STUDY AND ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR DETECTION OF DISTRACTED DRIVERS

by

Fangming Qu

A Thesis Submitted to the Faculty of

The College of Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, FL

MAY 2024

Copyright 2024 by Fangming Qu

STUDY AND ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR DETECTION OF DISTRACTED DRIVERS

by

Fangming Qu

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Mehrdad Nojoumian, Department of Computer and Electrical Engineering and Computer Science, and has been approved by the members of his supervisory committee. It was submitted to the faculty of the College of Engineering and Computer Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Mehrdad Nojoumian (Apr 9, 2024 16:55 EDT)

Mehrdad Nojoumian, Ph.D.
Thesis Advisor

Borko Furht (Apr 9, 2024 19:18 EDT)

Borko Furht, Ph.D.

Taghi Khoshgoftaar (Apr 10, 2024 09:25 EDT)

Taghi Khoshgoftaar, Ph.D.

Hari Kalva, Ph.D.
Chair, Department of Computer and Electrical Engineering and Computer Science

Stella N. Batalama, Ph.D.
Dean, The College of Engineering and Computer Science

Robert W. Stackman Jr., Ph.D.
Dean, Graduate College

April 16, 2024

Date

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my advisor, Dr. Nojoumian, for his invaluable guidance, mentorship, and support throughout this research project. I extend my thanks to Dr. Borko Furht and Dr. Taghi Khoshgoftaar for serving on my thesis committee and providing insightful feedback.

I am deeply grateful to my research partner, Nolan Dang, for his outstanding collaboration and co-authorship on our published paper, *"Comprehensive Study of Driver Behavior Monitoring Systems Using Computer Vision and Machine Learning Techniques"*.

I would like to express my sincere gratitude for the contributions that have significantly enhanced this work, emphasizing my commitment to research integrity and the collaborative spirit of academic inquiry. Zeren Shen, from the University of Waterloo, has provided an invaluable open-source PyTorch code, laying a solid foundation for my research [1]. Equally, the creators of the MoveNet model have contributed innovative code and deployment resources that have been instrumental in the development of this project [2]. By acknowledging these contributions, I underscore the importance of transparency and the collective nature of scientific advancement.

Finally, I would like to thank Florida Atlantic University (FAU) for providing the infrastructure, resources, and supportive research environment that made this work possible.

ABSTRACT

Author:	Fangming Qu
Title:	Study and Analysis of Machine Learning Techniques for Detection of Distracted Drivers
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Mehrdad Nojoumian
Degree:	Master of Science
Year:	2024

The rise of Advanced Driver-Assistance Systems (ADAS) and Autonomous Vehicles (AVs) emphasizes the urgent need to combat distracted driving. This study introduces a fresh approach for improved detection of distracted drivers, combining a pre-trained Convolutional Neural Network (CNN) with a Bidirectional Long Short-Term Memory (BiLSTM) network. Our analysis utilizes both spatial and temporal features to examine a broad array of driver distractions. We demonstrate the advantage of this CNN-BiLSTM framework over conventional methods, achieving significant precision (up to 98.97%) on the combined 'Union Dataset,' merging the Kaggle State Farm Dataset and AUC Distracted Driver Dataset (AUC-DDD). This research enhances safety in autonomous vehicles by providing a solid and flexible solution for everyday use. Our results mark considerable progress in accurately identifying driver distractions, pushing the boundaries of safety technology in AVs.

STUDY AND ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR DETECTION OF DISTRACTED DRIVERS

List of Tables ix

List of Figures x

1 Introduction 1

2 Literature Review of Self-Driving Car Technology 5

2.0.1 Autonomous Vehicle 5

2.0.2 Vision Systems and Machine Learning in AVs 7

2.0.3 Roles of AI and Machine Learning in Autonomous Vehicles . . 9

2.0.4 Driver Behavior Classification 10

2.0.5 Deep Learning in Autonomous Vehicles 11

2.0.6 Artificial Neural Network (ANN) 12

2.0.7 Convolutional Neural Network (CNN) 12

2.0.8 Recurrent Neural Network (RNN) 15

2.0.9 Long Short-Term Memory (LSTM) 16

2.0.10 Role of CNN and LSTM in Autonomous Vehicles 16

2.0.11 Bidirectional LSTMs (BiLSTM) 18

2.0.12 Time-Distributed Layer 19

2.0.13 Driver Behavior Classification 20

2.0.14 Hand Classification 22

2.0.15 Facial Classification 27

2.0.16 Body Posture Classification 32

2.0.17 Integration of Physiological Indicators Classification 37
2.0.18 Pose Monitor AI Program 43

3 Methodology 47
3.0.1 Leveraging CNNs and BiLSTMs for Distraction Detection . . 47
3.0.2 Data Augmentation in Enhancing Model Generalization . . . 47
3.0.3 Experimental Setup 48
3.0.4 Model Configurations 48
3.0.5 Assessing Performance 49
3.0.6 Challenges and Considerations 49
3.1 Datasets: State Farm and AUC-DDD, and rationale for merging . . . 50
3.1.1 State Farm Dataset 50
3.1.2 AUC Dataset 50
3.1.3 Union Dataset 50
3.2 Architectural design: CNN, Time Distributed layer, and BiLSTM . . 52
3.3 Experimental setup: Data preparation, training, and evaluation metrics 53
3.3.1 Preliminary Setup 53
3.3.2 Model Selection for Pose Estimation 54
3.3.3 Environment Configuration 55

4 Results and Analysis 56
4.1 Performance comparison: Baseline models v.s. proposed model 56
4.2 Impact of the BiLSTM layer and Union Dataset 60

5 Discussion 63
5.1 Analysis of the BiLSTM layer's effectiveness 63
5.2 Contributions and limitations of the Union Dataset 63
5.3 Comparison with existing literature 63

6 Conclusions and Future Work 66

6.1 Summary of findings 66

6.2 Implications for autonomous vehicle safety and AI 66

6.3 Suggestions for future research 67

Bibliography 68

LIST OF TABLES

2.1 Summary of research on Hand classification. 23

2.2 Summary of research on Facial classification. 29

2.3 Summary of research on Body Posture classification. 33

2.4 Comparison of combined classifying models. 40

4.1 Performance on Union Dataset 57

4.2 Performance on AUC Dataset 58

4.3 Performance on State Farm Dataset 59

LIST OF FIGURES

2.1 Visualization of CNN Sample based on the reference of [3] 13

2.2 Comparative Accuracy of CNN vs. CNN-LSTM Models Over Training Epochs[4]. 17

2.3 The Unrolled Bidirectional LSTM Structure [5]. 18

2.4 Detailed architecture with visualization of time-distributed layer [6]. . 19

3.1 Ten Classes of Driver Distraction training example from AUC-DDD [7] 51

4.1 Comparative Performance of Deep Learning Models on Union, AUC, and State Farm Datasets 56

4.2 Test Accuracy trends for CNN models with versus without BiLSTM . 60

4.3 Test Accuracy Across Models and Datasets 61

5.1 Training Time Efficiency Across Models and Datasets 64

CHAPTER 1

INTRODUCTION

Distracted driving poses a severe threat to road safety, claiming over 3,000 lives anually in the United States. Recent advances in AVs present new opportunities to enhance in-vehicle safety. A crucial aspect is understanding driver behavior, as inattention has been identified as a major contributor to accidents [8]. To address this, the survey paper examines current driver inattention detection methods, emphasizing the advantages of hybrid approaches that combine multiple indicators as well as highlighting the limitations of existing detection systems[9]. Building upon these insights, this thesis introduces an innovative CNN+BiLSTM with a time-distributed layer architecture for improved detection of distracted driving behaviors.

Recently, an increasing focus has been on creating self-driving or autonomous vehicles - vehicles that can operate without human intervention. This development has opened up new ways to increase safety in these vehicles. A key aspect is the capability to understand and keep an eye on what is happening inside the vehicle, particularly with the driver. A review conducted by researchers from Japan [8] shows that driver inattention was a leading cause of most traffic accidents. Researchers have extensively studied this issue, categorizing driver inattention into two primary types: distraction and fatigue. Detecting and mitigating driver inattention requires a multifaceted approach, incorporating subjective reports, driver biological indicators, physical measurements, driving performance assessments, and hybrid measures that combine multiple indicators. Among these, hybrid measures offer more reliable and accurate solutions compared to relying on a single measure. However, commercial products for driver inattention monitoring exist, and their effectiveness in actual

driving conditions may be limited. An ideal driver inattention monitoring system for safety enhancement integrates driver physical variables, driving performance metrics, and data from the In-Vehicle Information System (IVIS) while considering the driving environment. To address this challenge, this research aims to develop AI-based monitoring software for autonomous vehicles, contributing to a safer and more secure transportation landscape.

In this survey paper, an AI-based driving assistant is proposed that can see and interpret the inside of the vehicle using a branch of artificial intelligence algorithms, which allows computers to learn and make decisions from data: an AI-based offline monitoring system to boost the safety of these autonomous vehicles. This system is designed to assist the driver and issue warning alerts if the driver seems to be not paying attention to the road without any data privacy concerns. This thesis proposes a unique AI-based monitoring system designed specifically for the context of autonomous vehicles. Our evaluation goes beyond the survey's proposed assistant by incorporating driving performance data and IVIS integration for enhanced accuracy. By combining three distinct classification methods to detect fatigued drivers, a software system that works on an autonomous vehicle acts as an intelligent driving assistant. There are various machine learning methods, known as neural networks, for analyzing these behaviors. These include artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNN).

This paper begins by defining essential computer science concepts relevant to autonomous vehicles: autonomous vehicles, vision systems, machine learning, driver behavior classification, deep learning, convolutional neural networks, recurrent neural networks, and artificial neural networks. It then outlines the proposed system's framework, detailing the specific algorithms used for each function and their integration. Section four examines how different datasets will be employed to train the algorithm effectively, ensuring robust performance. Finally, the survey demonstrates

how these elements combine to form a comprehensive AI-based monitoring solution designed to improve autonomous vehicle safety. This research endeavors to make a significant contribution to the development of safer and more reliable transportation systems.

This system leverages artificial intelligence algorithms to interpret the driver's state and provide timely alerts. It aims to surpass existing systems by incorporating visual cues, driving performance data, and potentially information from the In-Vehicle Information System. Unlike current systems, which often miss the full complexity of these behaviors, our analysis, trained on a unique 'Union Dataset,' aim to better detect distracted driving behaviors more effectively than existing systems. Our goal is to significantly advance detection precision and adaptability in autonomous vehicle environments [10].

Traditional detection systems frequently fail to address the nuance of driver behaviors, leading to sub-optimal performance. Our analysis, validated by the University of Nottingham, surpasses current models, with the CNN+BiLSTM framework achieving an average classification accuracy of 92.7 percent [11]. Moreover, this combination has proven superior in identifying temporal sequences and patterns in driver behavior, essential for understanding distractions over time [12, 13].

The Union Dataset offers significant value for evaluating distraction detection models. Our analysis demonstrates that incorporating diverse distraction scenarios enhances the understanding of model performance in real-world applications. The increasing prevalence of distracted driving, especially among young people, highlights the critical need for data-driven analysis to inform the development of improved countermeasures [14, 15].

This study investigates whether a CNN+BiLSTM with time-distributed layer model can markedly improve the accuracy of distracted driving detection in AV systems beyond traditional methods. It outlines the key benefits of the architecture

– extracting spatial features (CNN's strength) and analyzing time-based patterns (BiLSTM's strength). Building on insights from successful complex driver posture identification, this research explores a unique approach with the potential to significantly improve distraction detection rates compared to current standards [16, 12, 17].

Our analysis aims to provide a more detailed understanding of temporal patterns in driver behavior and their impact on distraction detection effectiveness. We employ a CNN-BiLSTM architecture with a time-distributed layer, trained on the Union Dataset, to investigate these relationships beyond the capabilities of existing methods. This thesis demonstrates that our proposed approach substantially surpasses the capabilities of traditional detection methods. In summary, this research aims to significantly advance AV safety and establish a new benchmark for the automotive industry's safety solutions.

CHAPTER 2

LITERATURE REVIEW OF SELF-DRIVING CAR TECHNOLOGY

2.0.1 Autonomous Vehicle

We should first emphasize that driver behavior monitoring systems have been used in human-driven vehicles for a long time, for instance, by insurance companies through mobile apps or small hardware equipment. However, these systems are more critical for autonomous vehicles as well as vehicles that utilize advanced driver-assistance systems, i.e., using a level of autonomy, because they may face unpredictable situations requiring the driver's intervention.

An autonomous vehicle with levels of autonomy revolves around a vehicle that can operate without (or with minimum) human intervention. These vehicles are designed to navigate and drive themselves, relying on advanced technologies and systems rather than requiring a human driver. These vehicles contain complex computer systems that often utilize artificial intelligence. AI acts similarly to the brain, processing information and making decisions. It gathers data from its surroundings via sensors, which function as the vehicle's eyes, then uses it to navigate safely, just as a human driver would. Vehicles that drive themselves have the potential to change our world. AI technologies could transform how people and goods travel, advance military and security operations, and provide a new level of freedom to those unable to drive. Furthermore, ADAS and AVs are potentially expected to make roads safer and save fuel, providing better transportation options, especially for those who have difficulties in driving, and reshaping society's transportation approach entirely.

Autonomous vehicles offer potential benefits compared to human drivers, as re-

ported in [18]. Firstly, while AVs can reduce accidents, it's essential to recognize that both AVs and human drivers share equal responsibility for preventing accidents. The report notes that many traffic accidents result from unsafe driving behaviors or drowsy driving. Secondly, they reduce traffic congestion by optimizing traffic flow and improving communication among road vehicles through in-between vehicles telecommunication, addressing the issue of inefficient traffic management. Additionally, using AI technologies, AVs provides accessibility and mobility for individuals unable to drive, thus promoting privacy and providing cost savings for those who opt for AV transportation instead of owning a vehicle, addressing economic considerations.

There are also disadvantages to autonomous vehicles. For example, addressing the technological challenges of software and hardware systems is essential. This includes developing robust algorithms, ensuring effective communication between components, and enhancing overall system reliability and safety. Then, adopting AVs raises concerns about job displacement in the driving and transportation sectors, potentially resulting in unemployment and necessitating retraining or job transition programs. Also, ethical and legal considerations arise when adapting laws to accommodate autonomous technologies. Addressing liability in accidents, decision-making processes in critical situations, and establishing an ethical framework for AV behavior are crucial for public trust and safety. Additionally, cybersecurity threats must be considered, as hackers targeting AVs could gain control over operations and endanger passengers and other road users. Robust cybersecurity measures should be implemented to prevent such risks. Furthermore, privacy concerns of AV drivers require attention, with clear guidelines and safeguards to protect personal information collected by autonomous vehicles. Moreover, the reliance on infrastructure and connectivity challenges widespread adoption and effectiveness. Consistent and reliable support, including road markings, traffic signals, and communication networks, is vital for successfully integrating autonomous technologies on a large scale. Despite

these challenges, the overall advantages of AVs outweigh the disadvantages. These are in addition to other human-machine interaction (HMI) challenges such as cross-cultural expectations from self-driving cars [19, 20], passengers' trust [21, 22], social acceptability [23, 24], and customized autonomous driving technologies [25, 26].

While complete vehicle automation is yet to be commonplace soon, the interpretation of driver behavior is crucial for partially and conditionally automated vehicles. These vehicles, which require either the driver's readiness to regain control at any moment or their intervention when the vehicle cannot perform certain critical operations, are predicted to be the dominant form in the market until 2030. Since these systems are automated, they still heavily rely on human supervision and intervention [27, 28, 29].

