muscle activity through surface electrodes placed on the driver's skin. It provides useful data on muscle function, such as force production, recruitment order, and muscle activation patterns. By assessing abnormalities in muscle recruitment and activity, it is frequently used in clinical practice to diagnose neuromuscular disorders, including neuropathies, myopathies, and motor neuron diseases [5]. EMG is used in research to examine muscle biomechanics, motor control, and movement disorders in addition to its diagnostic uses.The main drawback of using EMG signals is its random and complex nature. Furthermore, the collected signals may change due to the structural and biological properties of the muscle. [6]

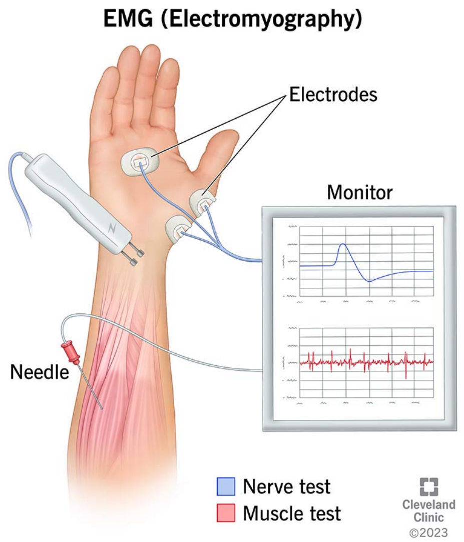


Figure 2.5 – Electromyography (EMG) System for Muscle and Nerve Testing

2.3.2 Behavioral Measures

In behavioral measures of drowsiness, observable behaviors that might be signs of tiredness are observed. These include facial expressions, eye closure, PERCLOSE (percentage of eyelid closure), eye blinking, yawning, and changes in head or eye movement. Behavioral metrics can be used in conjunction with physiological data to provide real-time insights regarding sleepiness levels. Their interpretation, nevertheless, can vary depending on the situation nature of the person . Behavioral Actions Numerous researches have employed behavioral assessments based on facial expressions, eye blinks, eye closure, and frequent yawning .With the use of recorded photos or videos, numerous machine learning algorithms have been crucial in detecting driver drowsiness in recent years. Researchers have used a variety of lightweight algorithms in an attempt to improve accuracy, decrease execution time, and lower cost.[7]

2.3.2.1 Facial Expressions

Drowsiness detection involves monitoring facial features that reflect a driver's level of alertness. Changes such as drooping eyelids, reduced blinking, and slackened jaw muscles are common indicators of fatigue. Eye states—open or closed—along with lip and facial movements, can provide crucial information about the driver's condition. A system utilizing a dataset of 2,904 classified images of facial, eye, and lip expressions helps detect these signs. By tracking mouth-to-eye ratios and feature point positions in video sequences(Figure 2.6), the system can assess drowsiness over time. When signs of sleepiness persist, a warning is issued to alert the driver. Tools such as facial profile detectors, the Dlib package, and the Local Binary Patterns Histogram (LBPH) method are used to enhance facial recognition accuracy. Monitoring these facial cues enables the system to identify fatigue and promote road safety by preventing drowsy driving incidents [8].

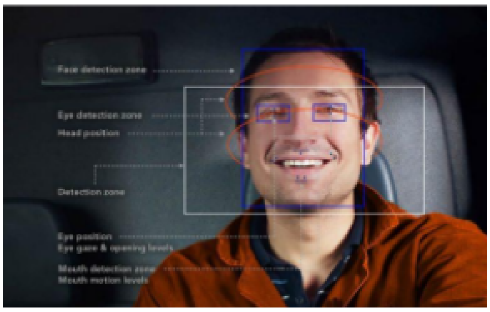


Figure 2.6 – Facial Expression Mapping for Emotion Recognition

2.3.2.2 Yawning Detection

Yawning is a well-known reflexive action involving opening the mouth wide and taking a deep breath, often linked to drowsiness, fatigue, or the need for more oxygen. Despite decades of research, its exact mechanism remains unclear, though theories suggest it may be related to an intrinsic release mechanism or empathy. Current yawning detection techniques rely on static images and lack temporal elements, making them less effective in distinguishing yawns from similar facial expressions. While yawning may help increase alertness, its biological basis is likely tied to changes in physiological states.(Figure 2.7).

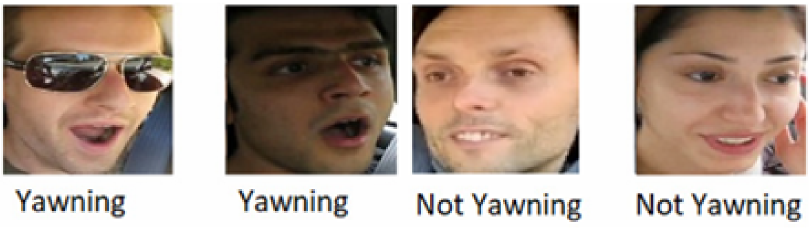


Figure 2.7 – Yawning Detection Using Facial Image Analysis

2.3.2.3 Eye-Based Measures

The eye- based measures categorized into the following categories:

A. Eye Blink Rate

The quick closure and opening of the eyelids is referred to as eyeblinking.In all eight investigations using blink category data, there was a positive correlation between mental weariness and the blink count, blink frequency, and normalised blink ratio[9].eye-blinks are characterized by positive deflection in the most anterior electrodes with a rapid amplitude fall-off at more posteriorly located electrodes.Traditionally, these artifacts are explained by an upward rotation of the eyeball during the lid closure, with the eyeball representing a dipole with corneal positivity and retinal negativity [10].

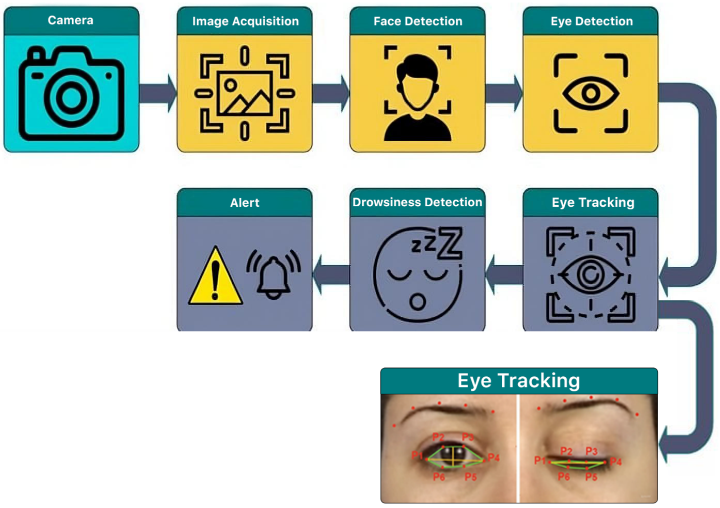


Figure 2.8 – Flowchart of Eye Tracking for Drowsiness and Attention Monitoring

B. Eye Closure

In order to determine a person's state of alertness or exhaustion, eye closure is a crucial signal utilised in drowsiness detection systems. Changes in eye closure behaviour, such as longer eyelid closure times or more frequent blinking, can indicate a decrease in attentiveness in sleepiness detection. Eye-tracking technology, which makes it possible to record and quantify eye closure events(Figure 2.9), is frequently used to analyse these patterns.Systems can objectively assess a person's degree of tiredness and provide prompt alarms or interventions to avoid potential mishaps brought on by diminished attentiveness by integrating eye closure data into drowsiness detection algorithms. An important part of improving the precision and efficacy of sleepiness detection systems is eye closure analysis.

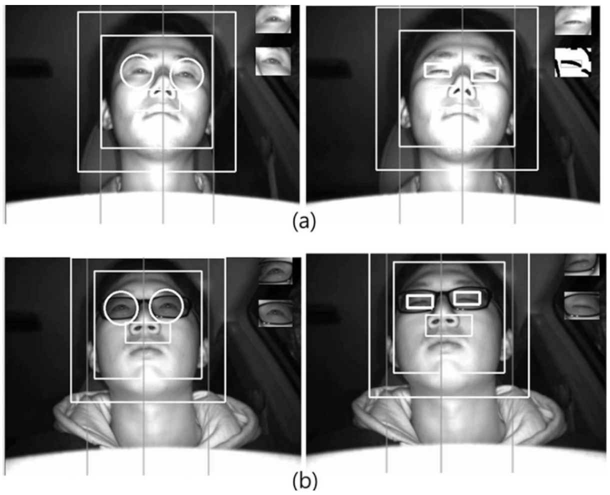


Figure 2.9 – Infrared-Based Eye with and without Glasses

C. PERCLOS (Percentage of Eye Closure)

The Percentage of Eye Closure (PERCLOS) measures drowsiness by calculating how often someone's eyes are at least 80% closed. In a study, twelve professional drivers (around 45 years old) completed two driving simulations: one after a normal night's sleep and another after 24 hours without sleep. Researchers tracked slow eye closure during the task and found that sleep-deprived drivers had significantly more eyelid closure, greater lane position variation, and more attention lapses compared to when they were well-rested. The results showed a strong link between eye closure and both attention lapses ($r = 0.68$) and lane position variability ($r = 0.61$). This method proved to be a reliable way to detect drowsiness, outperforming the drivers' own ratings of how sleepy they felt. It was especially effective in predicting lapses within the first 22 hours of waking time, making it a valuable tool to assess alertness in critical tasks like driving [11].

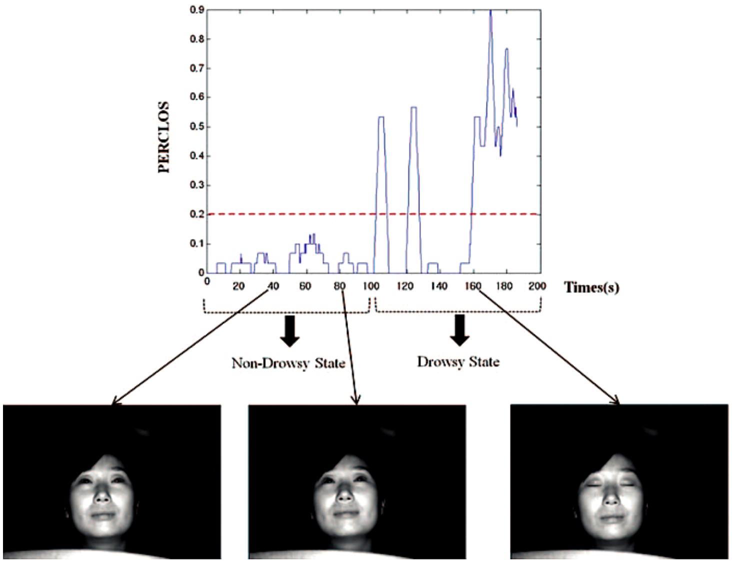


Figure 2.10 – PERCLOS Graph for Drowsiness Detection with Sample Face Images

2.3.2.4 Head Position Changes with eye movements

Head Position Changes with eye movements are frequently employed as objective indicators of drowsiness. To determine a person's state of alertness or drowsiness, these metrics track changes in patterns of head movements, like nodding or swaying, and eye movements, like blink rate, length, and eyelid closure. For example, drowsiness may be indicated by a slower eye blink rate or an increase in head nodding, as these behaviors are frequently linked to increased weariness and decreased attention(Figure 2.11). Quantitative information can be obtained to objectively assess sleepiness levels in real time by analyzing these motions using technology such as motion sensors or eye tracking systems. In addition to subjective evaluations and physiological markers, these objective measurements based on head and eye movements provide important information about a person's level of awareness .