2.0.2 Vision Systems and Machine Learning in AVs

In the vision system, cameras and sensors function as the eyes of the computer, with sophisticated software algorithms acting like a human brain to interpret the data they capture. These systems analyze images, breaking them down to understand each element, enabling the recognition of faces, objects, and navigation paths. Such vision systems empower machines to "envision" and "understand" their surroundings, playing a pivotal role in diverse applications such as robotics, security systems, autonomous vehicles, and mobile phones.

Driver distraction, a major contributor to vehicular accidents, is well-documented and can be effectively monitored by these vision systems. However, it's important to note that while these systems achieve impressive performance levels, there are ongoing concerns about their evaluation methodologies. Often, AI-based models in vision systems are assessed using datasets involving drivers who were part of the training set. This practice can lead to a potential 'memory' effect, where the models are fine-tuned to the characteristics of these specific drivers, raising questions about

their ability to generalize to new, unseen drivers [30, 31, 32, 33, 34].

To counter this, it is crucial to incorporate evaluation strategies that ensure these models can effectively adapt to diverse driving behaviors not represented in the training data. This involves using more heterogeneous datasets and applying rigorous cross-validation techniques. Such approaches are essential to evaluate the real-world applicability and robustness of these vision systems, ensuring they remain effective across a broader range of scenarios and driver behaviors. Recent reviews, like Ji's panoramic study, highlight various non-invasive approaches for detecting signs of fatigue using vision systems. These methods leverage video analysis of a driver's visual characteristics to identify fatigue levels. However, the success of these techniques hinges on the models' ability to generalize their learning to a wide array of drivers. The development of adaptable and broad-based AI models remains a key area of research in enhancing the effectiveness of vision systems in real-world applications [35].

Visual perception, fundamental to driving, relies heavily on visual sensors for data capture. However, this data contains abundant indirect information that machine vision and image understanding techniques handle. Smart vehicles, from advanced assistance systems to autonomous vehicles, leverage machine vision to distill and categorize video data, making it useful for driving. Techniques like convolutional neural networks play a crucial role in identifying specific objects in traffic and aiding in the mapping and positioning of self-driving cars. They also incorporate the discussion of real-time computing architectures backed by real-world experiments. However, the field of vision systems is witnessing an inventive shift where, instead of trying to identify objects universally, the system should adapt its method based on the size and context of the object under observation. It requires the system to be adaptable enough to modify its strategy according to the target, using different techniques for smaller or less transparent objects than for larger, more detailed ones. More than

just theoretical, this concept has been tested and has outperformed other methods on popular benchmarks. The adaptable vision system reinforces that there is always room for innovation and improvement in performance, especially as vision systems tackle various real-world situations [36, 37].

2.0.3 Roles of AI and Machine Learning in Autonomous Vehicles

Machine learning is a sub-field of computer science that gives computers the capacity to learn from data and subsequently make informed decisions or predictions. This concept embodies the machine equivalent of a human brain, utilizing a variety of algorithms and statistical models to learn and adapt over time. Three primary types of learning exist in this context: supervised, unsupervised, and reinforcement learning.

The algorithm assumes a student-like role in supervised learning, with data presented as question-and-answer pairs serving as the tutor. Through exposure to this data, the algorithm learns the pattern of problem-solving and, over time, acquires the ability to solve similar problems independently. Supervised learning is analogous to a learning paradigm in which a student learns by being presented with a problem and the corresponding solution.

Unsupervised learning, in contrast, presents the algorithm with a dataset without the provision of predefined solutions. The focus falls on the algorithm to identify patterns, correlations, and relationships within the data. This method is similar to a student given tasks without an explicit solution, necessitating independent pattern recognition and problem-solving.

Reinforcement learning takes a different approach, embodying a process of repetitive learning through trials and errors, reminiscent of learning to ride a bicycle. Each attempt and subsequent failure offers the algorithm new insights, adjusting its future decisions based on gathered experiences.

The utility of machine learning pervades a multitude of sectors in contemporary

times. From facial recognition capabilities on smartphones to stock market trend predictions, its applications span sectors including healthcare, finance, retail, and transportation. Machine learning equips us with the tools to identify patterns within extensive data sets and informs data-driven decision-making processes.

In the following sections, this paper will comprehensively explore different learning types, their associated algorithms, and the vast array of applications where machine learning is harnessed. This survey aims to present an inclusive examination covering the extensive spectrum of machine learning. By examining the diverse algorithms, learning models, and machine learning applications, this paper aims to provide a holistic view of this rapidly evolving field.

2.0.4 Driver Behavior Classification

Recognizing and understanding the unique driving behaviors of individuals is crucial for enhancing drivhand movements, facial expressions, and body postureer awareness. This is where driver behavior classification draws its significance, acting as a sophisticated observer that continually monitors and analyzes the driver's actions, hand movements, facial expressions, and body posture. By leveraging advanced technologies including vision systems, machine learning algorithms, and sensor data, driver behavior classification aims to provide real-time feedback and promote safer driving practices.

Driver behavior classification encompasses multiple aspects. Firstly, hand classification monitors the driver's hand movements. For instance, if the system detects the driver's hand off the wheel, including reaching for a coffee cup, it gently reminds them to keep their hands on the wheel. Secondly, facial classification focuses on analyzing facial expressions that could indicate fatigue or distraction. If the driver's eyes frequently close, indicating drowsiness, the system can trigger an alert or activate the autonomous system if necessary. Lastly, body posture classification examines the

driver's posture and movements. If the driver starts slouching after extended periods of driving, it suggests the need for a break or seat adjustments to promote comfort. These driver behavior classifications serve as indicators of potential safety concerns. Considering the importance of transportation safety and the seamless integration of autonomous vehicles, it is essential to also address the legal perspectives and regulatory challenges associated with these advanced systems. As autonomous vehicles become more prevalent, their successful integration relies on the ability to identify and adapt to human driving behaviors. This adaptability ensures that autonomous vehicles can assimilate naturally into existing traffic flows, promoting safer and more efficient travel.

2.0.5 Deep Learning in Autonomous Vehicles

With the surge in computational capabilities and the breakthroughs in machine learning that began in the early 2010s, it became possible to develop more advanced visual algorithms, particularly those reliant on extensive convolutional neural networks. This era marked significant progress in deep learning, which substantially strengthened the development and reliability of autonomous vehicle technologies. A notable contribution to this field is the "You Only Look Once" (YOLO) framework by Joseph Redmon, which revolutionized visual machine learning approaches [38].

Before the advent of YOLO, the visual systems in autonomous vehicles primarily depended on edge detection and classic image processing methods, such as filtering, to identify objects in each frame. However, the introduction of computer vision breakthroughs like YOLO has empowered companies, notably Tesla, to fully embrace visual-based autonomous driving through their exclusive technology, Tesla Vision [39]. Despite this, other leading companies remain cautious, opting for a hybrid approach that combines cameras with LiDAR systems, rather than relying solely on visual systems.

2.0.6 Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are computational models directly inspired by the structural and functional characteristics of the human brain. They embody a network of interlinked artificial neurons that synergistically function to learn from data and generate predictions. Similar to the data processing mechanism of the human brain, an ANN accepts inputs, performs calculations, and yields outputs. ANNs can be compared to a well-coordinated team of professionals, each having a defined role, passing information in a relay. Each member takes input, performs a specific calculation, and forwards the resulting data to the next member. This process iterates until the final member delivers the output. During its training phase, an ANN adjusts its calculations based on provided examples, essentially learning through fine-tuning. It aims to find the most effective methodology to make accurate predictions or decisions.

A noteworthy aspect of ANNs is their ability to comprehend complex patterns and relationships in data, even revealing associations that may not be immediately apparent. As a result, ANNs are invaluable in diverse applications including image recognition, natural language understanding, and even autonomous driving. Integrating different types of ANNs, including recurrent neural networks (RNNs) for sequential data processing and convolutional neural networks (CNNs) for image analysis, has made significant advancements in domains like natural language processing, image recognition, and autonomous vehicles.

In conclusion, ANNs are potent tools for learning from data and making predictions or decisions. They are universally employed across many fields to unravel complex problems and augment our understanding of and interaction with the world.

2.0.7 Convolutional Neural Network (CNN)

Figure 2 below provides a visual representation of a CNN sample, as elaborated in reference [3], offering a graphical insight into the model's structure and function.

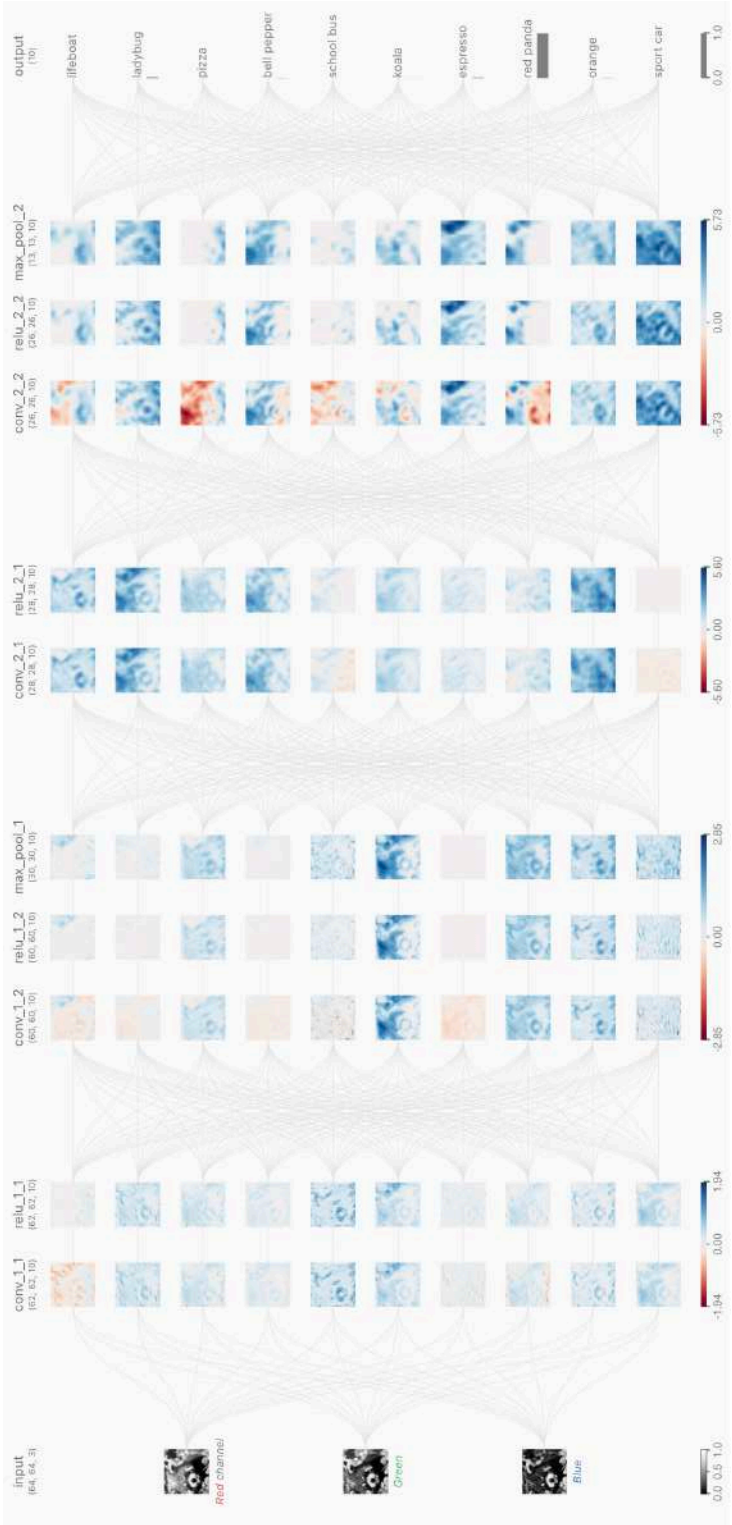


Figure 2.1: Visualization of CNN Sample based on the reference of [3]

Convolutional neural network stands as an essential component in the toolkit of deep learning techniques, especially in domains necessitating image understanding and interpretation. For humans, interpreting a picture is an intuitive process; however, for computers, the task is significantly more complex as it perceives the image as an array of tiny points or "pixels." CNNs facilitate computers in comprehending these pixels and their interrelationships.

CNN's operation begins with an input layer. The initial step is where the image data comes into the CNN model. In instances of a color image, the input layer receives information on the three primary colors: red, green, and blue. Following the input layer is the convolutional layer. During this phase, the layer traverses the image in small sections, referred to as filters or kernels, to construct a feature map. Feature maps enable CNN to recognize fundamental shapes or patterns, including lines or textures. Subsequently, the activation function, typically the ReLU (Rectified Linear Unit), is applied. This function augments the CNN's learning capabilities by nullifying any negative pixel values, thereby increasing the effectiveness of the CNN in identifying complex patterns. The subsequent pooling layer condenses the information derived from the convolutional layer by downsizing the feature map but preserving the essential features. The pooling process increases the efficiency of CNNs. Max pooling is a universally deployed method, which retains only the highest value from each section of the feature map. After several iterations of these steps, the fully-connected layer is activated. This layer assimilates the information collected thus far to arrive at a final decision, categorizing the image appropriately. Finally, the Softmax function generates a probability distribution for each category, indicating the likelihood of the image belonging to each class. Essentially, a CNN transforms the raw pixels of an image into a categorized output, enabling the computer to interpret the image. In essence, CNNs can identify complex patterns and objects within an image, similar to human visual interpretations.

2.0.8 Recurrent Neural Network (RNN)

Recurrent neural networks (RNNs) are powerful tools that allow computers to understand sequences of data, similar to a time radar. Sequences can range from sentences in text, frames in a video, or a series of numbers. RNNs enable computers to comprehend these sequences as a series of data points.

The initial step is the input layer. In the case of a sentence, the input layer receives vectors, which are representations of the input data. The next step is the recurrent layer, where RNNs pass information from one point to the next in the sequence. The sequential information exchange aids in understanding the order with the context of previous information. Sequentially, RNNs apply a comprehending function, including "ReLU", to the output of the previous layer, in order to understand the contextual meaning of the data by identifying intricate patterns within the sentence. Next, RNNs progress to the fully-connected layer, consolidating the knowledge into the decision-making process which involves tasks including prediction and classification. Finally, RNNs employ the Softmax function, which assigns a probability score to each possible outcome.

The potential of RNNs extends even further as researchers continually explore ways to enhance their capabilities. One notable advancement is integrating spatiotemporal graphs into RNNs, enabling the capture of intricate, high-level structures involving both space and time. The results obtained from this novel technique have surpassed existing methods, whether in understanding human movements or interpreting interactions among objects. This significant progress represents a substantial leap forward, providing a potent tool to augment machine learning models and laying the foundation for more precise predictions and analyses in complex scenarios [40].

2.0.9 Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) units, a form of artificial intelligence, embody a neural network capable of storing, learning, and recalling patterns over prolonged periods. LSTMs use their data inputs to predict sequences, a mechanism that could bring significant advancements in speech recognition, natural language processing, and time series prediction. A groundbreaking use of LSTMs lies in vehicle safety, where real-time driver distraction detection becomes possible through analyzing long-term patterns in driving and head tracking data. With an impressive accuracy rate of up to 96.6 percent, LSTM-based approaches outdo traditional methods like support vector machines and show remarkable utility in handling time-series data. This notable application can lead to advanced vehicle safety systems development, improving road safety and further reflecting LSTM units' potential in enhancing AI applications [41].

2.0.10 Role of CNN and LSTM in Autonomous Vehicles

These two key components, CNN for feature extraction and LSTM for classification, operate together to deliver a real-time surveillance solution. The CNN, specifically MobileNet V2 in this study, is responsible for extracting spatial features from input images. The partial features are how pixels are arranged and related to each other in a two-dimensional layout. The extraction process ensures that the essential visual information is drawn from each video frame. Then, the extracted features are sequenced and passed to the LSTM network for classification. Equipped with the ability to manage sequential data and long-term dependencies, the LSTM can process image features over time. Using its gates to control how information flows, the LSTM can learn patterns across the sequence of frames. Thus, combining feature extraction via CNN and sequence pattern learning through LSTM enables the system to recognize and classify activities effectively in real-time video streams. The CNN-

LSTM approach has shown promising results in real-world applications, detecting suspicious activities at 10-13 frames-per-second in real-time under various conditions. This study combined CNN and LSTM on a Raspberry Pi, demonstrating the possibility of a self-contained system using these two technologies. The study also includes a human action recognition (HAR) methodology that combines CNN and LSTM for optimal speed and precision, demonstrating the significant potential for real-time applications. The HAR approach achieved remarkable accuracy, reaching up to 98 percent on the Peliculas dataset and 91 percent on complex real-life datasets with variable backgrounds, thus showcasing improvements over earlier techniques. This study's findings are particularly relevant for the research into vision systems within autonomous vehicle cabins using machine learning, highlighting the practicality and real-time capability of a combined CNN-LSTM model in recognizing and classifying driver behaviors [42, 43, 42].

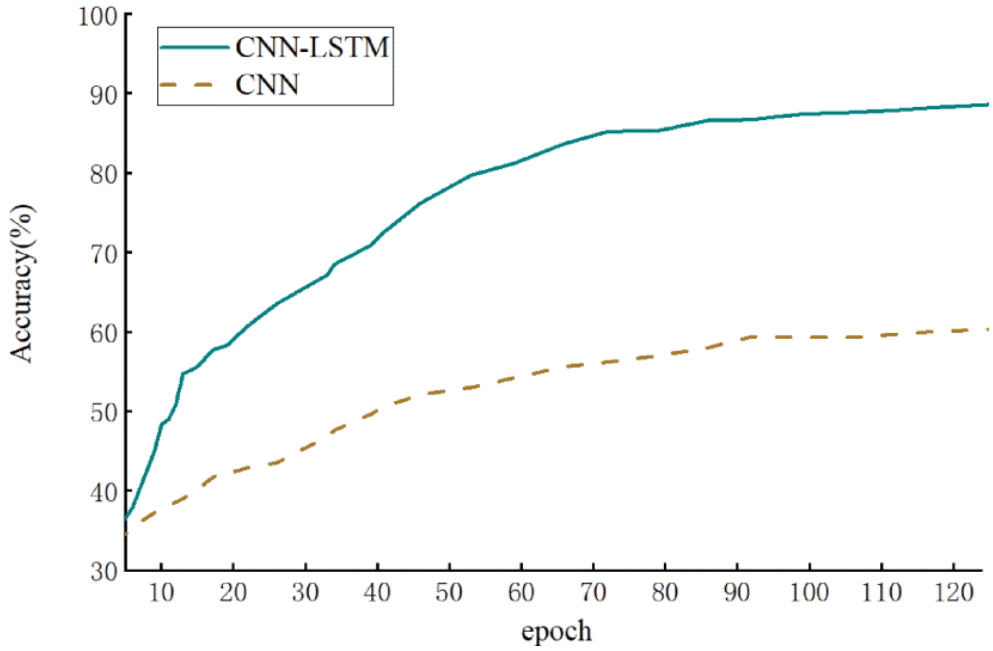


Figure 2.2: Comparative Accuracy of CNN vs. CNN-LSTM Models Over Training Epochs[4].

Building on CNN-LSTM algorithms in real-time video surveillance, similar strate-

gies surface in studies analyzing driver behavior in autonomous vehicles. From capturing smooth spatial patterns and fine-grained motion details to addressing scene and representation bias, these methods continue to enrich autonomous vehicle safety systems, demonstrating the innovative use of machine learning models in detecting driver behavior [44].

2.0.11 Bidirectional LSTMs (BiLSTM)

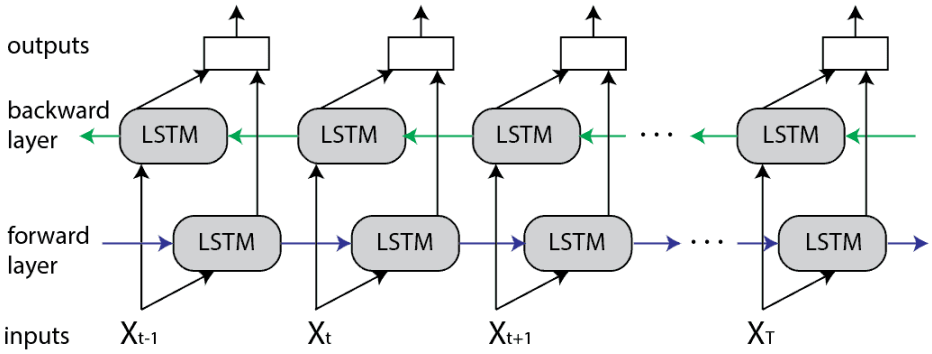


Figure 2.3: The Unrolled Bidirectional LSTM Structure [5].

Bi-LSTM extend the LSTM model by processing sequences in both forward and backward directions, enabling greater contextual understanding, as illustrated in Figure 2.3. This dual-path architecture allows BiLSTMs to capture not just the preceding elements in a sequence (as traditional LSTMs do) but also the succeeding elements. By doing so, BiLSTMs offer a more comprehensive understanding of the sequence's context, significantly improving the model's predictive accuracy for a wide range of applications, including speech recognition, text generation, and more complex time series analyses. The BiLSTM model has a unique structure. It integrates two separate LSTMs: one processes the input sequence in its original order, while the other processes it in reverse. The combined output of both LSTMs is then used to make predictions, offering a richer, more nuanced interpretation of sequential data compared to unidirectional LSTMs. This approach not only retains the advantages

of traditional LSTMs, such as handling long-term dependencies, but also introduces an additional layer of context sensitivity, thereby significantly advancing the field of sequential data analysis [45].

A study [46] conducted by Texa Tech University highlights that BiLSTMs, by processing data in both directions, significantly outperform unidirectional LSTMs in time series forecasting, improving accuracy by an average of 37.78%. Despite slower training and requiring more data batches, BiLSTMs' ability to capture complex sequential patterns justifies their selection over traditional LSTMs for advanced predictive modeling.

2.0.12 Time-Distributed Layer

Time-Distributed Layers wrap around existing layers. They apply these layers independently to each timestep in a sequence, ensuring consistent input-output dimensionality. By integrating Time-Distributed Layers, the model can discern subtle behavioral variations over long periods. This provides a detailed and comprehensive understanding of the subject [47].

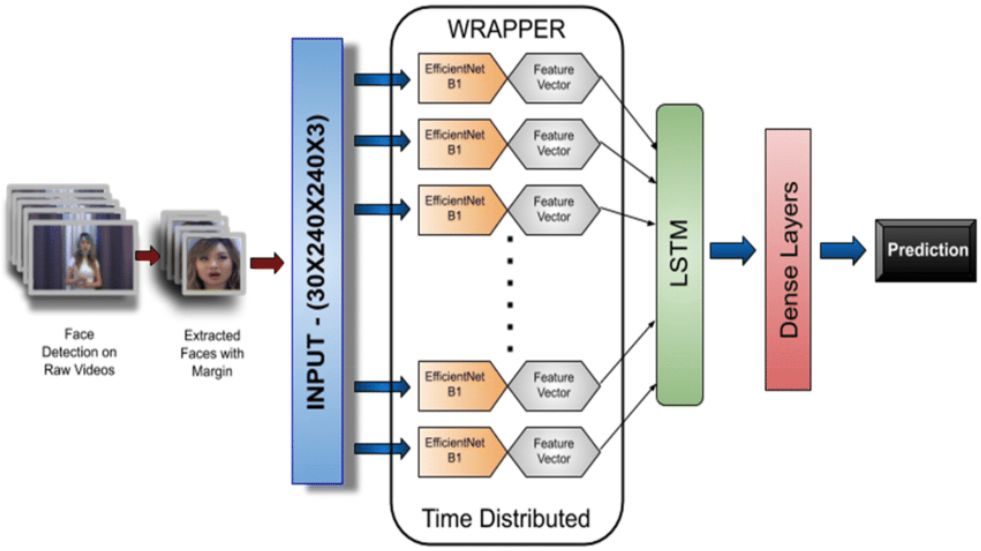


Figure 2.4: Detailed architecture with visualization of time-distributed layer [6].

2.0.13 Driver Behavior Classification

According to the World Health Organization (WHO) data for 2023, road traffic injuries emerge as a substantial global concern, accounting for approximately 1.3 million fatalities annually, with an additional 20 to 50 percent of cases resulting in non-fatal injuries that often lead to disabilities. It is noteworthy that over half of these road traffic deaths occur in low-income and middle-income countries, primarily affecting individuals with lower socioeconomic statuses who are more susceptible to being involved in such accidents[48].

In response to these alarming statistics, the National Highway Traffic Safety Administration (NHTSA) is actively addressing unsafe driving behaviors that are significant contributors to road accidents, injuries, and fatalities. These behaviors encompass drug-impaired driving, involving both alcohol and drugs; distracted driving, which includes activities like texting while driving; aggressive driving, characterized by behaviors such as tailgating and excessive speeding; drowsy driving; and the failure to use seat belts. These risky driving behaviors not only pose a grave threat to passenger safety but can also result in severe injuries and tragic fatalities [49].

Driver behavior classification is a pivotal aspect of studying human interaction with autonomous vehicles; it aims to distinguish the safe and risky driving behaviors. These behaviors are typically classified into two states in the datasets: safe and unsafe. "1" represents safe behaviors, and "0" denotes unsafe behaviors. This system utilizes binary classification. Safe behaviors include maintaining a steady speed, maintaining an appropriate gap from the vehicle in front, and regularly checking rear-view mirrors. In contrast, unsafe behaviors can involve texting, talking on the phone, operating the radio, drinking, reaching in back, doing hair and makeup, and talking to passengers. The causes of unsafe driving behavior are multifarious, ranging from driver fatigue and drowsiness to distraction and impairment due to substance use. Under specific circumstances, hostile driving or vehicular aggression could also lead to unsafe be-

haviors. Identifying the causes for unsafe driving is crucial to developing effective interventions and preventive measures. The classification of driver behaviors has the potential to indisputably enhance the protection of the driver. The classification can facilitate real-time monitoring and feedback, alerting drivers to potentially dangerous behaviors, and prompting corrective action. This classification additionally maintains the potential for long-term benefits, including providing driver education and training programs, as well as the ability to contribute to the design and development of enhanced, intuitive, and safer autonomous vehicle interfaces. Further, this classification also has the ability to feed into the development of advanced driver-assistance systems (ADAS), thereby enabling these systems to better understand and predict human behavior; this may also serve to prevent accidents.

In this context, the review presented in [33] offers an inclusive framework that addresses the regulation of autonomous vehicles and associated challenges, both in the United States and Europe. Regulatory issues, particularly those pertaining to the legal interpretation of driver behavior, present complex problems. However, the potential for autonomous vehicles to impressively enhance road safety in developed countries justifies these efforts. In anticipation of this emerging technology, countries in the European Union are being primed to adjust their legal structures, and they are collaborating with lawmakers and technical experts to establish unambiguous guidelines and practical solutions for optimization of the AVs on the road.

Despite the challenges posed by regulatory issues, driver behavior technology is making significant strides in understanding and improving the interaction between drivers and autonomous vehicles. Research in driver behavior technology now focuses on distraction and fatigue issues. Driver state-of-mind analysis involves a blend of self-reports, biological metrics, driving performance indicators, and hybrid methods. The latter combines multiple data sources for a clearer picture and shows more accurate results by cutting down on false alerts and keeping high-performance ratings [50].

Researchers have developed a new vision system using "vehicle dynamics data", which eliminates the need for cumbersome eye-tracking hardware. This approach with "support vector machines" (a machine learning model) has shown promising high classification rates, boosting the development of the next generation of autonomous driving aid systems. On another front, Kinect V2 sensors, commonly used in video gaming, have created a database of standard upper limb movements in healthy individuals. Initially proposed for rehabilitation, this method might help understand how drivers interact with vehicle controls. Understanding driving behavior has evolved with machine learning and deep learning models that draw on large-scale vehicle data. Deep learning methods, including neural networks, have shown high accuracy levels, signaling they may soon dominate driving behavior analysis as the technology progresses [51, 52, 53].

2.0.14 Hand Classification

In 2018, a recent incident involving a Tesla Model S striking a fire truck in "Autopilot mode," which is a system in vehicles that automates certain aspects of control, including maintaining speed and staying within lanes, with human supervision, highlights the danger of keeping hands off the wheel, even in autonomous vehicles. According to the National Transportation Safety Board (NTSB), the driver had his hands off the wheel for most of the trip, receiving multiple alerts to place his hands back on the wheel. The vehicle accelerated towards the driver-set cruise control speed and collided with the parked fire truck while the Autopilot system failed to detect the driver's hands on the wheel. This incident, along with previous fatal crashes involving Tesla vehicles, emphasizes the potential importance in hand classification in autonomous vehicles for drivers' safety [54, 55].

Hand classification in driver monitoring systems involves monitoring and analyzing a driver's hand movements while operating a vehicle. It holds a pivotal position

Ref.	Summary	Methodology	Relevance
[56]	Enhances action recognition using dense trajectories, improving understanding of physical actions in videos.	Dense trajectories	Enhances action recognition models for accurate action detection, relevant to driver behavior analysis in autonomous vehicles.
[57]	Combines temporal and spatial convolution in a new CNN model to learn spatiotemporal features from videos.	Spatiotemporal Multiplier Networks	Proposes Spatiotemporal Multiplier Networks (STMNs) for video data analysis within autonomous vehicle cabins, extracting important features for in-cabin analysis.
[58]	Uses EfficientNet, a highly efficient ConvNet family, to achieve state-of-the-art accuracy through systematic model scaling.	EfficientNet	EfficientNet's superior accuracy and efficiency are relevant for developing robust vision systems within autonomous vehicle cabins.
[59]	Utilizes MobileNets for efficient and lightweight deep neural networks design, suited for mobile and embedded vision applications.	MobileNets	MobileNets' efficient, lightweight architecture is suitable for real-time vision systems in autonomous vehicle cabins, minimizing computational resources.
[60]	Combines GoogLeNet and LSTM models to classify self-efficacy levels through human body gesture and movement recognition, achieving high accuracy.	CNN (GoogLeNet) and LSTM	Provides an effective approach to monitor and analyze driver behaviors, enhancing safety and efficiency within autonomous vehicles.
[61]	Uses a pre-trained Keras neural network to classify hand presence in a controlled hand-washing dataset, achieving perfect accuracy.	Neural Network using Keras	Demonstrates an effective approach for hand presence classification, potentially enhancing safety and efficiency by monitoring driver actions in autonomous vehicles.
[62]	Introduces a novel hard attention network for Driver Action Recognition (DAR), effectively recognizing driver behaviors in real-world conditions and reducing computational complexity.	Bidirectional LSTM (Bi-LSTM)	Investigates deep learning for driver behavior monitoring and action recognition, aligning with the goal of in-cabin analysis in autonomous vehicle cabins.
[63]	Uses a multi-camera framework for hand classification in driver monitoring systems, potentially enhancing traffic safety and reducing distracted driving.	RestNet CNN	Discusses a multi-camera framework for hand classification in driver monitoring systems, aligning with the topic of vision systems and machine learning analysis in autonomous vehicle cabins.
[64]	TPresents a CNN-based system for abnormal driving behavior recognition, emphasizing the importance of monitoring and preventing potential accidents caused by distractions.	CNN	Detects abnormal driving behaviors through physiological character classification using deep learning, contributing to understanding of vision systems for driver behavior analysis in autonomous vehicle cabins.