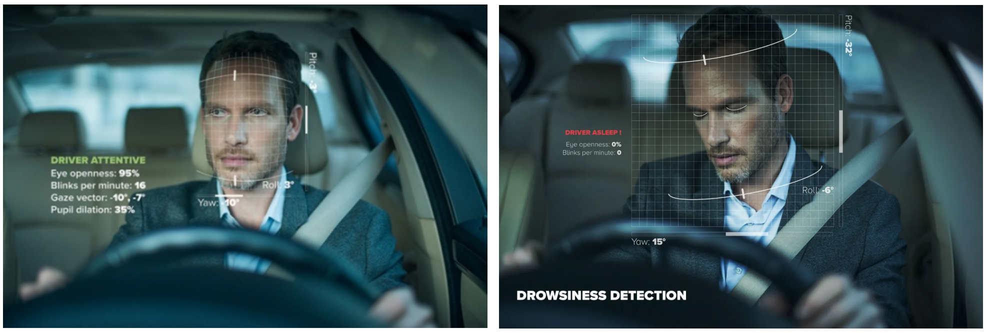


Figure 2.11 – Real-Time Driver Head Position Monitoring for Attention and Drowsiness Detection

2.3.3 Vehicle-Based Measures

Although detection of drowsiness can be based on physiological measures and Behavioral Measures, an alternative source of information is available from the vehicle system. For example, sleep deprivation can produce greater variability in driving speed and lane departures quite likely. Another important vehicle-based measure is steering wheel movement (SWM). When indicators of tiredness are identified, these vehicle-based measures can give the driver real-time feedback or alerts. They are frequently incorporated into advanced driver-assistance systems (Advanced Driver-Assistance System, or ADAS). Additionally, some systems might suggest rest periods or pauses to avoid collisions brought on by sleepy driving.

2.3.3.1 Lane Position Deviation

Driving behavior is significantly influenced by road alignment, especially by its cross-sectional elements such as shoulder width, lane width, and lane position. The influence of shoulder width, lane position, and lane width on driving behavior has been the subject of many researches. Studies have shown that the difference between the left and right tunnel wall is statistically significant on driving behavior[12].Driver inattention, distraction, drowsiness, or impairment are frequently linked to lane position deviation. Reduced attentiveness and a higher chance of lane departure or collision may be indicated by an increase in lane position deviation. Using sensors or cameras mounted on vehicles to monitor lateral lane deviation can help identify changes in driving behaviour associated with tiredness and initiate prompt interventions to avoid collisions.(Figure 2.12).

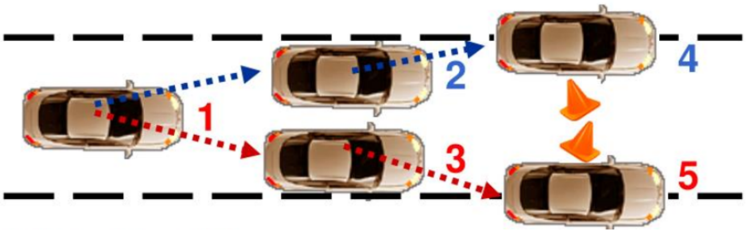


Figure 2.12 – Vehicle Path Deviation Scenarios in Lane Departure Analysis

2.3.3.2 Brake and Acceleration Response

Brake response time (RT) is a driving behaviour metric that has drawn the attention of researchers and is crucial for both road design and the accident lawsuit process. Brake RT is used, among other things, to evaluate stopping sight distance, which determines the level of alertness and attention of the driver. The legal process in accident litigation frequently tries to ascertain whether the participating driver responded to the approaching collision in a time that was "acceptable." Acceptability is defined as a specific percentile of the RT distribution that is believed to represent the driver population (or a pertinent fraction of it) under the relevant circumstances[13].The changes in a car's speed over time, as determined by acceleration and deceleration rates, are referred to as acceleration in drowsiness detection. These modifications can reveal information about driving behaviour since they represent the driver's control inputs, such as braking and throttle. Since tired drivers frequently exhibit delayed reactions, uneven speed control, or decreased acceleration rates, sudden accelerations or decelerations may be a sign of tiredness, distraction, or impairment. Drowsiness detection systems can detect unusual driving behaviour and provide alarms or interventions to improve road safety by examining acceleration trends.

2.3.3.3 Steering Wheel Movements

Using a steering angle sensor to measure steering wheel movements is a popular technique for identifying driver tiredness. The steering behaviour of the driver is recorded by this sensor, which is fixed to the steering column. Studies like Furlough and Graham's have demonstrated that when drivers are sleep deprived, they make less minor steering changes (between 0.5° and 5°) than when they are not driving. These minor motions are examined to determine tiredness and are essential for preserving lane position. Corrective steering actions were triggered in simulated tests by elements such as curving roads and side breezes. Although this approach has been used by automakers like as Nissan and Renault, its applicability is restricted in certain situations due to its strong reliance on road geometry and vehicle dynamics. Overall, steering behavior provides valuable insights into drowsiness but works best when combined with other detection systems [14].

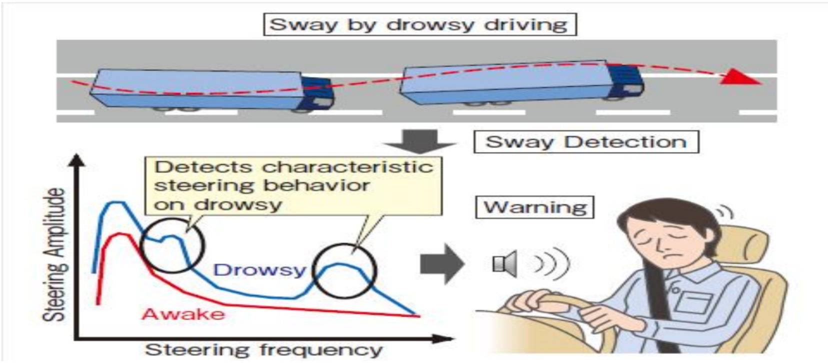


Figure 2.13 – Steering Movement Based Detection

2.3.4 Hybrid Approaches

Hybrid drowsiness detection methods, which combine physiological, behavioral, and vehicle-based signals, offer greater accuracy and reliability than single-signal approaches. They provide a more complete understanding of fatigue levels. However, few studies directly compare hybrid and individual methods, limiting our insight into their relative strengths. For instance, combining EEG and ECG signals can improve detection performance by capturing both brain and heart activity, making these systems more effective in preventing drowsy driving.

2.4 Modern Drowsiness Detection Methods

Modern methods of drowsiness detection use advanced technologies such as artificial intelligence, deep learning, wearable devices, computer vision, and machine learning to analyze facial signals such as eye movements, blinking patterns, and yawning to create effective, non-invasive systems. These systems, which are often integrated into cameras or wearable devices, can detect fatigue with high accuracy, for example (96-99%). Therefore, a study was conducted to evaluate fatigue and drowsiness by applying deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), and combining video-based analysis and classification methods such as random forests and neural networks, and reached an accuracy of up to 99% [15]. These methods are reliable, safe, and often do not require any additional equipment, making them practical for real-world applications such as driving or operating heavy machinery. This is done by alerting users in real time, making them effective solutions that prevent accidents in human life activities.

2.4.1 Deep Learning Approaches

Deep learning, a subset of artificial intelligence, plays a pivotal role in modern drowsiness detection systems by automatically learning patterns from complex data through multiple layers of artificial neural networks. These models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are inspired by the structure of the human brain and excel at identifying subtle signs of fatigue. Trained on large datasets using optimization techniques like backpropagation, these networks can make accurate predictions by analyzing intricate relationships within data. In the context of drowsiness detection, deep learning is often applied through computer vision techniques that monitor facial expressions, eye movements, and head posture in real time. For instance, CNNs are used to extract visual features such as blinking frequency, PERCLOS (percentage of eye closure), yawning, and head tilt from video feeds. Algorithms like Viola-Jones can detect and analyze head positions to infer levels of alertness. Moreover, integrating machine learning into these vision-based systems improves their performance under noisy or variable lighting conditions, making them highly reliable for use in driver monitoring applications. Through the synergy of deep learning and AI-driven vision systems, drowsiness detection has become more precise, enabling proactive interventions to enhance road safety..we take a example of deep learning for driver drowsiness detection in(Figure 2.14).

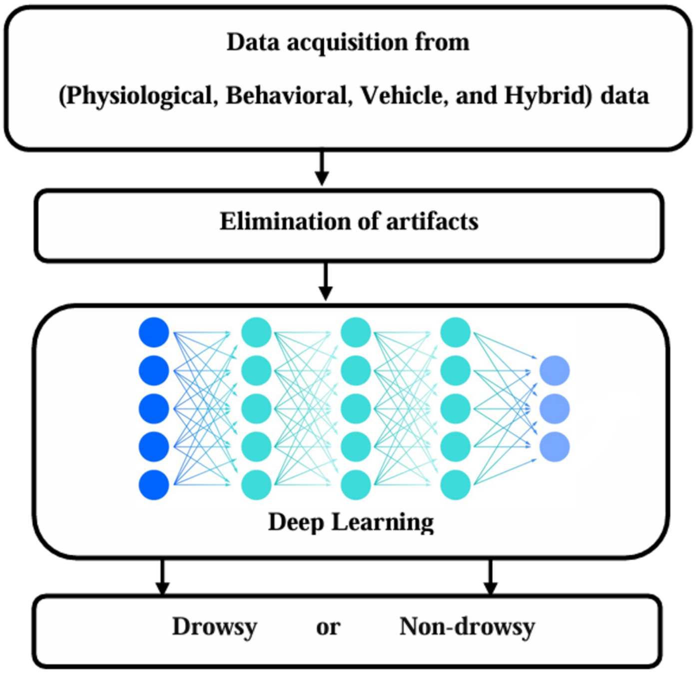


Figure 2.14 – Drowsiness Detection System Using Multimodal Data and Deep Learning

2.5 Comparison Between Traditional and Modern Approaches

- **Traditional Methods:** Traditional drowsiness detection methods, such as physiological, behavioral, and vehicle-based measures, have several limitations. They are often affected by measurement variability, sensitivity to noise, and challenges in real-world application. Physiological and behavioral measures may lack specificity, while vehicle-based indicators can be influenced by external factors like road conditions or driver habits. Additionally, extracting meaningful drowsiness indicators from traditional methods can be difficult due to environmental factors, sensor inconvenience, and driver movements.

- **Modern Methods:** Modern approaches leverage artificial intelligence (AI), deep learning, and advanced sensor technology to enhance accuracy and robustness. These methods improve upon traditional techniques by integrating machine learning models that analyze facial expressions, head posture, and eye movement more reliably, even in challenging conditions like variable lighting or occlusions. However, as seen in the table, modern methods still face limitations, such as difficulty in real-time video analysis, susceptibility to environmental conditions, and challenges in handling diverse facial characteristics,(Table 2.1) summarizes the key differences between traditional and modern approaches:

Aspect	Traditional Methods	Modern Methods
Accuracy	Easily affected by noise and signal instability	Improved noise handling, but still sensitive to poor data quality
Environmental Sensitivity	High impact on performance	More robust, but still affected
Measurement Consistency	Heavily dependent on sensor calibration and expert input	More standardized with automated analysis
Driver Adaptability	Less adaptable to individual differences	More adaptive through model personalization, though not perfect
Real-time Processing	Limited	Enhanced through deep learning algorithms
Implementation Challenges	Sensor inconvenience, driver discomfort	Avoids intrusive sensors, but requires high computation and raises privacy issues

Table 2.1 – Key Differences between Traditional and Modern Methods

- **Key Differences :**

Both approaches have strengths and weaknesses, but modern methods provide more precise and adaptable solutions while still facing implementation challenges. Future advancements aim to combine sensor fusion, AI optimization, and real-world adaptability to create more effective drowsiness detection systems.

2.6 Conclusion

Drowsiness detection methods include behavioral, physiological, and vehicle-based approaches. Behavioral methods, especially facial analysis, are preferred due to their non-intrusive nature and simplicity. With advances in deep learning and computer vision, these methods now offer high accuracy. This project focuses on using deep neural networks for facial-based drowsiness detection.

Chapter 3

Deep Learning for Drowsiness Detection

3.1 Introduction

Artificial intelligence (AI) aims to replicate human thinking and decision-making. It allows machines to learn from data and adapt to their environment. A key branch of AI is machine learning (ML), which enables systems to improve through experience without explicit programming.

As problem complexity grew, deep learning (DL) emerged as a powerful extension of ML, allowing models to automatically extract patterns from large datasets and achieve high performance in many scientific and technological fields. [16].

This chapter examines deep learning, how it differs from traditional machine learning, and how it processes large data to uncover complex patterns with high accuracy.

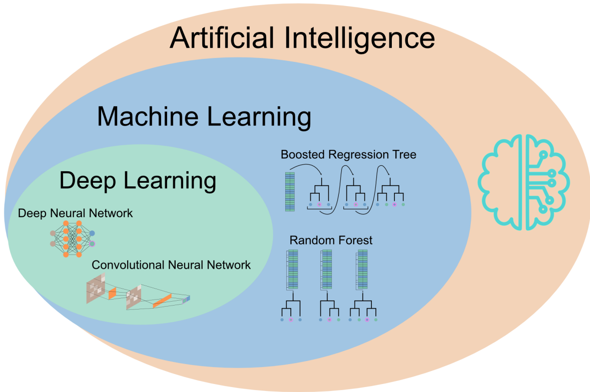


Figure 3.1 – The relationship between AI, ML, and DL

3.2 Deep Learning: Fundamentals and Concepts

3.2.1 What is Deep Learning?

Deep learning is a branch of machine learning that relies on artificial neural networks to automatically learn complex representations from data, without the need for manual feature engineering. These representations are extracted through multiple layers that hierarchically learn patterns directly from raw data. These models excel at processing large volumes of data and uncovering subtle nonlinear relationships. Such patterns include complex dependencies between features, hierarchical data representations, latent hidden characteristics, and intricate feature interactions that simpler models find difficult to interpret. Their widespread success is largely attributed to advances in computational techniques and the availability of large and diverse datasets [17].

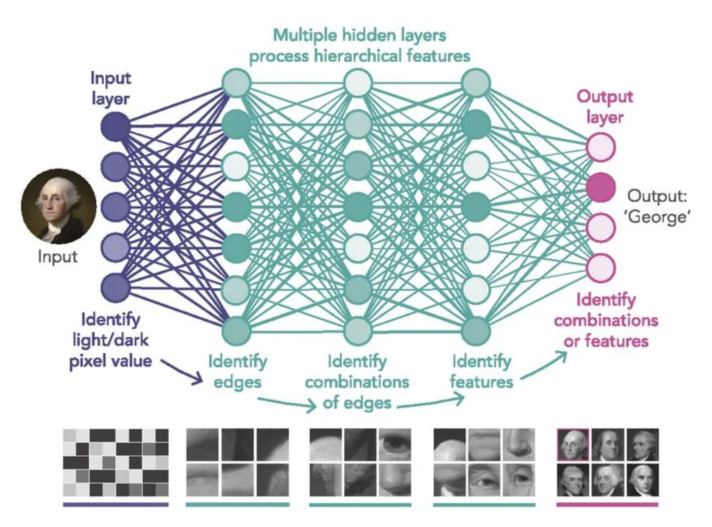


Figure 3.2 – Illustration of deep learning

3.2.2 Evolution from Machine Learning to Deep Learning

As AI advanced, traditional machine learning methods based on statistical models like decision trees and logistic regression faced challenges due to the need for manual feature extraction, especially with complex data such as images and text. Artificial neural networks (ANNs) gained attention but were initially limited by scarce data and low computing power.

The early 21st century brought breakthroughs with GPUs and large datasets, enabling training of deeper models. By 2006, combining support vector machines (SVMs) with computer vision improved automatic pattern recognition significantly.

This progress led to deep neural networks (DNNs) capable of learning directly from raw data with minimal human intervention. DNNs now outperform traditional methods across many fields [18].

Below is a picture that shows more evolution from machine learning to deep learning

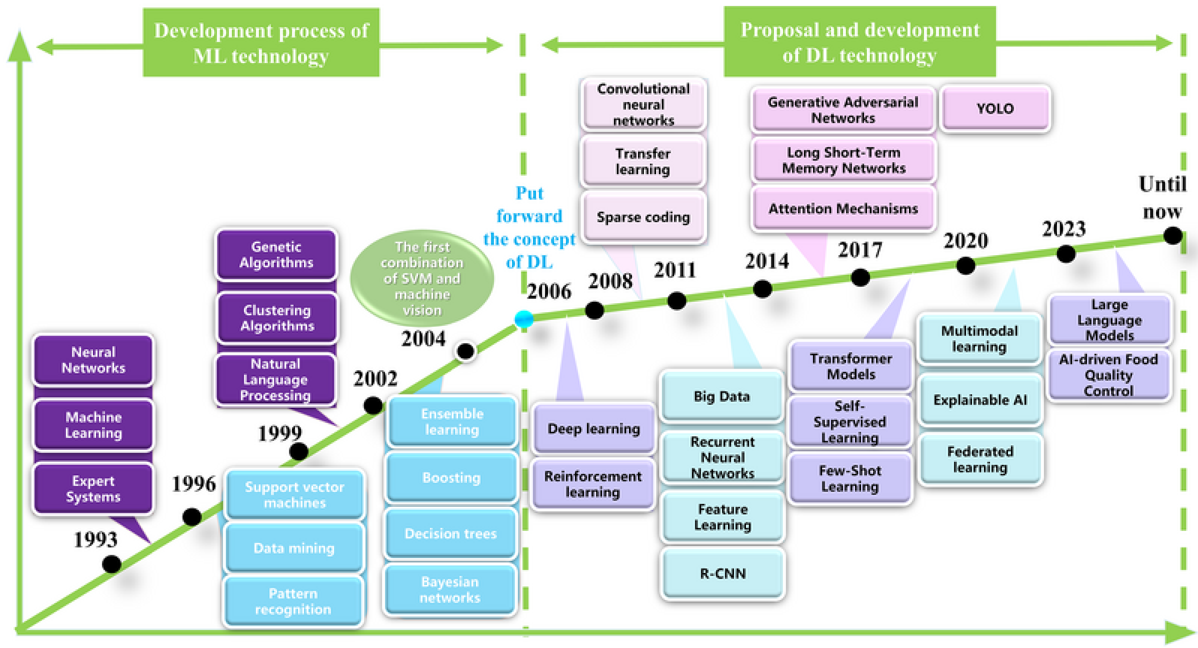


Figure 3.3 – Evolution from Machine Learning to Deep Learning

3.2.3 Machine learning vs deep learning: why does deep learning excels?

Deep learning surpasses traditional machine learning by automatically processing complex data without the need for manual feature engineering. Its deep architecture enables efficient learning from large-scale data, leading to superior performance in areas such as computer vision and natural language processing [19].

In the following table[20], we illustrate the key factors that make it more advanced:

Factors	Machine Learning	Deep Learning
Performance in complex tasks	Ineffective in processing complex data; suitable only for simple to moderately complex tasks	Effective in processing complex data such as image recognition
Accuracy of results	Accuracy depends heavily on the quality of extracted features; usually lower than deep learning	Very high accuracy, especially in image and video processing
Ability to handle big data	Limited; performs poorly with large volumes of unstructured data	Efficient at extracting hidden patterns in big data
Feature extraction	Features must be manually extracted, risking the loss of important information	Features are extracted automatically via neural networks, improving performance
Dealing with new data	Requires resetting features and retraining for new data	Capable of continuous learning and adapting to new data due to deep architecture

Table 3.1 – Comparison Between Machine Learning and Deep Learning

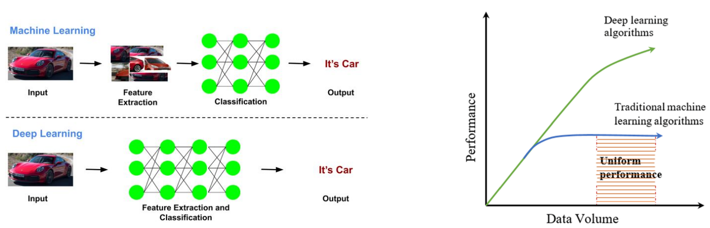


Figure 3.4 – Comparison Between Machine Learning and Deep Learning

Through this comparison, we recognise that deep learning offers superior capability in processing large and unstructured data, granting it a clear advantage over traditional machine learning methods, particularly in automatic feature extraction and performance improvement as data volume increases. With these advantages, we see it as the optimal solution for advanced applications we aspire to develop across various fields.