Table 2.1: Summary of research on Hand classification.

in evaluating the driver's behavior and ensuring their safety on the road. The classification system of hands provides valuable insights into the driver's actions and level of engagement by detecting whether the driver is holding the steering wheel properly, using turn signals appropriately, or reaching for objects within the vehicle. Deep learning models have emerged as promising approaches for recognizing specific hand actions and movements. In [65], with remarkable accuracy, a "pre-trained Keras Neural Network" was employed to classify hand presence: a pre-trained Keras Neural Network model was an already trained model on a large dataset; Keras, a Python Programming Language library, allows for the quickly building and testing these networks; Pre-trained, in this instance, means the model's initial weights come from an earlier training run, often from an embracing dataset like ImageNet. These pre-trained models already recognize common patterns, which can be adapted to new tasks, reducing training time and computational resources; Keras provides various pre-trained models, especially beneficial when the dataset is not large enough to train a whole network from the beginning. This model is able to distinguish between one hand on the wheel, two hands on the wheel, or no hands on the wheel. By utilizing this deep learning model and a carefully selected hand-classifying dataset comprising data from 30 volunteers, the system achieved an impressive 100 percent accuracy. Although the controlled nature of the dataset contributed to this high accuracy, it highlights the potential of deep learning models in accurately recognizing and classifying hand presence.

Another exploration area involves using a multi-camera framework for hand classification in driver monitoring systems. This approach enables more precise and thorough hand detection and tracking by utilizing multiple cameras mounted at different locations inside the cabinet strategically placed within the vehicle cabin. It enhances the system's ability to accurately classify hand movements and gestures, thereby seriously improving traffic safety and reducing accidents caused by distracted

driving. Multi-camera framework aligns with the broader objective of vision systems and machine learning analysis within autonomous vehicle cabins, ensuring a complete understanding of the hand classification and facilitating a more accurate hand classification. In summary, hand classification in driver monitoring systems is critical for assessing driver behavior and promoting safe driving practices. Deep learning models, including the pre-trained Keras Neural Network and RestNet CNN, demonstrate the potential for accurate hand presence classification [63].

A powerful approach to enhance on-wheel hand action recognition and prioritize driver safety is utilizing feature trajectories, which is the technique to track the path or progression of specific features or characteristics over time. Wang et al. propose a method that analyzes video actions using dense trajectories, which is a method in computer vision and video analysis that densely samples key points in a video sequence to capture motion information and track object movements. This approach efficiently evaluates hand motions and quick movements in hazardous circumstances, determining potential risks. Additionally, the implementation of convolutional neural network (CNN) models, including spatiotemporal multiplier networks (STMNs) introduced by Zolfaghari et al., emphasizes the importance of hand classification for driver safety. By combining temporal convolution with spatial convolution, STMNs offer an inclusive approach to analyzing spatiotemporal patterns in driver behavior, enabling the identification of unsafe driving actions hands-off-the-wheel [56, 58].

Moreover, advancements in efficient CNNs like EfficientNet and lightweight deep neural networks like MobileNets reinforce the significance of hand classification for in-cabin analysis and real-time monitoring, enhancing autonomous vehicles' safety and efficiency. The utilization of complex attention networks, as presented in Driver Action Recognition (DAR), further underscores the importance of hand classification for driver safety. By focusing exclusively on vital behavioral elements including the hand and head, this approach aligns perfectly with the analysis of data inside of

the cabinet, ensuring an exhaustive understanding of driver behavior. The studies above indicate that more than relying solely on hand classification for driver behavior detection is required as it fails to capture the full spectrum of driver behavior and intention. While hand movements provide valuable insights into driver actions, an all-inclusive understanding requires considering factors including head position, eye gaze, and body posture. Incorporating multiple factors leads to a better understanding of a driver's cognitive state, attention level, emotional response, and fatigue level [59, 66].

As technology advances, the focus on comprehending the holistic range of driver behavior has led to integrating these multiple factors into driver assessment models. Notably, the area of "Hand Classification" has seen significant breakthroughs, revolutionizing our grasp of driver interactions. Recent developments in the field of hand classification present intriguing innovations using machine learning and computer vision for understanding hand gestures and driver behavior. One study created an algorithm that tracks a driver's right hand and ear in real-time, processing video frame images to identify if the driver is distracted. With an impressive accuracy score of 74 percent, this algorithm can classify various actions, including everyday driving, touch screen interaction, and phone conversations. Another study formulated an algorithm that identifies hand gestures using three specific characteristics of a hand's shape, achieving 91 percent classification from a test set of 200 images [67, 68, 69].

Researchers in the realm of sign language innovated a system that interprets gestures by thinning a segmented image resulting in a communication breakthrough for sign language users. Further studies rolled out an Urdu Alphabet translation recognition system with an accuracy rate of 97.4 percent, proving extremely helpful for individuals with vocal and hearing disabilities. To improve human-computer interaction, a trailblazing system recognizes hand gestures as an alternative to classical mouse and keyboard inputs. This system uses the AdaBoost algorithm to identify the hand in a video feed and then applies multi-class support vector machines to

understand the gesture [70, 71, 72].

In a similar vein, an approach for recognizing moving hand shapes was developed. Keeping the focus on real-time image processing, researchers first extracted the hand region and then identified the hand's shape. In an effort to boost secure access, a hand image-based identification system was developed, achieving confident recognition in groups of about 500 people. Finally, the creation of a detailed video-based dataset stands as a pioneering venture for hand detection in varied driving settings. This dataset, encompassing various backgrounds, lighting conditions, users, and viewpoints, serves as a potent tool for fine-tuning machine learning algorithm performance. Notably, it also features annotations that offer detailed hand related insights, marking significant advancements in the field of hand classification [73, 74, 75].

2.0.15 Facial Classification

As the global economy rapidly expands, the transportation sector is also swiftly advancing. Specifically, heavy trucks stand out for their impressive cargo capacity and have become crucial in logistics and road transportation. In China, with 2022 sales projected at 1.2 million units, the country's total heavy truck ownership will reach 11.7 million by 2025. However, this growth has resulted in a corresponding rise in traffic incidents, often due to drowsy driving. Fatigue is notably prevalent among these drivers who undertake extensive drives to make ends meet, leading to exhaustion, decreased attention, and potential accidents. Data from the US National Highway Traffic Safety Administration (NHTSA) indicates that 91,000 crashes involved drowsy driving, leading to approximately 50,000 injuries and nearly 800 deaths. The traffic safety, sleep science, and public health communities generally agree that these figures underestimate the true impact of drowsy driving. Additionally, a Chinese study highlighted the propensity for such incidents to occur at any time, especially during early morning or mid-afternoon hours. These statistics underscore the serious threat posed

by drowsy driving, particularly with heavy trucks, marking it as a significant factor in critical traffic accidents. In recent years, there have been driver fatigue monitoring systems with facial detection technology and precise infrared sensors developed to effectively identify signs of driver drowsiness. The system monitors facial movements and detects subtle fatigue indicators, including increased blinking, drooping eyelids, or prolonged eye closures, often unnoticed by the drivers. The detection system remains unaffected by external variables, including the time of day, the presence of glasses, or reflective light. Enhanced by integrated pre-trained artificial intelligence with advanced facial recognition capabilities, the system operates even without a WiFi connection due to its inbuilt algorithms. The AI system is programmed to recognize and audibly alert drivers about signs of fatigue or distraction, providing real-time and reliable fatigue monitoring. This deployment of facial detection technology significantly boosts road safety by adeptly assessing drivers' drowsiness levels and reducing the risks associated with drowsy driving [76, 77, 78].

Facial Classification entails detecting and analyzing a driver's facial features and expressions, including yawning, blinking, or looking away from the road. Such analysis can offer insights into the driver's level of attention and alertness, which are vital for ensuring responsible driving. One facial classification approach employs the FaceNet system to efficiently carry out facial recognition, clustering, and verification. The Euclidean Embedding method simplifies complex data by representing it more linearly while keeping the critical relationships intact through a convolutional neural network (CNN) to tackle facial recognition challenges. By examining feature vectors, the FaceNet system can provide solutions that enhance facial recognition accuracy under various conditions. Another multiple-resolution cascade network method combines different layers with varying levels of detail to efficiently process and extract features from complex data, with high discriminative capabilities. This CNN Cascade technique uses a sequence of convolutional neural networks to progressively filter and

Ref.	Summary	Methodology	Relevance
[79]	Utilizes feature vectors in FaceNet for efficient face recognition, clustering, and verification tasks.	CNN-based Euclidean embedding	Explores facial recognition, presenting solutions to pose and illumination issues, relevant for enhanced biometric systems in autonomous vehicles.
[80]	Presents a multi-resolution cascade network to handle pose, expression, and lighting issues in facial recognition.	CNN Cascade with discriminative capabilities	Introduces a discriminative cascade system for efficient facial distinctions analysis, pertinent to vehicle cabin surveillance.
[81]	Explores appearance-based gaze estimation in non-lab conditions using the MPIIGaze dataset with 213,659 images from 15 participants.	Appearance-based gaze estimation	Investigates gaze estimation under everyday conditions, contributing to improved facial feature recognition within autonomous vehicle cabins.
[82]	Detects driver emotions unobtrusively via smartphone-captured contextual features, outperforming facial recognition by 7 percent and ensuring privacy.	YOLOv5, Microsoft Face Recognition, DeepLabV3, OpenCV	Unobtrusive Sensor Feed Pipeline analyzes driver emotions less intrusively, relevant to vision systems in autonomous vehicle cabins.
[83]	Proposes a hazardous driving image classification system using a modified ShuffleNet model, balancing speed and accuracy for real-time monitoring.	ShuffleNet	Proposes a solution for dangerous driving behavior monitoring, enhancing safety within autonomous vehicle cabins.
[84]	Suggests a deep-learning-based system for drowsiness detection using a novel CNN model for eye state classification, enhancing traveler protection.	CNN - Deep Driver Drowsiness Detector (4D) Model	Uses a deep-learning-based system for drowsiness detection via a novel CNN model, important for driver state analysis in autonomous vehicle cabins.
[85]	Offers a non-invasive approach for driver vigilance classification via deep learning HyMobLSTM model and transfer learning, analyzing facial and eye components.	HyMobLSTM model (MobileNetV3 and LSTM)	Contributes to driver behavior analysis inside autonomous vehicle cabins via a non-invasive vision system, improving safety and alertness monitoring.
[86]	Presents Hypo-Driver, a real-time driver inattention and fatigue detection system using multi-view cameras and biosensors, outperforming existing solutions.	Hypo-Driver system: fused through CNN, RNN-LSTM, and DRNN	Hypo-Driver system employs multimodal features for driver hypovigilance detection, aligning with vision-based systems in autonomous vehicles for safety enhancement.
[87]	Monitors driver behavior using image processing and computer vision techniques to prevent accidents, promising high accuracy and real-world application potential.	OpenCV, Support Vector Machine (SVM)	Describes real-time driver monitoring using computer vision techniques, relevant for understanding how such techniques can enhance safety and behavior analysis in autonomous vehicle cabins.

Table 2.2: Summary of research on Facial classification.

refine object detection. The result addresses challenges associated with pose, expression, and lighting. This cascade system uses a sequence of classifiers to progressively refine and improve the classification accuracy of an object or pattern recognition

task. In addition to these approaches, appearance-based gaze estimations further augment facial recognition in everyday situations. By concentrating on real-world scenarios, this method enhances the recognition of facial features and contributes to a more accurate evaluation of a driver's attention and alertness levels. Moreover, driver emotion detection systems have been explored by analyzing different types of information or characteristics surrounding a particular subject or situation, which are considered together to gain a more comprehensive understanding. One research experiment utilizes advanced machine learning models, including YOLOv5, Microsoft Face Recognition API classifier, VGG13-based image classifier, DeepLabV3 semantic segmentation, and OpenCV. The innovative unobtrusive sensor feed pipeline (USFP) developed in this research provides a less intrusive method for analyzing driver emotions inside an autonomous vehicle's cabin, contributing significantly to developing vision systems for autonomous vehicles [88, 89, 90].

A hazardous driving classification system based on a modified ShuffleNet lightweight model has been proposed. This system effectively reduces model complexity and increases operational speed without compromising classification accuracy, making it a potential solution for real-time monitoring of dangerous driving. Similarly, a deep-learning-based drowsiness detection system has been proposed using a novel CNN model to classify eye states. The HyMobLSTM model presents a non-intrusive method putting emphasis on analyzing facial features and eye localization in order to yield a more comprehensive interpretation. This model determines a driver's alertness by categorizing it into five levels based on head orientation and the eye position relative to the eyelids. Transfer learning extracts additional features from the driver's eyes, serving as input vectors for the LSTM network [62, 63].

Another real-time driver inattention and fatigue detection system, Hypo-Driver, utilizes multi-view cameras and biosignal sensors to extract hybrid features. The Hypo-Driver system uses a combination of CNNs, RNNs, and deep residual neural

networks (DRNN). This system achieves a high accuracy rate of 96.5 percent and outperforms other top-rated driver fatigue detection systems. This is achieved by extracting multimodal features and using deep learning models for driver's decreased alertness levels in individuals, often through the analysis of behavioral or physiological indicators. In addition to the above, another project uses OpenCV, an open-source computer vision library that provides tools and functions for image and video processing, as well as support vector machine (SVM), a machine learning algorithm used for the classification and regression tasks. The described systems above contribute to understanding how computer vision and machine learning techniques can enhance safety and behavior analysis in drivers, thereby improving the overall safety measures within autonomous vehicles [87, 91].

Building on the use of machine learning and computer vision for driver monitoring through hand classification, researchers have expanded into the realm of facial classification. This advancement is opening new avenues for improving driver safety through cutting-edge recognition and emotion perception technologies. One study explored how blocking facial features affects emotion perception, while another mapped facial features to emotional recognition. Significant strides were made in real-time driver distraction detection by analyzing visual cues from the face and tracking eye and head positions [92, 93, 94].

Furthermore, an AdaBoost algorithm-based system calculates gaze direction to assess if drivers maintain eye contact with the road. Another study improved distraction detection accuracy to 81.1 percent by analyzing eye activities and driving performance data. In contrast, others employed facial cues to develop highly precise classifiers for visual and cognitive distractions [95, 96, 97].

Researchers also optimized convolutional neural networks and introduced an unified face detection system through the wearable face recognition system, a notable development for blind and visually impaired individuals. The rapidly advancing

field of facial classification is creating breakthroughs in autonomous driving, human-computer interaction, and communication aids for the visually impaired [98].

In summary, these Facial Classification methodologies have been put into practice to advance facial classification tasks. These strategies provide a comprehensive approach to monitoring and analyzing driver behavior inside autonomous vehicle cabins, enhancing safety measures and contributing to the development of autonomous vehicles.

2.0.16 Body Posture Classification

Driving a vehicle demands prolonged periods of intense focus and repeated sitting posture and movements. These factors inevitably cause fatigue. When the driver experiences fatigue, their ability to maintain focus diminishes, and their reaction times may suffer, posing potential safety risks. Therefore, it is crucial to devise methods that can promptly and accurately assess the level of vehicle drivers' fatigue. This assessment should be conducted to ensure operational safety and efficiency without interfering with the driver's routine tasks. Body posture classification involves analyzing a driver's body posture and movements, including slouching, leaning, or sudden jerky movements. The result can provide insights into the driver's fatigue, distraction, or impairment level.