3.3 How deep learning works?

Deep learning relies on the use of deep artificial neural networks, known as Deep Neural Networks (DNNs), which are computational models consisting of multiple layers of interconnected nodes (neurons). These networks are used to extract high-level representations from raw data. DNNs are an extension of traditional artificial neural networks (ANNs), as they include more than one hidden layer, granting them greater capacity to model complex nonlinear relationships within the data [21].

3.3.1 Basics of neural networks

After moving beyond the biological inspiration, artificial neural networks (ANNs) are primarily defined by their layered structure and internal computational mechanisms, which process data through multiple layers. The typical model of these networks consists of three main layers:

a • Input layer: This layer is responsible for receiving the raw data. For example, if the input is an image, this layer consists of units that represent each pixel of the image. The input layer serves as the entry point for the data to be processed in the subsequent layers [22].

b • Hidden layers: The hidden layers play a crucial role in enabling the neural network to learn complex patterns.The process starts with the calculation of the weighted sum, where each input is multiplied by its corresponding weight, and the bias term is added. This adjustment helps modify the activation threshold, allowing the model to learn more intricate patterns. After this transformation, the result is passed through a non-linear activation function such as ReLU, Sigmoid, or Tanh, which introduces the necessary non-linearity to handle complex relationships within the data [21].

c • Output layer: This layer is responsible for producing the final output after the analysis of data in the hidden layers. The output could be a classification or a numerical value, depending on the specific task being performed [22].

Sewak and colleagues (2020) indicated that the performance of these models is influenced by several factors, including the number of layers, the type of activation functions used, as well as the method adopted for weight initialization and data normalization. The strength of DNNs lies in their ability to automatically learn hierarchical representations of data without the need for manual feature engineering, making them well-suited for various applications such as image classification, biosignal analysis, and pattern recognition [21].

We illustrate in the following figure the fundamental structure of an artificial neural network, including its main layers and the mechanism of data flow for processing information and generating outputs.

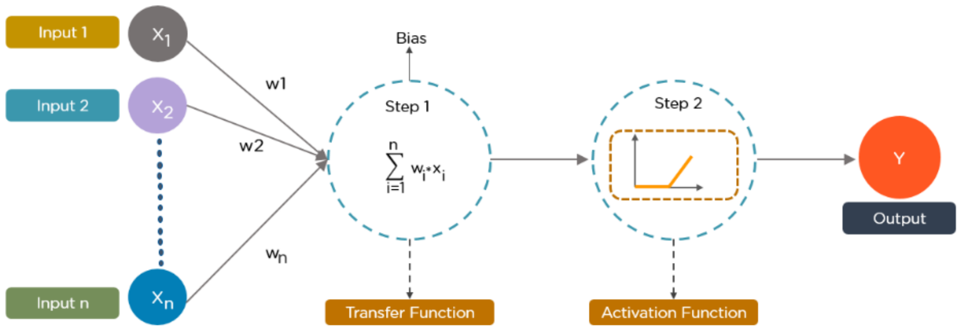


Figure 3.5 – Structure of an Artificial Neural Network

Neural networks rely on diverse training strategies that vary in how they handle data and the extent of prior knowledge available about the desired outputs. In certain scenarios, the model is trained using data that includes reference information, allowing it to fine-tune its internal parameters with precision. In other cases, the model is exposed to data without any prior output information and is required to autonomously identify latent structures and patterns. These training approaches are chosen based on the nature of the task and the availability of resources, and they play a critical role in determining the performance of neural networks across various practical applications [23].

3.4 Convolutional neural networks (CNNs)

The Convolutional Neural Network (CNN) was first introduced by Kunihiko Fukushima and later developed by Yann LeCun, who integrated it with the back-propagation algorithm to recognize handwritten digits and perform document analysis. LeCun's system was ultimately employed for reading handwritten checks and postal codes.

CNNs rely on convolutional and pooling layers. Convolutional layers are responsible for filtering input data to extract useful features. These layers contain learnable parameters, allowing the filters to adjust automatically to extract the most relevant features for a given task. Multiple convolutional layers are used sequentially, each extracting increasingly abstract features from the input. Pooling layers provide limited invariance to translation and rotation, while also reducing memory usage, which enables deeper network architectures [24].

3.4.1 Architecture of CNNs

Convolutional Neural Networks (CNNs) are primarily composed of two main blocks:

• **Feature Extraction Block:** responsible for extracting important features from input data using operations such as convolution and pooling.

• **Classification Block:**responsible for classifying the extracted features into target classes using fully connected layers.

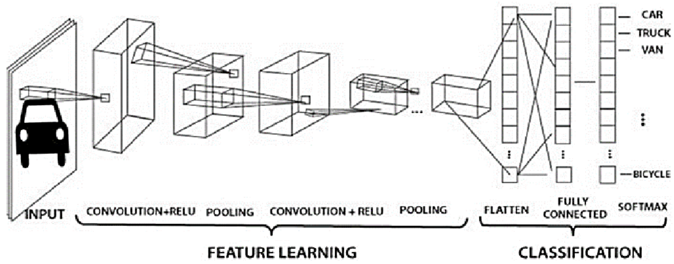


Figure 3.6 – CNN Architecture

The process begins with extracting important features from the data using specialized layers, which include:

a • Convolutional Layer (CONV):

The convolutional layer performs a mathematical operation called convolution on the input data. Convolution is used to determine the degree of overlap between the input data and a filter (or kernel). In CNNs, the kernel slides across the input matrix, performing element-wise multiplication followed by summation to produce transformed feature maps. For example, convolving a 3×3 region from the upper-right corner of an image with a specific kernel might result in a value such as 77 (Figure 3.7) . This operation enables various image processing functions such as edge detection, sharpening, and blurring.

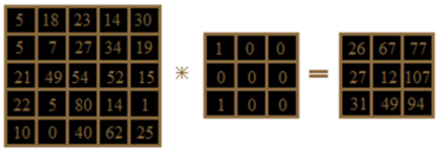
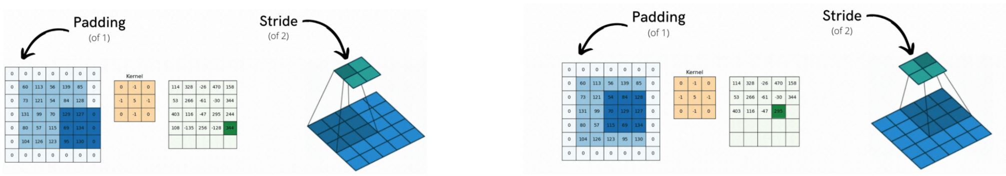


Figure 3.7 – A basic representation of the two-dimensional convolution process

By stacking multiple convolutional layers, CNNs can learn hierarchical representations starting from raw pixels, then progressing to edges, primitive shapes, parts, and eventually entire objects. After each convolution, a non-linear activation function, typically the Rectified Linear Unit (ReLU), is applied to enhance learning by setting negative values in the feature maps to zero [24].

Several key hyperparameters significantly influence the performance of convolutional layers, including [23]:

- **Depth:** Refers to the number of filters used, determining how many distinct features the network can learn.
- **Stride:** Indicates how many pixels the filter moves across the input, affecting the spatial dimensions of the output.
- **Padding:** Zero-padding helps preserve spatial dimensions and retain edge information.



(a) **(b)**

Figure 3.8 – Padding and Stride Mechanisms in Feature Extraction

b • Pooling layer (POOL)

The pooling layer reduces the spatial dimensions of activation maps while retaining critical information. Common pooling strategies include [24]:

- **Max pooling:** Retains the maximum value within each pooling window, helping to extract the most prominent features.
- **Average pooling:** Calculates the average of values within each pooling window, resulting in a smoother representation of the data.

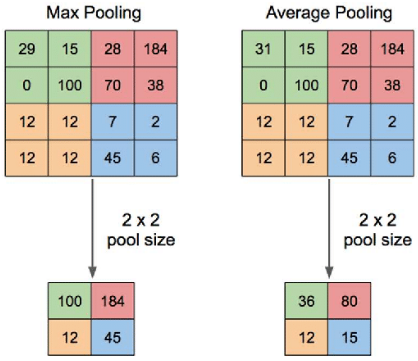


Figure 3.9 – Example of max and average pooling

Pooling enhances the model's invariance to translation and deformation and reduces the number of parameters and computations, improving overall efficiency.

After feature extraction, the processed feature maps are passed to the classification part, which includes:

c • Fully Connected Layer :

vector through flattening. This vector is then fed into fully connected layers, where each neuron connects to every neuron in the next layer. These layers analyze and combine learned features to generate final predictions. The output of the final fully connected layer typically passes through a softmax activation function, converting raw outputs into probability distributions over target classes [24].

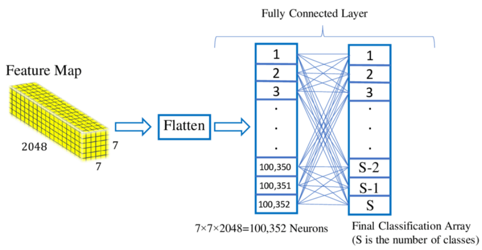


Figure 3.10 – Fully Connected layers

This hierarchical structure, characterized by successive convolutional, pooling, and fully connected layers, enables CNNs to progressively extract and integrate visual features. Owing to this organized and multi-layered design, CNNs achieve high accuracy and efficiency in a wide range of intelligent visual applications, including facial recognition and autonomous driving.

3.4.2 Activation Functions in CNNs

Activation functions are essential components in Convolutional Neural Networks (CNNs), as they introduce non-linearity into the model. This non-linearity enables the network to learn and approximate complex patterns and relationships in data that linear transformations alone cannot capture [25].

Below are the most commonly used activation functions in deep learning, each with unique characteristics and applications:

3.4.2.1 Sigmoid function

Also known as the logistic function, it transforms any real-valued input into a range between 0 and 1. It is commonly used to convert model outputs into probabilities [25]. It is mathematically defined as:

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.3}$$

Its curve resembles the letter S, allowing it to smoothly and consistently represent probabilistic outputs.