Firstly, in [107], researchers from China have found a deep learning technique: by extracting features related to upper body posture, including the head, neck, chest, shoulders, and arms, from images captured of train drivers to detect drivers' fatigue level. In [108], the researchers from Beijing Jiaoton University introduced a method for detecting the fatigue state of drivers by analyzing their upper body postures extracted by OpenPose framework and a Deep Belief Network - Back Propagation Neural Network ("DBN-BPNN") model. The model takes a "9-dimensional principal eigenvector" of the driver's upper body posture as input: the 9-dimensional principal

Ref.	Summary	Methodology	Relevance
[99]	Proposes an ensemble model using deep convnets for Human Body Posture Recognition (HBPR), relevant for posture-based analysis in autonomous vehicle cabins.	Bivarial Deep Convnet Model	Uses deep convnets for human body posture mapping, relevant for individual passenger analysis in autonomous vehicle cabins.
[100]	Classifies seven common driving activities using pre-trained CNN models, contributing to driver behavior analysis within autonomous vehicles.	AlexNet, GoogLeNet, ResNet50	Classifies driving activities and body postures using CNNs, contributing to driver behavior analysis in autonomous vehicles.
[101]	Proposes a drone surveillance system for human behavior analysis, including posture analysis via OpenPose, relevant for outdoor surveillance and interaction with autonomous vehicles.	Pose Estimation, OpenPose, DeepSort, YOLO	Analyzes and classifies body postures and behaviors using multiple algorithms, applicable for passenger behavior analysis in autonomous vehicles.
[102]	Utilizes PoseNet for real-time 6-DOF camera relocalization from single RGB images, relevant for vehicle cabin monitoring and driver pose estimation.	23-layer deep CNN (convolutional neural network) trained in an end-to-end manner to regress the 6-DOF camera pose.	Handles challenging lighting conditions and motion blur, contributing to the development of robust vision systems for in-cabin analysis in autonomous vehicles.
[103]	Uses MoveNet to predict subject-specific joint angle profiles for different walking conditions, applicable for pedestrian behavior analysis around autonomous vehicles.	MoveNet	Uses MoveNet's user-specific prediction capabilities from minimal input data, showcasing potential for individualized passenger analysis in autonomous vehicles.
[104]	Presents D3-Guard, a real-time drowsy driving detection system using built-in smartphone audio devices and LSTM networks, applicable for drowsiness monitoring in autonomous vehicle cabins.	LSTM networks	D3-Guard, an acoustic sensor-based system for drowsiness detection, offers insights for vision systems in autonomous vehicles, emphasizing alertness monitoring and machine learning techniques like LSTM networks.
[105]	Introduces BiRSwinT network for fine-grained driver behavior recognition, offering enhanced driver action learning and accuracy in driver behavior analysis.	Bilinear full-scale residual Swin-Transformer network (BiRSwinT)	BiRSwinT's approach to fine-grained driver behavior recognition contributes to driver behavior analysis in autonomous vehicle cabins, enhancing detection of subtle behaviors.
[106]	Proposes a driver distraction detection system using a blend of deep learning and machine learning models, relevant for enhancing roadway safety through distraction monitoring.	DenseNet and Genetic Algorithms (GA)	The real-time driver distraction detection via a Hybrid Genetic Deep Network aligns with the objectives of in-cabin driver behavior analysis in autonomous vehicles.
[90]	TLeverages MobileNetV2 for efficient classification of driver distraction behaviors, demonstrating potential for reducing accidents caused by distracted driving in autonomous vehicles.	MobileNetV2	Uses MobileNetV2 for driver distraction classification, providing valuable insights for machine learning-based vision systems in autonomous vehicle cabins.

Table 2.3: Summary of research on Body Posture classification.

eigenvector is like the main road on a map with nine different directions, which provides the most efficient route to capture the essential features in a dataset or system.

Next, the model applies a forward Restricted Boltzmann Machine (RBM) learning algorithm to reconstruct the eigenvector and extract high-level distribution features. The DBN-BPNN model includes four levels for classifying fatigue states. Results from the experiment demonstrate an average detection accuracy of 92.7 percent using the DBN-BPNN model, indicating the method's high accuracy in detecting fatigue among drivers. MoveNet, furthermore, is a deep neural network designed to predict subject-specific joint angle profiles for various walking speeds and slopes, minimizing input data requirements. MoveNet's ability to predict highly user-specific profiles from minimal input data shows the potential for using similar approaches in vision systems analyzing the interior of autonomous vehicle cabins. By understanding and adapting to individual passengers' needs and preferences, MoveNet can contribute to a more personalized and comfortable ride in autonomous vehicles.

In fact, Beijing Institute of Technology researchers introduced D3-Guard, a system that detects driver drowsiness in real-time using the audio capabilities of a smartphone. It identifies unique sound patterns from behaviors like yawning and steering and uses long short-term memory (LSTM) networks for efficient detection. With an average accuracy of over 93 percent in real-world testing, D3-Guard suggests that sound-based detection can complement or even replace vision-based systems in self-driving cars. The scaling method of the system preserves the original aspect ratio of images or videos during resizing, ensuring high-quality output. Furthermore, another study proposes the Residual Swin-Transformer (BiRSwinT), a network that recognizes ten fine-grained driver behaviors. BiRSwinT employs a dual-stream structure to process and analyze various data types, performing exceptionally well on the AUC V1 and V2 datasets. This dual-stream design allows for the simultaneous processing

of global and local cues of driver actions, enhancing the detection of subtle behaviors and improving the overall safety of autonomous vehicles [104, 106, 83].

Another highly accurate system for detecting driver distraction consists of a blend of deep learning and machine learning models, fine-tuned by a genetic algorithm. This system adapts to new datasets in real-time, aiming to enhance traffic precautions through the Hybrid Genetic Deep Network. This model uses principles from genetic algorithms and deep neural networks. It utilizes evolutionary techniques to optimize the architecture and processes of deep learning networks. This approach aims to enhance performance or efficiency when studying driver behavior within self-driving vehicle cabins [90].

Furthermore, the research from National Tsing Hua University (NTHU) [39] underscores the proficiency of the MobileNetV2 model in categorizing driver activities, achieving an impressive blend of speed and precision while preserving low computational demands – an essential characteristic for mobile system implementation. The research leveraged two distinct datasets for their experiment: a 10-class dataset from State Farm and a 2-class dataset. The clearly defined features in the State Farm dataset allowed the model to successfully differentiate between two classes, resulting in superior predictive accuracy. However, the NTHU drowsiness dataset, in its realistic depiction of driver behavior, offered a more authentic training environment, fostering progress toward real-world applications. In the context of mounting traffic fatalities worldwide, specifically in areas like Malaysia where distraction-induced accidents are prevalent, the application of deep learning techniques, specifically convolutional neural networks presents a promising avenue for efficient identification and classification of distracted driving behavior. Therefore, it contributes to the broader objective of promoting safety in autonomous vehicles [85].

While machine learning techniques, particularly convolutional neural networks, promise swift identification and classification of distracted driving behavior, innova-

tions extend beyond this realm to improve driver safety. These advanced systems now consider other vital parameters, including head and body movements, to gauge a driver's alertness, paving the way for comprehensive driver behavior assessment. In the pursuit of creating safer roads, advanced systems are analyzing drivers' alertness, including their head and body movements, to prevent accidents due to fatigue or distraction.

Procedures including the integration of the Microsoft Kinect range camera's capabilities of capturing and analyzing 3D shapes of drivers, and fitting a human skeleton model to this data have been beneficial in evaluating nuanced driving behaviors across varying demographics. When combined with machine learning techniques like K-means clustering, SVMs, and HMMs, the result is a highly accurate recognition of driving-related actions and postures [109, 110].

Algorithms like Part Affinity Fields (PAFs) have proven efficient in detecting 2-dimensional poses of multiple individuals in images, setting a benchmark for pose detection. The use of tools like head trackers and vision-based foot behavior analysis, along with video sequence trajectories, is enhancing the accuracy in action recognition and prediction of drivers' foot behavior. This data contributes to body posture classification as a significant component of autonomous vehicle safety. By detecting diversions or measuring the driver's head orientation, these advancements promise a safer future for autonomous driving [111, 112].

In summary, these body posture classification methodologies contribute to the advancement of safer driving. These strategies collectively provide a comprehensive approach for monitoring and analyzing driver behavior inside autonomous vehicle cabins, contributing to the development thereof and safer roads worldwide.

2.0.17 Integration of Physiological Indicators Classification

Incorporating physiological indicators, including hand gestures, facial expressions, and body postures, is essential for developing an all-in-one driver monitoring system. This section outlines strategies for the physiological classification approaches to establish a cohesive driver monitoring system. To further elaborate on the integration of the classifications, it is important to note that driver behavior classification serves as a key component in predicting and preventing risky driving scenarios. Not only can driver behavior classification provide real-time assistance to human drivers, but it is also instrumental in shaping the development of autonomous vehicles. It is worthwhile to recognize that human interaction with self-driving cars is an emerging research area with profound implications for road safety, traffic efficiency, and overall driving experience. The classification schema that divides driving behaviors into safe and unsafe categories offers a practical and simplified representation of the complexities involved in everyday driving. This classification system is the basis for analyzing and predicting driver behavior. By assigning a binary value of "1" for safe driving behaviors, and "0" for unsafe driving behaviors, researchers can create a streamlined and consistent method of collecting and analyzing data. This data, in turn, provides valuable insights that can help develop various interventions to enhance road safety.

In [113], researchers shed light on the significance of different body parts, in the perception of emotions. While not directly addressing the process of integrating hand, facial, and body posture classifications, the research provides valuable insights for developing an emotion detection system. The study highlights the importance of different body parts in accurately perceiving emotions, suggesting that emotion classification is an essential component in a multi-modal system for a comprehensive analysis. By incorporating these insights, a unified framework can be developed that accounts for the importance of hands, employs a multi-modal approach, harnesses shared mechanisms, and addresses challenges including the body inversion effect,

leading to the creation of a robust and accurate system for emotion recognition.

Another study [114], conducted by Chinese researchers, emphasizes the integration of deep learning-based segmentation to isolate the driver's body parts, including the head and hands, which play critical roles in identifying distraction. Two segmentation architectures, Human Body Parts Segmentation (HBPS) and Cross-Domain Complementary Learning (CDCL), were investigated. Despite similar performance on the Pascal VOC dataset, the CDCL model performed significantly better under low light conditions in the study's specific dataset, efficiently segmenting critical body parts even in challenging lighting scenarios. This model facilitated the elimination of irrelevant image regions and concentrated on hands and head-related regions essential for safe driving. The system achieved an impressive average accuracy of over 96 percent on the authors' dataset and 95 percent on the public AUC dataset, indicating its substantial potential in developing comprehensive driver assistance systems by integrating physiological indicators for driver behavior classification.

The AWAKE (System for Effective Assessment of Driver Vigilance and Warning According to Traffic Risk Estimation) project, adopted by the European Union, emphasizes the importance of combining driver state and performance measures to detect driver fatigue effectively. The project endeavors to demonstrate the viability of driver vigilance monitoring systems, considering both technological and non-technical aspects. It employed mainly the driver state measures, including the eyelid movement and changes in steering grip, and driver behavior metrics like lane tracking, usage of accelerator and brake, and steering position. These measures were then compared against traffic risk evaluations derived from digital navigation maps, anti-collision devices, driver gaze sensors, and odometer readings. The project's output is a set of design guidelines for evaluating driver vigilance and warning signals, which, despite leaving many research questions unanswered, are likely to influence the future implementation of fatigue detection devices significantly [115].

The proposed method in [116] integrates detection and tracking algorithms to monitor distracted driving behavior based on facial and hand movements. The facial detection and tracking involve using the Viola-Jones algorithm to detect the driver's face and an algorithm described in reference [117] to detect key facial features like eyes, lips, and forehead. The centroids of the forehead and lips are tracked using the KLT tracker algorithm. Hand detection focuses on a localized search region, typically the lower-right or lower-left quarter of the frame, and employs a hand detection algorithm from reference [118]. The centroid of the hand is tracked using the KLT tracker as well. The tracking algorithms continuously estimate the displacement between consecutive centroids and calculate the tracking error based on feature differences. If the tracking error exceeds certain thresholds, reinitialization is performed by redetecting the respective body part. The method emphasizes the significance of simultaneous tracking of these body parts in capturing distracted driving behaviors. By analyzing the trajectories and patterns of facial and hand movements, specific distracted driving behaviors including talking on the phone, eating, or texting can be recognized. This proposed method leverages the simplicity and effectiveness of the algorithms, taking into account the constrained setting of driving and marginal deformations of body parts. The integration of two physiological indicators together, hand gestures and facial expressions, enhances the understanding of distracted driving behaviors and contributes to the development of comprehensive driver monitoring systems.

This study [120] investigates how a driver's body and head characteristics can influence the categorization of driving tasks, beginning with evaluating depth information from facial landmarks and joints. The precision of task classification demonstrated substantial differences when relying exclusively on either head or body signals. The model, trained only with two-dimensional information like head rotation and joint coordinates, showed accuracy levels comparable to those trained with complete features. However, the classification accuracy decreased when only using head

Method	Year	Dataset	Feature	Algorithm	Accuracy
[113]	2023	Bochum Emotional Stimulus Set	Full body, particularly the hand	Isolated Body Part Emotion Recognition algorithm	64.7%
[117]	2022	ORL Face Database from ATT and FEI dataset	Face and Hand	SVM	98.03%
[114]	2021	Driver Monitoring Dataset and AUC	Full body	Two different pre-trained CNNs, VGG-19 and Inception-v3	96%
[119]	2019	QVGA ToF image sequences	Body Key Points	3D CNN-LSTM	85%

Table 2.4: Comparison of combined classifying models.

pose information. While the distracted driving behaviors were successfully detected, it was challenging to differentiate safe driving behaviors with similar head positions. In other words, using only body features (the coordinates of the hand, wrist, elbow, and shoulder joints) resulted in weaker detection of mirror-checking behaviors but a higher degree of accuracy for detecting distraction behaviors. So, the head and body characteristics are vital for comprehensively classifying driving tasks. Though there was a slight dip in the overall detection accuracy, the selection of 18 features, which includes yaw, pitch, roll, nose, hand, and shoulder coordinates, provided a reasonable balance between accuracy and computational speed. The final result demonstrates

the potential of such a system in effectively combining physiological indicators together for the classification of driver behaviors using the unification of various body part classifications.

In the study referenced as [119], the researchers built a model of the driver's posture classification consisting of nine key points – left/right shoulders, left/right elbows, left/right hands, left/right hips, and the right knee. They selected a combination of various body parts for their perceivability in denoting driving behaviors. The team trained fully convolutional neural networks using their dataset to calculate the pivotal points for each frame of the body independently. Then, they transposed the data into three-dimensional camera coordinates using depth imagery, resulting in a real-time 3D rendering of the driver's physical stance. The focus is shifted from individual actions to the real-time tracking of a combination of various body parts, thereby enhancing the depth and precision of physiological indicator-based classification of driver behaviors.

While research like [119] illustrates how real-time 3D imaging and tracking of various body parts can enhance driver behavior analysis, the potential benefits of autonomous vehicles extend beyond improved safety measures. These breakthroughs not only revolutionize transportation policies and systems but also necessitate a thorough understanding of the legal regulations governing autonomous vehicles. This understanding is crucial to adapt the existing road traffic laws and navigate the regional differences in these regulations. Autonomous vehicle technologies potentially lower transportation costs and increase accessibility, particularly for those with mobility limitations. With a focus on the communication between autonomous vehicles and infrastructure, opportunities arise to develop efficient routing systems. Such technologies can revolutionize transportation policies. Meanwhile, in the U.K., connected and autonomous vehicles are triggering a transformative change in the economy, promising benefits like improved safety, reduced congestion, and increased productivity.