3.4.2.2 Tanh (Hyperbolic Tangent)

The hyperbolic tangent function, or tanh, is similar to the sigmoid but outputs values in the range $[-1,1]$ [25]. It is defined as:

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2.4}$$

This function is also S-shaped but centered around zero, which allows it to produce outputs with a mean closer to zero, often resulting in better convergence during training compared to the sigmoid.

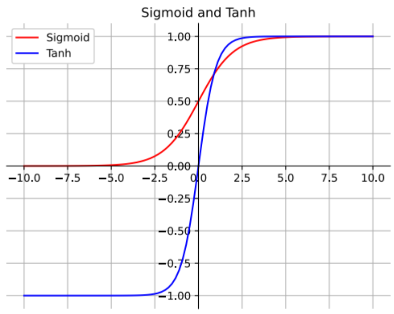


Figure 3.11 – Graphical Sigmoid and Tanh Activation Functions

3.4.2.3 ReLU (Rectified Linear Unit)

The ReLU function has become the most widely used activation function in modern deep learning models. It introduces non-linearity by converting all negative input values to zero while allowing positive values to pass through unchanged [25]. It is defined as:

$$f(x) = \max(0, x) \tag{2.5}$$

The figure below illustrates the behavior of the ReLU function

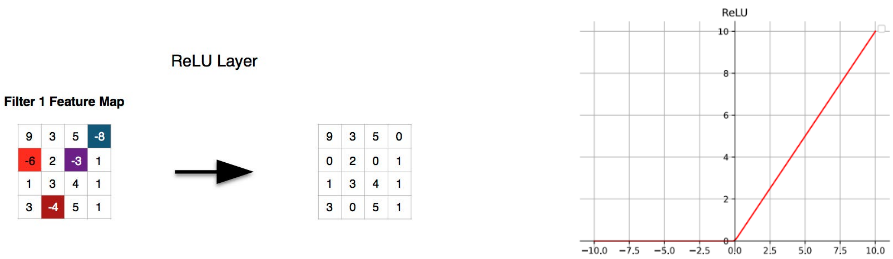


Figure 3.12 – Example and Curve ReLU Function

3.4.2.4 Softmax function

The softmax function is primarily used in the output layer of neural networks for multi-class classification problems. It converts raw output scores (logits) into a probability distribution over classes, where the sum of all probabilities is equal to 1.

The softmax function highlights the largest values in the input vector while suppressing the smaller ones, making it easier to interpret the most likely class [25]. The diagrams below demonstrate how softmax transforms a set of raw scores into a normalized probability distribution:

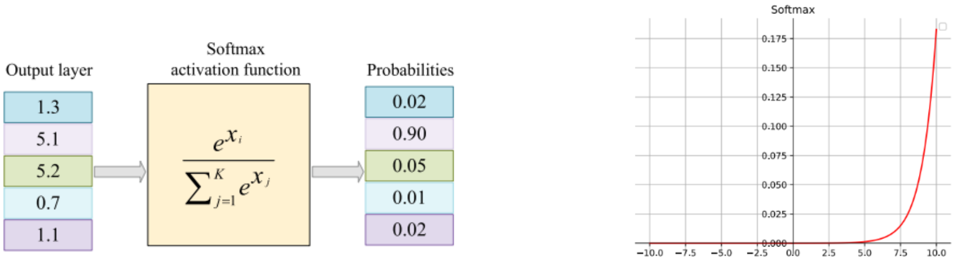


Figure 3.13 – Softmax Function

3.5 Transfer Learning

3.5.1 Concept of Transfer Learning

Transfer learning is an advanced technique in artificial intelligence that involves transferring knowledge and skills acquired from a source task to a related target task. This process resembles how humans leverage prior experiences to accelerate the acquisition of new skills rather than starting from scratch, individuals draw upon previously acquired knowledge to enhance learning speed and efficiency .

According to Pan (2020), transfer learning makes use of neural architectures that have already been developed using large and diverse datasets. These knowledge-rich models reduce the need for extensive labeled data specific to the new task, which is especially valuable in data-scarce scenarios [26].

The mechanism of transfer learning operates through several key strategies, including:

- **Using a pretrained model as a fixed feature extractor:** In this method, a new linear classifier is added as the final layer (classifier layer) and trained only on the new dataset, while the feature extraction layers remain fixed and unchanged. This approach is effective when the new dataset is small but similar to the original training data.

- **Fine-tuning the entire model:** In this approach, the classifier layer is replaced by a new fully connected layer, and the entire network is retrained using the new dataset. This includes extending the backpropagation process to the upper layers, allowing all weights to be adjusted to better fit the new task [27].

3.5.2 Benefits of Transfer learning

Neural models used in transfer learning rely on diverse datasets, providing generalizable patterns that can be adapted to new tasks. One major advantage is faster training convergence, as prior knowledge offers a strong initialization, reducing time and computational cost especially when labeled data is scarce [27] [26]. These models also improve performance and generalization, often achieving higher accuracy compared to models trained from scratch. Furthermore, they enable effective learning with fewer labeled examples, making them suitable for resource-limited domains[26]. Finally, their reusable architecture allows easy adaptation to multiple tasks without rebuilding from scratch, lowering development costs and accelerating deployment[27].

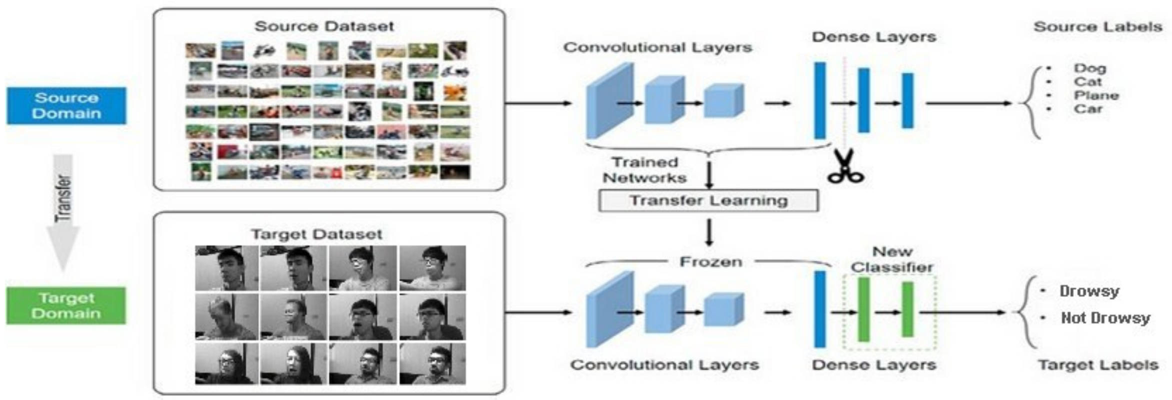


Figure 3.14 – Transfer Learning for Drowsiness Detection

3.6 Popular CNN Architectures in Drowsiness Detection

Convolutional Neural Networks (CNNs) include a wide range of architectures, with the most common ones being:

3.6.1 ResNet(2015)

Residual Network (ResNet), developed by Kaiming He et al., won the ILSVRC 2015 competition and marked a major breakthrough in deep neural network design. It is distinguished by the use of skip connections that allow certain layers to be bypassed, which helps mitigate the vanishing gradient problem during training and enables the construction of very deep networks. Well-known variants include ResNet-50, ResNet-101, and

ResNet-152, with 50, 101, and 152 layers respectively. The architecture also incorporates techniques such as Batch Normalization and Global Average Pooling to enhance training efficiency and reduce computational complexity [24].

3.6.2 DenseNet (2017)

Densely Connected Convolutional Network (DenseNet), introduced by Gao Huang et al., received the Best Paper Award at CVPR 2017. It features a unique architecture in which each layer is directly connected to all preceding layers in a feed-forward fashion, encouraging feature reuse and facilitating gradient flow. This design enhances learning efficiency and helps to alleviate the vanishing gradient problem. DenseNet has proven effective on several challenging datasets such as CIFAR-10, CIFAR-100, SVHN, and ImageNet, showing significant improvements over previous state-of-the-art architectures in object recognition tasks [24]. Common variants include DenseNet-121, DenseNet-169, and DenseNet-201, with networks comprising between 121 and 201 layers, incorporating interleaved convolutional layers, skip connections, pooling, normalization, and fully connected layers.

3.6.3 EfficientNet (2019)

EfficientNet, introduced in 2019, is an advanced convolutional neural network model known for achieving an effective balance between classification accuracy and computational efficiency. The model is based on a compound scaling strategy that systematically and simultaneously increases the network's depth, width, and input image resolution, leading to significant performance improvements without excessive growth in model size or computational requirements. EfficientNet is highly resource-efficient, allowing it to achieve high accuracy while reducing energy consumption and processing time compared to larger and more complex models [28]. Its most well-known variants include EfficientNet-B0 through EfficientNet-B7, as well as the improved version EfficientNetV2. Typically, the architecture consists of around 10 to 30 layers, including convolutional, pooling, normalization, and fully connected layers, along with advanced activation functions. These characteristics make it suitable for lightweight deep learning applications that demand high performance in image classification and recognition.

3.7 Conclusion

In this chapter, we have examined the theoretical foundations of deep learning, from the core principles of neural networks to their advanced applications. We placed particular emphasis on the analytical structure of Convolutional Neural Networks (CNNs), meticulously deconstructing their complex layers and precise operational mechanisms. Furthermore, we explored the transfer learning methodology adopted in our research. Our discussion also included an analysis of leading architectural models in this field. In the next chapter, we will transition to practical implementation, presenting and closely examining the experimental results obtained.

Chapter 4

Results and discussions

4.1 Introduction

This chapter presents the experimental results of our deep learning-based approach to detecting driver drowsiness. The study focuses on evaluating the effectiveness of several advanced convolutional neural network (CNN) architectures, specifically MobileNet, ResNet50, DenseNet201, and EfficientNetV2. These models were selected for their strong performance in image-based classification tasks and applied to the dataset using a transfer learning approach. The trials compared each CNN's performance in identifying signs of drowsiness, offering insights into their suitability for real-time driver monitoring systems.

4.2 Proposed Framework

The proposed framework(Figure 4.1) for drowsiness detection consists of four main stages: dataset DDD used ,dataset preparation, feature extraction, and classification. Initially, facial images are collected to build the dataset, which includes samples labeled as either drowsy or alert. During the preprocessing stage, the images are resized, rotated to introduce variation, and augmented to increase dataset diversity and improve model performance. Next, deep features are extracted using transfer learning, aided by pseudo-labeling, to utilize both labeled and unlabeled data effectively. These features are then used to train a classification model capable of distinguishing between drowsy and non-drowsy states. Finally, the system outputs a binary classification indicating whether the subject is drowsy or alert.