Vital innovation and research capabilities in the U.K. automotive sector help leverage these benefits efficiently. From a legal perspective, a professional understanding of autonomous vehicles' legal regulations is crucial because it aids the discussions related to modifying existing road traffic laws and the navigation of the variations in the regulations across different regions, including the U.S. and Europe. In vision-based human action recognition or labeling image sequences, varied advancements focus on image representation and the subsequent classification process. Despite current challenges and limitations, these advancements uncover potential areas for further exploration and improvement. Finally, the accuracy in recognizing subtle driver behaviors can improve significantly by using a network like BiRSwinT. This network combines global shape appearances and local discriminative cues of driver actions in its structure, effectively identifying multi-scale, local lines and can help drive future research in recognizing driver behaviors [121, 122].

Developing a proficient and effective driver monitoring system requires the concurrent examination and integration of numerous physiological indicators. The singular analysis of distinct body parts, including facial expressions, hand movements, and overall postures, can indeed yield meaningful insights into a driver's actions. However, this individualized focus might need to look into the larger, more complex picture of driving behavior due to the multifaceted nature of human actions and responses. By integrating the findings from the analysis of various body parts, researchers can achieve a more complete understanding of a driver's behavior. This comprehensive view enables them to design more precise, well-rounded interventions and assistance systems. The full-body analysis holds the potential to uncover nuanced and complex driving behaviors, enhancing the capability to predict and prevent risky scenarios that might otherwise remain undetected. In summary, the core of driving behavior is not encapsulated solely in the isolated movements of the hands or the face. Instead, it is embodied in the complex interactions among all body parts. As the future is ap-

proaching, marked by autonomous vehicles and sophisticated driver-assist systems, a comprehensive, full-body analysis becomes increasingly significant in promoting safer and more efficient roads. In the light of this, focusing on the interplay of full-body indicators becomes a crucial step toward a future characterized by increasing safety and efficiency in driving.

2.0.18 Pose Monitor AI Program

The Pose-Monitor AI Application is an open-source project that evaluates the user's body posture and provides real-time feedback to improve posture. The system utilizes image processing techniques to distinguish between proper and improper postures, generating a score based on the evaluation. If the rating falls beneath a pre-established limit after 30fps, the system warns the user; if the score remains below the threshold after another 30fps, it alerts the user to adjust their posture, using either a familiar voice or a more severe tone if necessary. Incorporating the Pose-Monitor AI Application into the Proposed AI system allows for adequate assessment of a driver's posture and behaviors in an autonomous vehicle. This integration enables the AI system to leverage existing image processing techniques and real-time feedback mechanisms to understand the driver's overall condition better. Consequently, the Proposed AI system can detect potential indications of fatigue, distraction, or impairment, ultimately improving safety in autonomous vehicles. Additionally, the open-source nature of the Pose-Monitor AI Application ensures that the AI system remains customizable and adaptable, permitting developers to continuously refine the algorithms and techniques used to analyze a driver's body posture. This adaptability is crucial for tailoring the AI system to address the specific needs of various autonomous vehicle manufacturers and user groups.

With the potential to be added based on the hand classification and facial classification algorithms [116], the Pose-Monitor AI Application emerges as a practical

solution, leveraging real-time feedback of the driver's driving state. The technology's continuous monitoring at a rate of 30 frames per second ensures rapid detection of any shifts in the driver's physiological states or overall behavior. These changes could indicate the onset of fatigue or distractions, triggering the system to alert the driver or activate autonomous controls for enhanced safety.

The functionality of the Pose-Monitor AI Application extends beyond simple posture monitoring. It uncovers insightful behavioral indicators tied to the driver's physical disposition. For instance, drivers' subtle body adjustments can signal anxiety or unease with the autonomous vehicle's decisions. This understanding of drivers' behavior empowers the system to respond to drivers' needs proactively. When integrated with other AI modules, like emotion detection [123], the Pose-Monitor AI-based Application contributes to a sweeping human behavior classification system. This integrated system can offer a more accurate and in-depth understanding of the driver's state, considerably improving the autonomous vehicle's interaction with its human occupants.

The performance of the Pose-Monitor AI Applications in driver behavior analysis sets the stage for exploring cutting-edge machine learning and AI systems aiming to foster road safety. Distinct research approaches form an intriguing landscape of advanced tools; these range from real-time warning systems and neural network-based activity recognition to novel hand pose estimation using 3D Convolutional Neural Networks. Delving into the individual contributions of these unique studies reveals their remarkable potential in expanding the scope of AI and machine learning in understanding driver behavior and thus enhancing autonomous vehicle safety. The proposed AI Monitor Program makes extensive use of machine learning and AI systems to enhance road safety. An activity recognition system based on deep CNNs successfully identifies seven common driver activities. Four of these activities are classified as normal driving tasks and the rest as distractions, achieving an impressive 91.4

percent accuracy rate. Another approach uses CNN to develop a real-time warning system for driver distraction detection. A unique approach inhabits the use of cellular neural networks and capacitive sensors on the steering wheel for real-time stress level monitoring, improving detection accuracy by up to 92 percent [124, 125, 126].

Understanding driving behaviors like car-following, lane-changing, and risky driving can be improved using sensor data, onboard vehicle computer data, and feature extraction methods. Deep-learning models have shown exceptional accuracy in identifying these behaviors, hence, implying their potential as a primary tool for understanding driver behavior. There also exists a new approach for real-time hand pose estimation using 3D CNNs, which enhances real-time monitoring of human activity. An open-source tool, VGG Image Annotator (VIA), which operates in any web browser, provides an efficient way to manage labeled data required for AI systems. Finally, a unique application tracks gaze direction to guide an automated surveillance system and represents a novel approach in AI surveillance. Each research study offers unique insights and significant contributions to autonomous vehicle safety, extending the potential use of AI and machine learning in understanding driver behavior [127, 128, 129].

In conclusion, the open-source nature of the Pose-Monitor AI Application opens up avenues for customization. Developers can adapt and enhance the system to cater to specific requirements, creating a versatile AI that fits various vehicle models, driving conditions, and user preferences. This ability to customize is pivotal in a rapidly evolving landscape of autonomous vehicle technology. Furthermore, the Pose-Monitor AI Application supports data-driven adjustments. Over time, accumulated posture data can guide refinements in AI algorithms, fostering a more nuanced understanding of human behavior patterns. This continuous learning process boosts the system's overall performance and safety measures. Incorporating the Pose-Monitor AI Application into a vision system for autonomous vehicles, as part of a broader machine

learning-based approach, can yield safer and more interactive autonomous driving experiences.

CHAPTER 3

METHODOLOGY

3.0.1 Leveraging CNNs and BiLSTMs for Distraction Detection

This research leverages the power of Convolutional Neural Networks to identify patterns within images and Bidirectional Long Short-Term Memory networks to understand time-based relationships in driver behavior. Data augmentation is used to expand data variability. Various CNN architectures (InceptionV3, ResNet, MobileNetV3, EfficientNet) are tested with and without BiLSTM, for a comparative analysis of their effectiveness. Performance is evaluated using accuracy and test loss. We employ a CNN+BiLSTM architecture to analyze complex patterns in distracted driving behavior. We introduce a novel 'Union Dataset,' created by combining the strengths of the AUC and State Farm datasets, to provide a more comprehensive training data source. We utilize three datasets seperately: the AUC dataset (12,537 images), the State Farm dataset (22,458 images), and a Union Dataset (34,995 images) for comprehensive training and evaluation.

3.0.2 Data Augmentation in Enhancing Model Generalization

To strengthen model robustness and prevent overfitting, we employ data augmentation techniques. These image transformations simulate diverse driving conditions, artificially expanding our dataset and enhancing model generalizability for real-world scenarios. We apply random rotations, shifts, shear transformations, zooms, and horizontal flips to the Union Dataset's images, uniformly resized to 224x224 pixels.

3.0.3 Experimental Setup

After enhancing our dataset with data augmentation, we followed these steps in our experimental setup:

1. **Data Augmentation:** As detailed in the previous section.

2. **Train the Model:** We trained the models using the augmented datasets, with a focus on capturing the complex behavioral indicators of distraction associated with distracted driving behaviors.

3. **Test the Model:** The trained models were then evaluated using a separate set of images to accurately assess their performance in detecting distracted driving behaviors.

4. **Plot Experiment Results:** We plotted the results from the testing phase to visualize the models' efficacy across different scenarios.

5. **Compare with Different Models and Datasets:** We conducte performance comparisons between various model configurations and datasets to identify the most effective approach.

6. **Draw Experimental Conclusion:** Insights and conclusions were drawn based on comparative analysis, which will guide future research directions.

3.0.4 Model Configurations

We explore a variety of configurations, including:

- **InceptionV3 + BiLSTM** - This setup aims to reduce grid size and employing label smoothing, in conjunction with BiLSTM to analyze temporal patterns.

- **ResNet + BiLSTM** - Utilizes ResNet's capability for residual learning together with BiLSTM to recognize long-term dependencies.

- **MobileNet + BiLSTM** - Leverages MobileNet for its efficiency, paired with BiLSTM, rendering it suitable for real-time applications.

- **EfficientNet + BiLSTM** - Combines EfficientNet's scalable architecture with BiLSTM to improve processing of temporal information.

For each configuration, we assess the impact of integrating BiLSTM layers. This analysis help us leverage the unique strengths of CNN architectures in detecting distracted driving behaviors.

3.0.5 Assessing Performance

Performance is evaluated based on accuracy and test loss as primary metrics. These metrics indicate the models' ability to accurately identify distracted driving behaviors. Our analysis of existing models using diverse datasets demonstrates their effectiveness in handling a wide range of driving scenarios, while variations in test loss offer insights into each model's reliability and prediction confidence. This assessment is essential in highlighting the specific advantages of each setup, facilitating the development of more precise models for distracted driving detection.

3.0.6 Challenges and Considerations

Our research faces several challenges, including finding the appropriate CNN models, troubleshooting the code during the experimental phase, and selecting the model's dataset. Our research acknowledges several challenges, including the representativeness of the Union Dataset and the potential for models' complexity to induce overfitting. To address these issues, we employ a wide range of data augmentation, rigorous validation, and careful data collection and labeling practices to mitigate biases and promote dataset diversity.

3.1 DATASETS: STATE FARM AND AUC-DDD, AND RATIONALE FOR MERGING

To comprehensively evaluate the effectiveness of the chosen approach, we leverage the strengths of three complementary datasets: the State Farm Dataset, the AUC-DDD Dataset, and the Union Dataset. The latter combines the first two, creating a comprehensive resource for assessing distracted driving.

3.1.1 State Farm Dataset

Originating from the State Farm Distracted Driver Detection competition, this dataset contains 22,458 annotated images showcasing various distracted driving behaviors. It is a key asset for identifying and understanding unsafe driving actions across diverse scenarios.

3.1.2 AUC Dataset

Developed through a collaborative research initiative between the American University in Cairo, the Technical University of Munich, and Valeo Egypt, the AUC Distracted Driver Dataset includes 12,537 annotated images. This dataset expands the research scope by including diverse driving behaviors not found in the State Farm Dataset, enhancing our ability to understand and detect distractions [7].

3.1.3 Union Dataset

The Union Dataset emerges from the integration of the State Farm and AUC-DDD Datasets, totaling 34,995 annotated images. This merged dataset aims to provide a more comprehensive representation of distracted driving behaviors by leveraging the strengths of both datasets. It encompasses a wide variety of distractions, demographic variables, and driving conditions, serving as a vital tool for developing and evaluating AI models in driver behavior detection.

The datasets are categorized into ten classes representing different driving behaviors, ranging from attentive driving to various distractions. The datasets contain distracted behaviors, such as using a phone, engaging with the entertainment system, personal grooming, and interaction with passengers. This categorization facilitates a detailed analysis of driver behavior, aiding in the creation of AI models adept at identifying and classifying a broad array of unsafe driving actions.

Our analysis of driver distraction detection models suggests that existing methods may not fully capture real-world complexities, leading to decreased accuracy.



Figure 3.1: Ten Classes of Driver Distraction training example from AUC-DDD [7]

Each dataset categorizes ten behaviors into ten classes (C0 to C9):

1. C0: Safe Driving – Full attention is on driving.
2. C1: Texting (Right Hand) – Engaging in texting with the right hand.
3. C2: Calling (Right Hand) – Making a call with the right hand.
4. C3: Texting (Left Hand) – Engaging in texting with the left hand.
5. C4: Calling (Left Hand) – Making a call with the left hand.
6. C5: Radio Interaction – Adjusting the radio settings.
7. C6: Drinking – Consuming a beverage.
8. C7: Reaching Back – Reaching for something in the back seat.

9. C8: Grooming – Engaging in personal grooming.

10. C9: Conversing – Talking to a passenger.

Our research employs a refined 10-class classification system to facilitate the development of CNN-based driver distraction detection models. This system is designed to recognize a wide spectrum of driver behaviors. As illustrated in Figure 3.1, this analysis employs a two-category classification system to distinguish between safe and unsafe driving behaviors for in-depth investigation. Only Category C0 is considered safe, whereas the other categories (C1 to C9) signify unsafe driving behaviors.

3.2 ARCHITECTURAL DESIGN: CNN, TIME DISTRIBUTED LAYER, AND BILSTM

This analysis employs Time Distributed Layers, CNNs, and BiLSTMs to extract both spatial and temporal information from driving action sequences. This integrated approach offers a deeper understanding of driver behavior compared to earlier methods, particularly for tasks like distraction detection where sequential context is critical.

We explore various CNN architectures, including InceptionV3, ResNet, MobileNetV3, EfficientNet, MoveNet, and DenseNet. MoveNet's reduced accuracy and DenseNet's higher processing cost made them less suitable for our real-time application, where both speed and precision are essential. This exploration guide us towards the following architectures, chosen for their balance of efficiency and accuracy:

- **InceptionV3+BiLSTM:** Enhances feature extraction through InceptionV3's efficient design and adds BiLSTM for analyzing how driving actions change over time.

- **ResNet+BiLSTM:** Leverages ResNet's residual learning framework to facilitate the training of deep networks, which is paired with BiLSTM to capture long-term dependencies.

- **MobileNetV3+BiLSTM:** MobileNetV3's efficient design for real-time analysis on mobile devices and adds BiLSTM for understanding temporal patterns.
- **EfficientNet+BiLSTM:** Takes advantage of EfficientNet's scalable architecture, which uniformly scales all dimensions of the model, integrating BiLSTM to effectively process temporal information.

Time Distributed Layers are employed in all these models to maintain consistent dimensionality for inputs and outputs across sequences, enabling the network to process and produce outputs for each timestamp. This capability is crucial for capturing nuances and variations in driver behavior over time, providing a deeper and more comprehensive understanding.

3.3 EXPERIMENTAL SETUP: DATA PREPARATION, TRAINING, AND EVALUATION METRICS

Our experiment leverage the powerful computational capabilities of a MacBook Air (M2 chip, 16GB RAM, 512GB SSD) to accelerate model development and analysis. Python, TensorFlow, and Keras provide the essential tools for data handling, image processing, and model building. We chose Visual Studio Code for its robust Python support.

The following steps outline the preparation of our development environment:

3.3.1 Preliminary Setup

Library Imports

To support the various aspects of driver behavior detection, including data handling, image processing, and machine learning tasks, several libraries were imported:

- **Data and Visualization**: `pandas` for data manipulation, `numpy` for numerical operations, and `matplotlib.pyplot` & `seaborn` for data visualization played a

key role in in analyzing and presenting the data.

- **Image Processing**: `opencv-python (cv2)` and `ImageDataGenerator` were used for image preprocessing and augmentation, preparing the data for model training in a way that boosts the model's ability to generalize from the training data.

- **Deep Learning**: TensorFlow and TensorFlow Hub were essential for accessing a wide range of pre-trained models and deep learning functionalities. Models such as `ResNet50`, `InceptionV3`, `EfficientNet`, `MobileNetV3-Small` were investigated for their potential application in detecting driver behavior through pose estimation.

3.3.2 Model Selection for Pose Estimation

For this experiment focusing on driver behavior detection, we selecte MoveNet, a TensorFlow Lite pose estimation model known for its strong performance in predicting human joint locations from RGB images. MoveNet is available in two variants, Lightning and Thunder, catering to different performance and accuracy needs. Lightning is designed for faster, real-time applications on limited hardware, making it suitable for quick assessments. On the other hand, Thunder, chosen for this experiment, offers higher accuracy in pose estimation, crucial for the detailed analysis required in detecting driver behaviors. This distinction allows us to tailor the experiment to our specific requirements for precision and computational efficiency.

Establishing a Streamlined Workflow

Efficient project organization is crucial for maintaining a streamlined workflow. The structure includes a dedicated directory for the TensorFlow example scripts, enhancing the project with valuable resources for pose estimation and ensuring an organized system for managing scripts, models, and data.

3.3.3 Environment Configuration

After importing the necessary libraries, we optimize the training dataset's file path. This streamlined the use of external scripts for pose estimation, enhancing our driver behavior analysis.

CHAPTER 4

RESULTS AND ANALYSIS

4.1 PERFORMANCE COMPARISON: BASELINE MODELS V.S. PROPOSED MODEL

Model	Dataset	Test Accuracy (%)	Test Loss	Training Time (s/epoch)
ResNet + BiLSTM	Union	98.83	0.0468	1429
ResNet	Union	98.6	0.1784	1154
InceptionV3 + BiLSTM	Union	99.6	0.014	678.5
InceptionV3	Union	99.6	0.022	657.5
EfficientNet + BiLSTM	Union	83.02	0.529	810
EfficientNet	Union	12.62	2.66	1214
MobileNetV3 + BiLSTM	Union	12.61	2.405	229
MobileNetV3	Union	22.01	3.178	232.5

Model	Dataset	Test Accuracy (%)	Test Loss	Training Time (s/epoch)
ResNet + BiLSTM	AUC	97.76	0.0898	391
ResNet	AUC	99.04	0.0263	450.5
InceptionV3 + BiLSTM	AUC	64.8	1.441	254.5
InceptionV3	AUC	57.17	2.376	255
EfficientNet + BiLSTM	AUC	23.22	2.271	355
EfficientNet	AUC	17.69	3.509	733
MobileNetV3 + BiLSTM	AUC	9.89	2.366	108.5
MobileNetV3	AUC	10.78	2.759	100.5

Model	Dataset	Test Accuracy (%)	Test Loss	Training Time (s/epoch)
ResNet + BiLSTM	State Farm	99.02	0.0479	773.5
ResNet	State Farm	99.24	0.0348	746
InceptionV3 + BiLSTM	State Farm	99.06	0.0314	412.5
InceptionV3	State Farm	99.24	0.0348	383
EfficientNet + BiLSTM	State Farm	52.58	1.764	481
EfficientNet	State Farm	20.05	2.85	802.5
MobileNetV3 + BiLSTM	State Farm	9.45	2.313	131
MobileNetV3	State Farm	10.56	3.457	123.5

Figure 4.1: Comparative Performance of Deep Learning Models on Union, AUC, and State Farm Datasets

This study performs a comprehensive evaluation comparing the performance of baseline CNN models with a CNN-BiLSTM architecture incorporating temporal analysis. The baseline architectures include InceptionV3, ResNet, MobileNet, and EfficientNet, each renowned for their unique feature extraction capabilities and operational efficiencies in image recognition tasks.

This analysis investigates the integration of BiLSTM layers with CNN architectures to enhance temporal data analysis, a crucial aspect for understanding the sequence-dependent nature of distracted driving behaviors.

The comparison is based on several performance metrics, including testing accuracy and testing loss across the AUC Dataset, State Farm Dataset, and the Union Dataset.

Table 4.1: Performance on Union Dataset

Model	**Test Accuracy (%)**	**Test Loss**	**Time (s/epoch)**
ResNet + BiLSTM	98.83	0.0468	1429
ResNet	98.6	0.1784	1154
InceptionV3 + BiLSTM	99.6	0.014	678.5
InceptionV3	99.6	0.022	657.5
EfficientNet + BiLSTM	83.02	0.529	810
EfficientNet	12.62	2.66	1214
MobileNetV3 + BiLSTM	12.61	2.405	229
MobileNetV3	22.01	3.178	232.5

Table 4.1 provides compelling evidence for the value of both BiLSTM layers and the Union Dataset in enhancing distracted driving detection. The integration of BiLSTM layers leads to considerable improvements in accuracy across all models, a trend particularly pronounced on the diverse and comprehensive Union Dataset. This underscores the critical role of temporal analysis – where the BiLSTM's ability

to interpret sequences of images markedly improves detection capabilities – in understanding complex driving behaviors that evolve over time. The Union Dataset's rich variety further amplifies model generalization capabilities, demonstrating the significance of comprehensive training data for real-world application. Notably, the ResNet+BiLSTM configuration emerges as a standout performer, achieving consistently high accuracies and underscoring the effective synergy between advanced feature extraction and temporal pattern recognition.

Table 4.2: Performance on AUC Dataset

Model	**Test Accuracy (%)**	**Test Loss**	**Time (s/epoch)**
ResNet + BiLSTM	97.76	0.0898	391
ResNet	99.04	0.0263	450.5
InceptionV3 + BiLSTM	64.8	1.441	254.5
InceptionV3	57.17	2.376	255
EfficientNet + BiLSTM	23.22	2.271	355
EfficientNet	17.69	3.509	733
MobileNetV3 + BiLSTM	9.89	2.366	108.5
MobileNetV3	10.78	2.759	100.5

While the Union Dataset demonstrates clear benefits, the AUC Dataset reveals challenges and the potential for dataset-specific adaptation. Table 4.2 reveals the challenges and complexities introduced by dataset variability. Despite the integration of BiLSTM layers, models demonstrate lower overall accuracies on the AUC Dataset compared to the Union Dataset. This highlights the influence of dataset characteristics on model robustness and emphasizes the need for further exploration of dataset-specific factors. The contrast with the Union Dataset findings underscores the importance of diverse training data for maximizing accuracy. Interestingly, the AUC Dataset shows greater efficiency gains when BiLSTM layers are integrated,

suggesting potential benefits for deployment on resource-limited devices. Future research should focus on domain-specific dataset adaptation or fine-tuning strategies to address the limitations revealed by the AUC Dataset and fully realize the potential of BiLSTM across diverse driving scenarios.

Table 4.3: Performance on State Farm Dataset

Model	**Test Accuracy (%)**	**Test Loss**	**Time (s/epoch)**
ResNet + BiLSTM	99.02	0.0479	773.5
ResNet	99.24	0.0348	746
InceptionV3 + BiLSTM	99.06	0.0314	412.5
InceptionV3	99.24	0.0348	383
EfficientNet + BiLSTM	52.58	1.764	481
EfficientNet	20.05	2.85	802.5
MobileNetV3 + BiLSTM	9.45	2.313	131
MobileNetV3	10.56	3.457	123.5

The findings from Table 4.3, regarding the State Farm Dataset, further corroborate the results obtained with the Union Dataset, demonstrating the substantial and consistent benefits of integrating BiLSTM layers within CNN frameworks across a variety of datasets. This alignment signifies the broad applicability of temporal analysis, made possible by BiLSTM, in detecting distracted driving behaviors within diverse driving contexts. The ResNet+BiLSTM model stands out, offering both speed and accuracy, which highlights its potential for real-world mobile deployment. This balance between computational efficiency and detection efficacy is crucial for embedded systems within vehicles. The consistency of improvement across datasets underscores the transformative potential of temporal analysis. Future research should focus on model compression techniques and smaller BiLSTM variants to further optimize this balance between BiLSTM's benefits and the need for efficient real-time operation.

This would significantly advance the integration of these sophisticated methods into automotive safety systems, paving the way for advancements in autonomous driving solutions.

The results show a marked improvement in model performance with the addition of BiLSTM layers. For instance, while the ResNet baseline model achieve commendable accuracies, the ResNet+BiLSTM configuration outperformed the standalone ResNet model across all datasets. Specifically, the ResNet+BiLSTM model reachs near-perfect accuracies, significantly higher than the baseline ResNet model, particularly on the Union Dataset. This shows the cooperative effect of combining advanced feature extraction with temporal pattern recognition.

This comparative analysis reveals the superiority of the proposed CNN+BiLSTM model over the baseline models, highlighting the value added by integrating BiLSTM layers. The results demonstrates the importance of considering temporal dynamics alongside spatial features in detecting distracted driving behaviors, a critical aspect that baseline CNN models alone fail to address adequately.

4.2 IMPACT OF THE BILSTM LAYER AND UNION DATASET

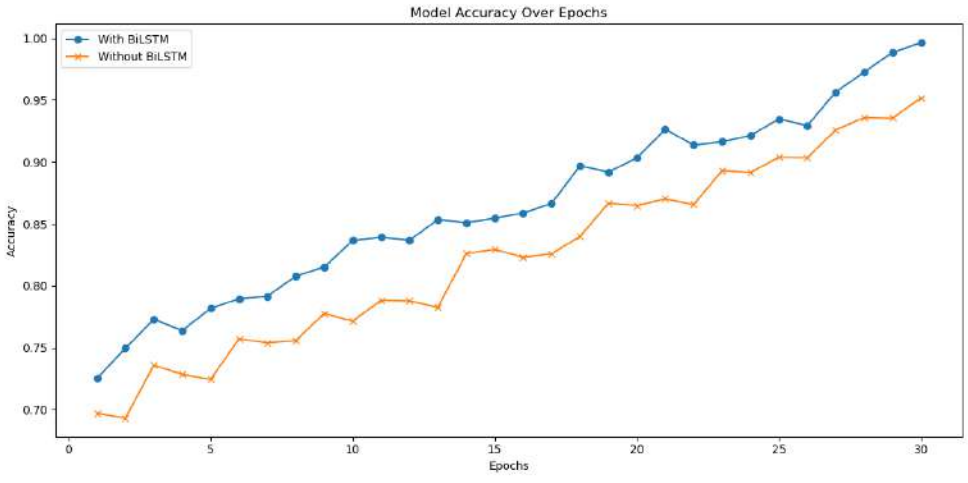


Figure 4.2: Test Accuracy trends for CNN models with versus without BiLSTM

The integration of BiLSTM layers within CNN architectures demonstrably enhances the models' ability to capture and analyze temporal sequences of driving behavior. BiLSTM layers enable the model to process not just individual images but sequences of images, thereby recognizing patterns over time, which is fundamental in distinguishing between various forms of distracted driving.

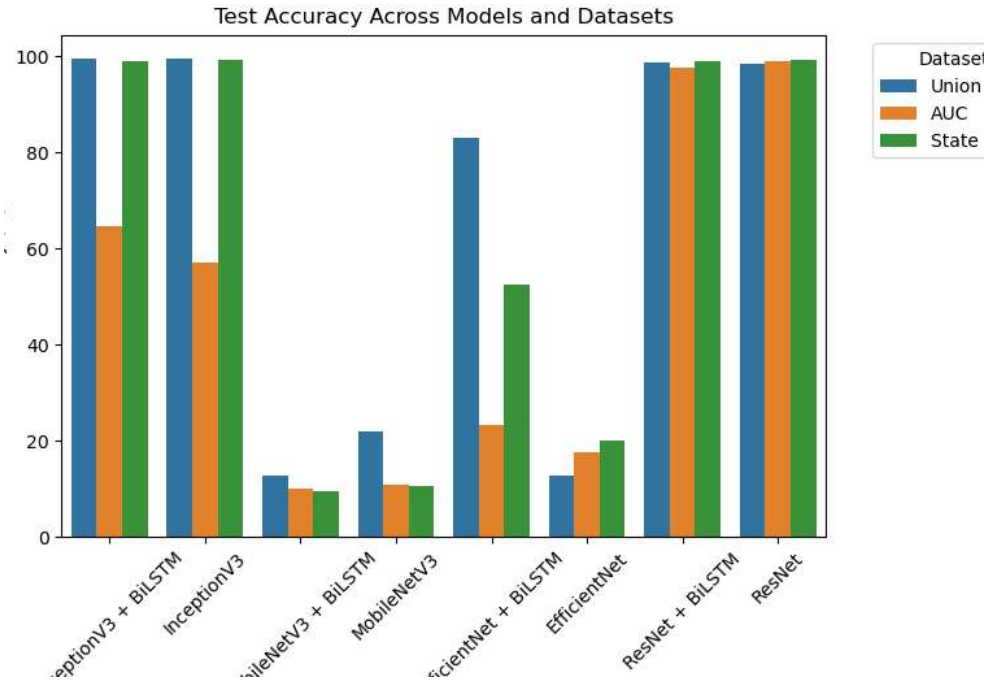


Figure 4.3: Test Accuracy Across Models and Datasets

Analysis of performance metrics, as shown in Figures 4.2 and 4.3, highlights the effectiveness of a BiLSTM layer for modeling temporal patterns in driver behavior. The models equipped with BiLSTM layers consistently demonstrate superior accuracy and lower loss rates compared to their counterparts without BiLSTM. This improvement is attributable to the BiLSTM's capability to capture long-term dependencies and sequential patterns that are common in distracted driving scenarios.

Moreover, the Union Dataset plays a pivotal role in enhancing the models' performance. By amalgamating the AUC and State Farm datasets, we create a more diverse and comprehensive dataset, covering a wider range of distracted driving be-

haviors across different demographics and driving conditions. This diversity is instrumental in training more robust and generalizable models. The Union Dataset's diversity and volume offer a valuable training environment for analyzing the effectiveness of CNN+BiLSTM architectures. This rich dataset potentially enhances the models' ability to generalize to unseen scenarios, as suggested in Figure 5.1.

The synergy between the BiLSTM layers and the Union Dataset is evident in the models' improved performance metrics. Models trained on the Union Dataset with BiLSTM integration exhibit higher predictive accuracy, showcasing the combined effect of diverse training data and advanced temporal analysis. This indicates that both the BiLSTM layer and the use of a comprehensive Union Dataset are crucial for developing highly effective models for detecting distracted driving behaviors, thereby contributing to the advancement of driver monitoring systems in autonomous vehicles.

CHAPTER 5
DISCUSSION

5.1 ANALYSIS OF THE BILSTM LAYER'S EFFECTIVENESS

Our analysis demonstrates that integrating a BiLSTM layer enhances model performance, as depicted in Figure 4.2. This advancement highlights the value of analyzing temporal sequences for understanding distracted driving behaviors that change over time. The disparity in performance between models with and without the BiLSTM layer, highlighted in our results, confirms the hypothesis that temporal dependencies play a crucial role in accurately identifying distracted driving patterns.

5.2 CONTRIBUTIONS AND LIMITATIONS OF THE UNION DATASET

Our analysis underscores the critical role of the Union Dataset in enhancing model generalization. Figures 4.3 and 5.1 demonstrate how its diverse scenarios improve model performance on unseen data. Its diverse range of scenarios and behaviors provides a comprehensive training environment, leading to higher accuracy and robustness in real-world applications. However, despite its contributions, the Union Dataset also presents limitations, such as the increased computational demand and training time, which could hinder scalability and efficiency in some contexts.