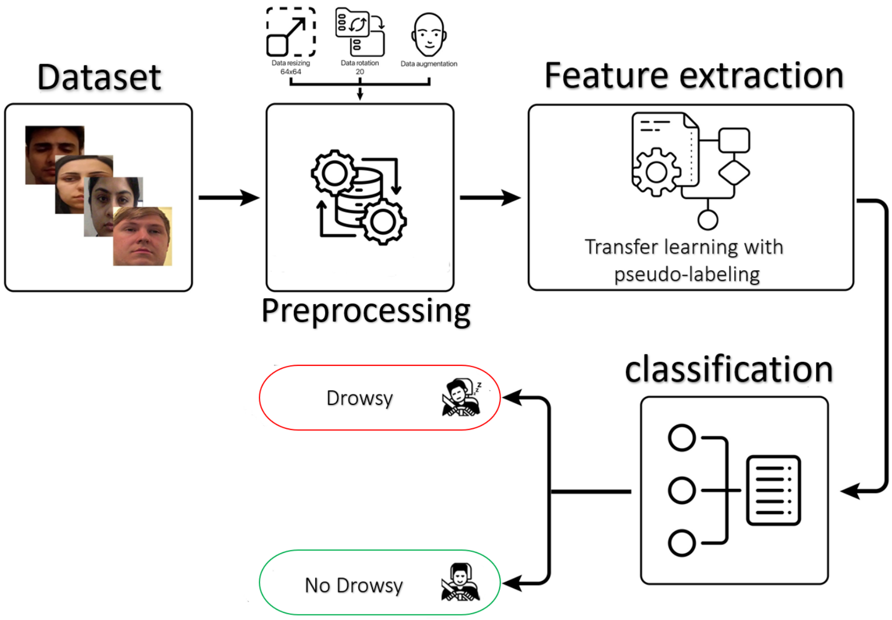


Figure 4.1 – An overview of the proposed drowsiness detection method

4.2.1 Dataset Description

The UTA-RLDD (University of Texas at Arlington Real-Life Drowsiness Dataset) was developed for multi-stage drowsiness detection, aiming to identify not only obvious signs but also subtle micro-expressions linked to early drowsiness. These micro-expressions are instinctive and difficult to fake, making the dataset more realistic and valuable. It includes approximately 30 hours of RGB video data from 60 participants—each contributing one video for three states: alertness, low vigilance, and drowsiness, totaling 180 videos (Figure 4.2). The participants, aged 20 to 59 and from diverse ethnic backgrounds, recorded themselves using their cell phones or webcams in real-life environments at a frame rate of under 30 fps, mimicking common camera conditions. This makes UTA-RLDD the largest and one of the most realistic datasets available for early drowsiness detection [29].

The Kaggle version focuses on a more compact and practical image-based representation. In this dataset, facial images were extracted from the original videos using VLC Media Player, followed by face detection and cropping using the Viola-Jones algorithm to isolate the region of interest (ROI). This preprocessing step ensured that only relevant facial features were retained for training the model. The resulting data set contains over 41,790 RGB images, uniformly resized to 227×227 pixels and organized into three folders:

- drowsy : images labeled as exhibiting signs of drowsiness.
- no drowsy : images labeled as showing alertness(Figure 4.3)[30].

To simplify the classification task, we adopted a binary labeling strategy, where:

• Label 0 (Drowsy) corresponds to images in the drowsy/ folder.

• Label 1 (No Drowsy) corresponds to images in the no drowsy/ folder.

Figure 4.2 – Sample frames from the UTA-RLDD dataset in the alert (first row), low vigilant (second row) and drowsy (third row) states

Figure 4.3 – Sample frames from the DDD dataset (extracted from UTA-RLDD) in the alert (first row) and drowsy (second row) states

4.2.2 Face Preparation

Prior to training, all facial images underwent several preprocessing steps to ensure compatibility with the input requirements of the deep learning models and to improve overall performance. The dataset was divided into four subsets: 12,549 images for training, 8,367 for validation, 8,367 for testing, and 12,550 images as an unlabeled set (Figure 4.4), which can be used for further experiments, such as semi-supervised learning. Class labels were balanced across all subsets to maintain an even distribution of drowsy and non-drowsy samples.

To prepare the images for model input, all images were resized either to 224×224 pixels or 64×64 pixels, depending on the specific model architecture used. Higher-resolution inputs (224×224) were used with deeper models, while lighter models or experimental configurations used the smaller 64×64 size for faster computation and reduced complexity.

Since the original image data was in BGR format (as read by OpenCV), a color space conversion to RGB was applied to meet the requirements of deep learning frameworks such as PyTorch and TensorFlow, which expect RGB inputs. Additionally, data augmentation was applied to improve model robustness. In particular, a random rotation of up to 20 degrees was introduced to simulate natural variations in head pose and facial orientation, which are common in real-life driving scenarios.

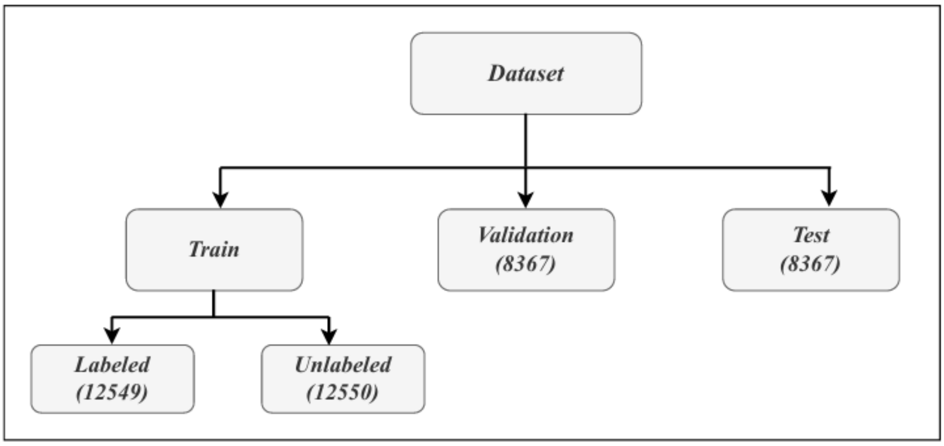


Figure 4.4 – split of the dataset

4.2.3 Feature Extraction and Classification Using Transfer Learning (TL)

To support the development of intelligent transportation systems and enhance road safety, this study proposes a comprehensive framework for driver drowsiness detection using transfer learning techniques. The framework relies on four advanced deep models—DenseNet201, ResNet50, EfficientNetV2, and MobileNet and employs a pseudo-labeling mechanism to improve training efficiency. The process begins with face detection and data preprocessing. Then, pseudo-labeling is employed to support both feature extraction and classification stages, leading to improved model performance.

The feature extraction phase benefits from the pre-trained weights of each model, where most of the feature layers are initially frozen to preserve general visual knowledge learned from large-scale datasets such as ImageNet. Pre-processed facial images are passed through the deep feature layers of the models to extract rich visual representations that highlight drowsiness-related facial cues such as eyelid closure, yawning, or head tilt.

- **In DenseNet201**, the pretrained convolutional layers remain frozen and are used to extract high-level visual features that are directly transferred to the classification head.
- **In ResNet50**, the entire model is fine-tuned, allowing its feature layers to adapt to the domain-specific characteristics of drowsiness detection.
- **In EfficientNetV2-S and MobileNetV2**, full fine-tuning is also applied, allowing the feature extraction and classification layers to be updated for optimal performance.

Model	Strategy	Classification Head	Dropout	The Train
ResNet 50	Full fine-tuning	Dense(512) Dropout(0.3) Dense(2)	0.3	Full model
MobileNet	Full fine-tuning	Dense(512) Dropout(0.2) Dense(2)	0.2	Full model
DenseNet201	Frozen pretrained layers	Dense(1024) Dense(512) Dense(2)	None	Head only
EfficientNetV2	Full fine-tuning	Dropout(0.3) Dense(2)	0.3	Full model

Table 4.1 – Model Configurations and Training Parameters

A key innovation of this framework is the integration of pseudo-labeling directly into the feature learning process. After an initial training round using labeled data, the model is used to generate predictions on the unlabeled samples. Only those predictions with high confidence are selected and treated as pseudo-labeled data, allowing the model to self-generate additional supervision. These pseudo-labeled samples are then passed through the feature extraction pipeline again, enriching the training distribution and enabling the model to learn from a broader dataset without manual annotation.

In the classification stage, the extracted feature maps are passed to a custom classification head tailored for binary classification (alert vs. drowsy). This head consists of fully connected (dense) layers activated by ReLU functions and regularized with dropout layers to prevent overfitting. In the end, a Softmax activation function is used to output class probabilities.

Pseudo-labeling also plays a critical role in this classification phase. As the model generates confident predictions for unlabeled data, these pseudo-labeled examples are added to the training set, helping the classifier generalize beyond the limited labeled dataset.

To improve the model's robustness, data augmentation techniques are applied during preprocessing. These include resizing facial images to 64×64 pixels, applying random rotations of ± 20 degrees to simulate natural head movements, and normalizing pixel intensities using the ImageNet mean and standard deviation.

Parameter	Value
Image size	64 × 64 pixels
Rotation range	±20 degrees
Normalization	ImageNet mean
Optimizer	Adam
Learning rate	0.001
Batch size	128
Framework	PyTorch Lightning

Table 4.2 – Preprocessing and Training Setup

Training begins with a small labeled dataset using the Adam optimizer (learning rate 0.001, batch size 128) within the PyTorch Lightning framework. After each training iteration, the model evaluates the unlabeled dataset and generates soft predictions. Samples that exceed a confidence threshold are assigned pseudo-labels and added to the training set. This pseudo-labeling-enhanced pipeline gradually increases the diversity and size of the training data, significantly improving model generalization with minimal labeling cost.

With repeated training cycles, the quality of pseudo-labels improves, and the model becomes increasingly confident in its predictions. Over time, this results in a more robust feature space and improved classification accuracy. Finally, the model is evaluated on a separate test set that was not used during training to assess its real-world performance. By combining transfer learning with pseudo-labeling in a unified pipeline, the proposed framework offers a scalable, efficient, and accurate solution for driver drowsiness detection based on facial imagery, particularly suitable for contexts with limited labeled data.

4.3 Evaluation Metrics

To comprehensively assess the performance of the proposed drowsiness detection model, a set of standard evaluation metrics was employed. These include **Accuracy, Precision, Recall, F1-Score**, and the **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**. Each of these metrics provides a different perspective on the model's predictive ability, especially in binary classification scenarios such as distinguishing between drowsy and alert driver states.

- **Accuracy :** Accuracy measures the proportion of correctly classified instances both positive and negative relative to the total number of predictions. It is a general indicator of overall model performance but may be insufficient on its own in imbalanced datasets.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.1}$$

- **Precision :** Precision, also known as Positive Predictive Value, evaluates the proportion of correctly predicted positive cases among all cases predicted as positive. It is especially relevant when the cost of false positives is high.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3.2}$$

- **Recall (Sensitivity) :** Recall quantifies the model's ability to correctly identify all relevant positive instances. It is defined as the ratio of true positives to the total number of actual positive instances. High recall is critical when missing a positive case (e.g, undetected drowsiness) carries serious risk.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3.3}$$

- **F1-Score :** The F1-Score is the harmonic mean of precision and recall, providing a single metric that balances both. It is particularly useful when there is an uneven class distribution or when both false positives and false negatives are critical.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.4}$$

- **AUC (Area Under the Curve) :** The Area Under the Curve (AUC) represents a comprehensive metric that quantifies the overall performance of a classification model based on the ROC curve. A higher AUC value reflects a stronger ability of the model to distinguish between positive and negative classes across varying decision thresholds.

Where:

- TP: True Positives
- TN: True Negatives

- FP: False Positives
- FN: False Negatives
- TPR: True Positive Rate
- FPR: False Positive Rate

These metrics collectively provide a robust framework for evaluating model performance, particularly in safety-critical applications such as real-time driver monitoring, where both detection accuracy and error minimization are crucial.

4.4 Work environment

- **Hardware environment :**

• PC: DESKTOP-0INKGUO

• Memory (RAM): 8.00 GB.

• Processor: Intel(R) Core(TM) i5-8250U CPU @ 1.80GHz

• System type: 64-bit operating system.

• During the training phase, high-performance GPUs were utilized via cloud computing, including:

o NVIDIA Tesla P100

o NVIDIA A100

- **Software environment :**

• Deep learning models were developed and implemented using the Python programming language.

• The PyTorch Lightning framework was employed to construct neural networks and manage the training process.

• Training and experimentation were conducted on cloud-based platforms, specifically:

o Google Colab Pro+

o Kaggle Notebooks

These platforms provide an integrated environment for model development, with direct support for GPU acceleration.

4.5 Experimental Results and Analysis

In this section, we present the quantitative results obtained from the experiments conducted using four different pre-trained convolutional neural network models: ResNet, MobileNet, DenseNet, and EfficientNet. The transfer learning technique was applied to adapt these models to the task of driver drowsiness classification.

The performance was evaluated using a set of common classification metrics: Accuracy, Precision, Recall, F1 Score, and AUC. The results demonstrate the effectiveness of the adopted deep learning approaches, with certain models showing superior performance in specific aspects (Table 4.3).

Model	**Accuracy**	**precision**	**Recall**	**F1 score**	**AUC**
ResNet 50	0.9877	0.9950	0.9782	0.9865	0.9978
MobileNet	0.9983	0.9967	0.9997	0.9982	0.9999
DenseNet201	0.9971	0.9974	0.9964	0.9969	0.9996
EfficientNetV2	0.9998	1.0000	0.9995	0.9997	1.000

Table 4.3 – model performance applying transfer learning

After analyzing the results of the four models ResNet50, MobileNet, DenseNet201, and EfficientNetV2, it is clear that all models demonstrated excellent performance in classifying driver drowsiness from images, reflecting the effectiveness of the transfer learning strategy in this context.

Focusing specifically on accuracy, the MobileNet model achieved the highest accuracy of 99.83%, along with an outstanding recall of 99.97% and a strong F1 score of 99.82%. These results indicate a high balance between correctly identifying positive and negative cases, making it the most suitable for applications requiring fast execution and low resource consumption, such as those on mobile devices and in real-time systems. In contrast, the ResNet50 model showed good performance but recorded the lowest accuracy among the four models at 98.77%, with a slightly lower recall of 97.82%, reflecting a relatively higher error rate compared to MobileNet, which may affect its reliability in sensitive operational environments.

The DenseNet201 model stood out with consistently excellent performance across all metrics, achieving an accuracy of 99.71% and recall of 99.64%, which can be attributed to its architectural design that facilitates information flow within the network and minimizes feature loss.

Finally, the EfficientNetV2 model outperformed all others, achieving the highest accuracy of 99.98% and the highest AUC value of 100%, indicating its superior ability to discriminate between classes. This model also offers an optimal balance between performance and accuracy with fewer computational parameters, making it an ideal choice for resource-constrained systems without sacrificing prediction quality.

For a more detailed understanding of each model's performance, confusion matrices and ROC curves were presented to illustrate the models' accuracy and error patterns.

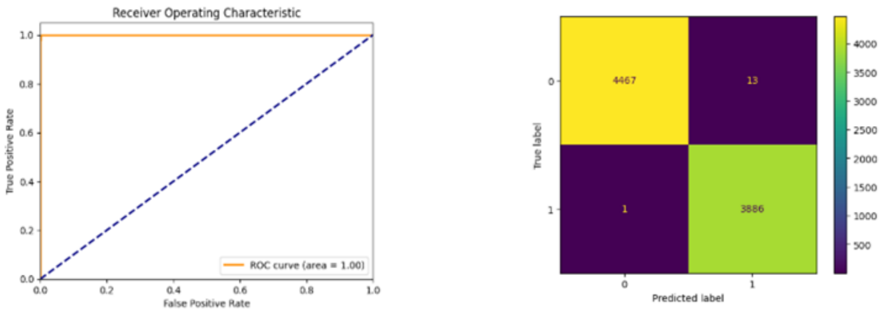


Figure 4.5 – Results using MobileNet

The evaluation metrics for the MobileNet model indicate very strong performance. The ROC curve attains an AUC of 1.00, reflecting complete separation between the two classes across threshold values.
From the confusion matrix, 4,467 negative samples were correctly classified, with a false positive count limited to just 13. On the positive side, 3,886 instances were identified correctly, while only a single positive sample was missed. This small number of errors indicates a well-balanced trade-off between sensitivity and precision, resulting in an overall low misclassification rate.
These results suggest that the MobileNet model is well-suited to binary classification tasks that require high accuracy and reliability.

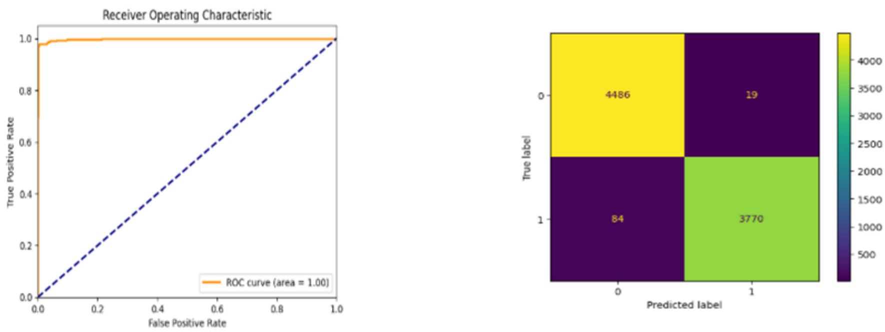


Figure 4.6 – Results using ResNet50

The ResNet50 model shows strong results across the evaluation metrics. The ROC curve, with an AUC of 1.00, indicates clear separation between the two classes with virtually no overlap.
The confusion matrix reveals that most samples were classified correctly: 19 negative cases were incorrectly flagged as positive, and 84 positive cases were missed. This distribution suggests a good balance between sensitivity and precision, with a relatively low overall error rate.
Taken together, these outcomes indicate that ResNet50 is a suitable choice for binary classification tasks where minimizing errors is important.

The figure illustrates the performance of the DenseNet201 model through the ROC curve and the confusion matrix. The ROC curve shows a near-perfect performance, with the curve closely hugging the top-left corner of the graph. This shape indicates that

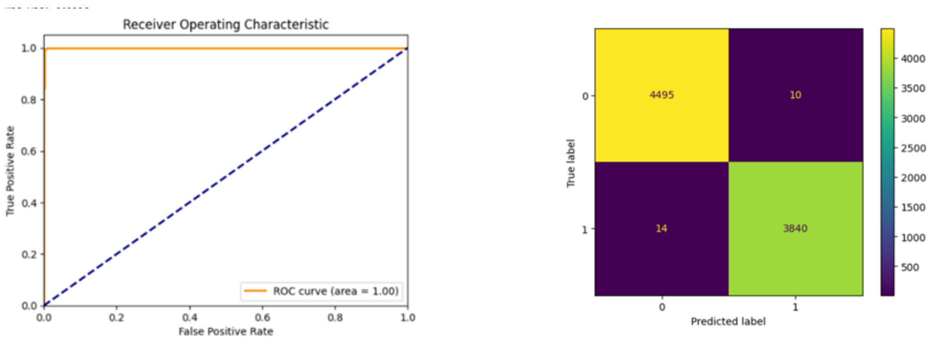


Figure 4.7 – Results using Densenet201

DenseNet201 achieves a very high true positive rate while maintaining a low false positive rate. The area under the ROC curve being approximately 1 confirms the model's strong discriminative power in classifying between drowsy and alert states.

The confusion matrix complements this result by revealing that the majority of predictions are correctly classified, as seen by the dominant diagonal values. Misclassifications are minimal, indicating that DenseNet201 maintains high precision and recall across both classes. This strong balance reflects the model's robustness and reliability in real-world drowsiness detection tasks.

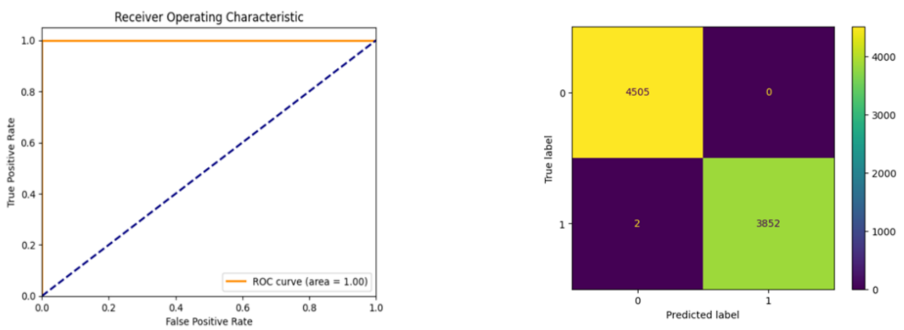


Figure 4.8 – Results using EfficientNetV2

The performance of the EfficientNetV2 model demonstrates exceptional effectiveness across key evaluation metrics. The ROC curve achieves an AUC of 1.00, indicating perfect separability between the two classes. This result confirms the model's ability to distinguish accurately between drowsy and alert states with no overlap, reflecting strong discriminative power and high reliability.

The confusion matrix further highlights this strength. All samples from the "alert" class were classified correctly, with zero false positives. For the "drowsy" class, only two samples were misclassified as alert, resulting in a minimal number of false negatives. This extremely low error rate indicates that the model maintains a near-perfect balance between precision and recall, ensuring both sensitivity and specificity. Such performance is particularly valuable in safety-critical applications like driver drowsiness detection, where minimizing misclassification is crucial.

Overall, EfficientNetV2 proves to be a highly robust and accurate model, delivering near-flawless classification results in this context.

4.6 Comparison with Previous Work

Ref	Dataset	Method	Reported Accuracy
2021 [31]	UTA-RLDD	DenseNet201	99.00%
2024 [32]	UTA-RLDD	EfficientNetV2, ResNet50	94.03%
2024 [33]	UTA-RLDD, YawDD	MobileNet	98.44%
2025 (our project)	UTA-RLDD	ResNet50	98.77%
		DenseNet201	99.71%
		MobileNet	99.83%
		EfficientNetV2	99.98%

Table 4.4 – Comparison of Reported Accuracy on UTA-RLDD and YawDD Datasets

Table 4.4 presents a comparative analysis between the proposed system and several recent studies using the UTA-RLDD dataset for drowsiness detection. The study in [31] achieved 99% accuracy using DenseNet201, while our DenseNet201-based model slightly outperformed it, reaching 99.71%. Similarly, the models presented in [32] used EfficientNetV2 and ResNet50 and obtained a maximum accuracy of 94.03%, whereas our implementation of these models achieved 99.98% and 98.77%, respectively. In [33], MobileNet was applied using both UTA-RLDD and YawDD datasets, yielding 98.44% accuracy.

Our MobileNet implementation surpassed this with a performance of 99.83%. These results demonstrate the effectiveness and robustness of our proposed system, particularly in optimizing deep learning architectures through fine-tuning and enhanced preprocessing techniques.

4.7 Conclusion

This chapter, which presents the experimental results, highlights the strong capabilities of all evaluated convolutional neural network models in detecting driver drowsiness. MobileNet and EfficientNetV2 stood out for their balance between accuracy and efficiency, making them ideal choices for practical applications and real-time monitoring systems. This study confirms the value of transfer learning in developing effective and resource-efficient systems for drowsiness detection.

Chapter 5

General conclusion and future work

Detecting driver drowsiness through facial image analysis has recently garnered significant attention due to its potential to enhance road safety and reduce accidents caused by fatigue. However, accurately identifying drowsiness remains a challenging task because of the subtle and variable nature of facial fatigue indicators under different lighting conditions, head poses, and facial structures. This research addresses this challenge by developing an intelligent system capable of distinguishing between alert and drowsy states with high precision, using advanced deep learning techniques and transfer learning with pseudo-labeling.

The study began with a review of fundamental concepts in deep learning, transfer learning, and convolutional neural networks, establishing a solid theoretical foundation for the system design. We implemented several state-of-the-art models, including DenseNet201, EfficientNetV2, ResNet50, and MobileNet, leveraging both pre-trained weights and training from scratch to improve performance on a real-world dataset featuring complex challenges.

Utilizing the UTA-RLDD dataset, which is considered one of the most comprehensive datasets for early drowsiness detection, provided a robust evaluation environment. Our models demonstrated outstanding performance, surpassing or closely matching the accuracy reported in recent studies, achieving accuracy rates of up to 99.98%. These results confirm the effectiveness of transfer learning in adapting existing deep models to new tasks with limited data while maintaining high accuracy.

This research emphasizes the importance of extracting distinctive features from facial images related to fatigue indicators and employing fine-tuning techniques to optimize the model's performance in response to new data characteristics. The proposed system lays the groundwork for future enhancements through the integration of multimodal data or advanced feature extraction methods.

Future perspectives for the development of drowsiness detection systems will leverage advanced AI techniques, such as generative models like GANs, to augment training data and enhance system performance. Multimodal transformer architectures will enable the simultaneous analysis of visual, auditory, and physiological signals, leading to improved accuracy. Federated learning enhances privacy by training models on devices without requiring data to be shared centrally. Edge computing with IoT devices enables real-time processing within vehicles, reducing latency and preserving user privacy. Ultimately, adaptive alert systems will deliver dynamic and personalized warnings to enhance driver safety while minimizing distractions.

Bibliography

[1] B. C. Tefft *et al.*, *Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009-2013*. AAA Foundation for Traffic Safety Washington, DC, 2014.

[2] V. Saini and R. Saini, "Driver drowsiness detection system and techniques: a review," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 4245–4249, 2014.

[3] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness detection using heart rate variability," *Medical & biological engineering & computing*, vol. 54, pp. 927–937, 2016.

[4] X. Zhu, W.-L. Zheng, B.-L. Lu, X. Chen, S. Chen, and C. Wang, "Eog-based drowsiness detection using convolutional neural networks," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 128–134.

[5] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque, "Sensor applications and physiological features in drivers' drowsiness detection: A review," *IEEE sensors Journal*, vol. 18, no. 8, pp. 3055–3067, 2018.

[6] M. Mahmoodi and A. Nahvi, "Driver drowsiness detection based on classification of surface electromyography features in a driving simulator," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 233, no. 4, pp. 395–406, 2019.

[7] C. C. Liu, S. G. Hosking, and M. G. Lenné, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *Journal of safety research*, vol. 40, no. 4, pp. 239–245, 2009.

[8] C.-C. Lien and P.-R. Lin, "Drowsiness recognition using the least correlated lbph," in *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, 2012, pp. 158–161.

[9] T. Bafna and J. P. Hansen, "Mental fatigue measurement using eye metrics: A systematic literature review," *Psychophysiology*, vol. 58, no. 6, p. e13828, 2021.

[10] M. Iwasaki, C. Kellinghaus, A. V. Alexopoulos, R. C. Burgess, A. N. Kumar, Y. H. Han, H. O. Lüders, and R. J. Leigh, "Effects of eyelid closure, blinks, and eye movements on the electroencephalogram," *Clinical Neurophysiology*, vol. 116, no. 4, pp. 878–885, 2005.

[11] M. L. Jackson, S. Raj, R. J. Croft, A. C. Hayley, L. A. Downey, G. A. Kennedy, and M. E. Howard, "Slow eyelid closure as a measure of driver drowsiness and its relationship to performance," *Traffic injury prevention*, vol. 17, no. 3, pp. 251–257, 2016.

[12] S. Liu, J. Wang, and T. Fu, "Effects of lane width, lane position and edge shoulder width on driving behavior in underground urban expressways: A driving simulator study," *International journal of environmental research and public health*, vol. 13, no. 10, p. 1010, 2016.

[13] H. Summala, "Brake reaction times and driver behavior analysis," *Transportation human factors*, vol. 2, no. 3, pp. 217–226, 2000.

[14] S. A. Vyas, A. A. Suke, and M. Gedam, "Real-time intelligent alert system on driver's hypo-vigilance detection using template matching technique," *Int. J. Recent Innov. Trends Comput. Commun*, vol. 3, pp. 665–671, 2015.

[15] Y. Alotaibi and B. Alotaibi, "Driver drowsiness detection using deep learning techniques: A review," *IEEE Access*, vol. 8, pp. 179 960–179 973, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.3028850

[16] H. Reese, "Understanding the differences between ai, machine learning, and deep learning," *URL: https://www. techrepublic. com/article/understandingthedifferences-betweenaimachine learninganddeeplearning*, vol. 11, no. 02, p. 2023, 2017.

[17] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, "Deep learning: a primer for radiologists," *Radiographics*, vol. 37, no. 7, pp. 2113–2131, 2017.

[18] H. Deekshith Shetty, M. J. Varma, S. Navi, and M. R. Ahmed, "Diving deep into deep learning: history, evolution, types and applications," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 2278–3075, 2020.

[19] A. Picon, A. Alvarez-Gila, U. Irusta, and J. Echazarra, "Why deep learning performs better than classical machine learning," *Dyna Ing. Ind*, vol. 95, pp. 119–122, 2020.

[20] J. Brownlee, "Difference between machine learning and deep learning," https://machinelearningmastery.com/difference-between-a-machine-learning-and-deep-learning/, 2016.

[21] M. Sewak, S. K. Sahay, and H. Rathore, "An overview of deep learning architecture of deep neural networks and autoencoders," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 1, pp. 182–188, 2020.

[22] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research," *Journal of pharmaceutical and biomedical analysis*, vol. 22, no. 5, pp. 717–727, 2000.

[23] A. A. Elngar, M. Arafa, A. Fathy, B. Moustafa, O. Mahmoud, M. Shaban, and N. Fawzy, "Image classification based on cnn: a survey," *Journal of Cybersecurity and Information Management*, vol. 6, no. 1, pp. 18–50, 2021.

[24] M. Coşkun, Ö. Yıldırım, A. Uçar, and Y. Demır, "An overview of popular deep learning methods," *European Journal of Technique (EJT)*, vol. 7, no. 2, pp. 165–176, 2017.

[25] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, p. 99, 2024.

[26] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[27] S. J. Pan, “Transfer learning,” *Learning*, vol. 21, pp. 1–2, 2020.

[28] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, “Plant leaf disease classification using efficientnet deep learning model,” *Ecological Informatics*, vol. 61, p. 101182, 2021.

[29] R. Ghoddoosian, M. Galib, and V. Athitsos, “Uta real-life drowsiness dataset (uta-rldd),” https://sites.google.com/view/utarldd/home, accessed: 2025-05-26.

[30] K. Kutlu, “Drowsiness detection,” https://www.kaggle.com/datasets/kutaykutlu/drowsiness-detection, 2019, accessed: 2025-05-26.

[31] A. H. Hasan, A. A. Yasir, and M. J. Hayawi, “Driver drowsiness detection based on the densenet 201 model,” *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 13, pp. 3682–3692, 2021.

[32] A. Rajak, P. Hatwar, A. Tiwari, G. Sahu, and R. Tripathi, “Efficient light-weight deep learning models for drowsiness detection,” in *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)*. IEEE, 2024, pp. 1–6.

[33] D. H. Vazquez-Santana and A. J. Argüelles-Cruz, “Non-intrusive drowsiness detection for accident prevention using a novel cnn,” *Computación y Sistemas*, vol. 28, no. 3, pp. 1167–1181, 2024.