5.3 COMPARISON WITH EXISTING LITERATURE

Our research builds upon existing literature emphasizing the importance of combining spatial and temporal data analysis for detecting complex behaviors such as distracted driving. The higher performance metrics achieved when integrating a BiLSTM layer

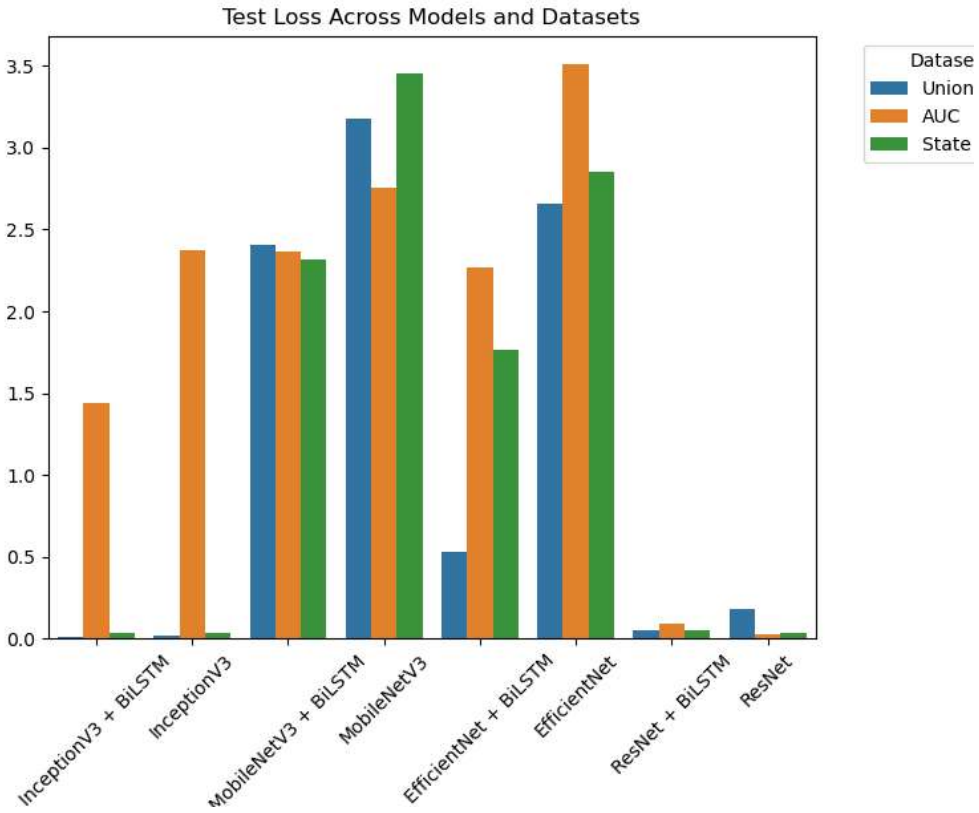


Figure 5.1: Training Time Efficiency Across Models and Datasets

with CNN architectures, particularly on the Union Dataset, underscore the potential benefits of temporal analysis for distraction detection.

This analysis shows significant performance gains with the ResNet+BiLSTM architecture compared to a standalone ResNet-50 baseline. On the State Farm dataset, it achieved 87.92% accuracy [130] aand on the AUC-DDD dataset, 87.7% accuracy [131]. Our analysis demonstrates that the EfficientNet+BiLSTM configuration yields higher accuracies on these datasets compared to previously reported results for standalone EfficientNet models. These comparisons highlight the value of the BiLSTM layer in enhancing model performance.

While the Union Dataset improved model generalization, it's important to acknowledge potential trade-offs, such as the increased computational demand and training time (Figure 5.1). Our work extends current research by demonstrating

the effectiveness of combining BiLSTM layers with a diverse dataset, offering advancements in accuracy for real-world autonomous vehicle safety systems.

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

6.1 SUMMARY OF FINDINGS

This research conclusively demonstrates that integrating BiLSTM layers greatly improves the performance of CNN models in detecting distracted driving behaviors. The Union Dataset's comprehensive nature has further augmented the models' effectiveness, enabling superior generalization capabilities compared to using single-source datasets. Notably, the ResNet+BiLSTM model emerged as the standout configuration, offering high accuracy and stability across diverse testing scenarios.

6.2 IMPLICATIONS FOR AUTONOMOUS VEHICLE SAFETY AND AI

Improving distracted driving detection holds the potential to directly save lives by enabling AI systems to prevent accidents caused by inattentive drivers. The implications of our findings extend to the broader context of autonomous vehicle safety and artificial intelligence. By offering tools to improve the accuracy and reliability of distracted driving detection, this research has the potential to support the development of safer, more intelligent autonomous driving systems. The integration of temporal data analysis through BiLSTM layers represents a significant step forward in understanding and mitigating driver distractions, a critical factor in preventing accidents and improving road safety.

6.3 SUGGESTIONS FOR FUTURE RESEARCH

In our ongoing research efforts, we are building upon the foundation laid by our recently published survey paper, *Comprehensive Study of Driver Behavior Monitoring Systems Using Computer Vision and Machine Learning Techniques*. Our initial work highlighted a limitation in existing distracted driving datasets, such as State Farm and AUC-DDD, due to their unrealistic camera angles. To address this, we are developing a unique dataset featuring realistic camera positions and a broader range of distraction behaviors (e.g., yawning, prolonged eye closure). We anticipate this dataset will significantly improve the real-world adaptability of distraction detection models in autonomous vehicles.

Additionally, we are exploring the Vision Transformer, a novel machine learning model. By fine-tuning it for distracted driving detection, we seek to gain deeper insights into driver states and enhance detection accuracy. Furthermore, we propose a binary classification approach (attentive vs. unattentive driving) to streamline the detection process and improve detection rates without sacrificing accuracy [132].

Building upon the insights of our initial study, our ongoing research addresses limitations and explores new avenues for applying AI techniques to enhance autonomous vehicle safety.

BIBLIOGRAPHY

[1] Z. Shen, "Distracted Driver Detection Project using PyTorch." `https://github.com/Followb1ind1y/Distracted-Driver-Detection-Project`, 2024. Accessed: 2024-03-28.

[2] TensorFlow Hub, "Movenet: An introduction to the model and deployment." `https://www.tensorflow.org/hub/tutorials/movenet`, 2023. Accessed: 2024-03-28.

[3] J. W. et al., "Cnn explainer." Web, 2023. Accessed: 2023-09-12.

[4] Y. Cai, R. Zhao, H. Wang, L. Chen, Y. Lian, and Y. Zhong, "Cnn-lstm driving style classification model based on driver operation time series data," *IEEE Access*, vol. 11, pp. 16203–16212, 2023.

[5] Y. Zhao, "Understanding rnn, lstm, and bidirectional lstm." `https://dagshub.com/blog/rnn-lstm-bidirectional-lstm/`, 2023. Accessed on 2023-02-20.

[6] A. Singh, A. S. Saimbhi, N. Singh, and M. Mittal, "Deepfake video detection: a time-distributed approach," *SN Computer Science*, vol. 1, no. 4, p. 212, 2020.

[7] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, "Real-time distracted driver posture classification," *arXiv preprint arXiv:1706.09498*, 2017.

[8] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 596–614, 2011.

[9] F. Qu, N. Dang, B. Furht, and M. Nojoumian, "Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques," *Journal of Big Data*, vol. 11, no. 1, p. 32, 2024.

[10] National Center for Statistics and Analysis, "Distracted driving in 2021," tech. rep., National Highway Traffic Safety Administration, May 2023. Research Note. Report No. DOT HS 813 443.

[11] J. M. Mase, P. Chapman, G. P. Figueredo, and M. T. Torres, "A hybrid deep learning approach for driver distraction detection," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1–6, IEEE, 2020.

[12] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 12, p. 1330, 2017.

[13] J. Mafeni Mase, P. Chapman, G. P. Figueredo, and M. Torres Torres, "Benchmarking deep learning models for driver distraction detection," in *Machine Learning, Optimization, and Data Science: 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020, Revised Selected Papers, Part II 6*, pp. 103–117, Springer, 2020.

[14] Centers for Disease Control and Prevention, "Youth Risk Behavior Surveillance System," Feb. 2024. Accessed 12 February 2024.

[15] P. Cañas, J. D. Ortega, M. Nieto, and O. Otaegui, "Detection of distraction-related actions on dmd: An image and a video-based approach comparison.," in *VISIGRAPP (5: VISAPP)*, pp. 458–465, 2021.

[16] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, M. N. Moustafa, *et al.*, "Driver distraction identification with an ensemble of convolutional neural networks," *Journal of Advanced Transportation*, vol. 2019, 2019.

[17] J. Wang, Z. Wu, F. Li, and J. Zhang, "A data augmentation approach to distracted driving detection," *Future internet*, vol. 13, no. 1, p. 1, 2020.

[18] S. Bagloee, M. Tavana, M. Asadi, *et al.*, "Autonomous vehicles: challenges, opportunities, and future implications for transportation policies," *Journal of Modern Transportation*, vol. 24, pp. 284–303, 2016.

[19] S. Tolbert and M. Nojoumian, "Cross-cultural expectations from self-driving cars," *Preprint (Version 1) available at Research Square*, 2023.

[20] J. Craig and M. Nojoumian, "Should self-driving cars mimic human driving behaviors?," in *3rd International Conference on HCI in Mobility, Transport and Automotive Systems (MobiTAS)*, LNCS 12791, pp. 213–225, Springer, 2021.

[21] S. Shahrdar, C. Park, and M. Nojoumian, "Human trust measurement using an immersive virtual reality autonomous vehicle simulator," in *2nd AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, pp. 515–520, 2019.

[22] S. Shahrdar, L. Menezes, and M. Nojoumian, "A survey on trust in autonomous systems," in *Computing Conference (CC)*, pp. 368–386, Springer, 2018.

[23] C. Park and M. Nojoumian, "Social acceptability of autonomous vehicles: Unveiling correlation of passenger trust and emotional response," in *4th International Conference on HCI in Mobility, Transport and Automotive Systems (MobiTAS)*, LNCS 13335, pp. 402–415, Springer, 2022.

[24] C. Park, S. Shahrdar, and M. Nojoumian, "EEG-based classification of emotional state using an autonomous vehicle simulator," in *10th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 297–300, 2018.

[25] M. Nojoumian, "Adaptive driving mode in semi or fully autonomous vehicles," 2022. US Patent 11,221,623.

[26] M. Nojoumian, "Adaptive mood control in semi or fully autonomous vehicles," 2021. US Patent 10,981,563.

[27] K. Kirkpatrick, "Still waiting for self-driving cars," *Communications of the ACM*, vol. 65, no. 4, pp. 12–14, 2022.

[28] J. Leech, G. Whelan, M. Bhaiji, M. Hawes, and K. Scharring, "Connected and autonomous vehicles-the uk economic opportunity," *KPMG. Available online: https://www. smmt. co. uk/wp-content/uploads/sites/2/CRT036586F-Connected-and-Autonomous-Vehicles-% E2*, vol. 80, 2015.

[29] S. International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," *SAE international*, vol. 4970, no. 724, pp. 1–5, 2018.

[30] J. Connor *et al.*, "The role of driver sleepiness in car crashes: a systematic review of epidemiological studies," *Accident; Analysis and Prevention*, vol. 33, no. 1, pp. 31–41, 2001.

[31] K. Hayashi, K. Ishihara, H. Hashimoto, and K. Oguri, "Individualized drowsiness detection during driving by pulse wave analysis with neural network," in *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pp. 901–906, 2005.

[32] T. Ito, S. Mita, K. Kozuka, T. Nakano, and S. Yamamoto, "Driver blink measurement by the motion picture processing and its application to drowsiness detection," in *Proceedings. The IEEE 5th International Conference on Intelligent Transportation Systems*, pp. 168–173, 2002.

[33] L. Fletcher, N. Apostoloff, L. Petersson, and A. Zelinsky, "Vision in and out of vehicles," *IEEE Intelligent Systems*, vol. 18, no. 3, pp. 12–17, 2003.

[34] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 4, pp. 205–218, 2003.

[35] Q. Ji, "Non-invasive techniques for monitoring human fatigue," dtic document, University of Nevada, Reno, NV, USA, 2003.

[36] B. Ranft and C. Stiller, "The role of machine vision for intelligent vehicles," *IEEE Transactions on Intelligent vehicles*, vol. 1, no. 1, pp. 8–19, 2016.

[37] D. Park, D. Ramanan, and C. Fowlkes, "Multiresolution models for object detection," in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pp. 241–254, Springer, 2010.

[38] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[39] A. Sarda, S. Dixit, and A. Bhan, "Object detection for autonomous driving using yolo [you only look once] algorithm," in *2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV)*, pp. 1370–1374, IEEE, 2021.

[40] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," in *Proceedings of the ieee conference on computer vision and pattern recognition*, pp. 5308–5317, 2016.

[41] M. Wollmer, C. Blaschke, T. Schindl, B. Schuller, B. Farber, S. Mayer, and B. Trefflich, "Online driver distraction detection using long short-term memory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 574–582, 2011.

[42] A. Kulshrestha, L. Chang, and A. Stein, "Use of lstm for sinkhole-related anomaly detection and classification of insar deformation time series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4559–4570, 2022.

[43] H. Abbasimehr and R. Paki, "Improving time series forecasting using lstm and attention models," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 1, pp. 673–691, 2022.

[44] A. Koesdwiady, S. M. Bedawi, C. Ou, and F. Karray, "End-to-end deep learning for driver distraction recognition," in *Image Analysis and Recognition: 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings 14*, pp. 11–18, Springer, 2017.

[45] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[46] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of lstm and bilstm in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 3285–3292, 2019.

[47] J. Brownlee, "How to use the timedistributed layer for long short-term memory networks in python." `https://machinelearningmastery.com/timedistributed-layer-for-long-short-term-memory-networks-in-python/`, 2019.

[48] W. H. Organization, "Road safety." Web, 2023. Accessed: 2023-09-02.

[49] N. H. T. S. Administration, "Risky driving." Web, 2023. Accessed: 2023-09-02.

[50] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE transactions on intelligent transportation systems*, vol. 12, no. 2, pp. 596–614, 2010.

[51] F. Tango and M. Botta, "Real-time detection system of driver distraction using machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 894–905, 2013.

[52] J. Darby, M. B. Sánchez, P. B. Butler, and I. D. Loram, "An evaluation of 3d head pose estimation using the microsoft kinect v2," *Gait & posture*, vol. 48, pp. 83–88, 2016.

[53] D. Zhao, Y. Zhong, Z. Fu, J. Hou, M. Zhao, *et al.*, "A review for the driving behavior recognition methods based on vehicle multisensor information," *Journal of Advanced Transportation*, vol. 2022, 2022.

[54] K. Allen, "Tesla model s in autopilot mode in utah crash; driver had hands off wheel." Web, 2018. Accessed: 2023-09-02.

[55] N. T. S. Board, "Collision between a sport utility vehicle operating with partial driving automation and a crash attenuator," investigation report, National Transportation Safety Board, 2018. Accessed: 2023-09-02.

[56] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories," in *CVPR 2011*, pp. 3169–3176, 2011.

[57] C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal multiplier networks for video action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4768–4777, 2017.

[58] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[59] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[60] A. Mujahid, M. Aslam, M. U. G. Khan, A. M. Martinez-Enriquez, and N. U. Haq, "Multi-class confidence detection using deep learning approach," *Applied Sciences*, vol. 13, no. 9, 2023.

[61] R. Bakshi, "Hand hygiene video classification based on deep learning," *Name of the Journal*, vol. Volume number, no. Issue number, p. Page range, 2021.

[62] I. Jegham, I. Alouani, A. B. Khalifa, and M. A. Mahjoub, "Deep learning-based hard spatial attention for driver in-vehicle action monitoring," *Expert Systems with Applications*, vol. 219, p. 119629, 2023.

[63] R. Greer, L. Rakla, A. Gopalan, and M. Trivedi, "(safe) smart hands: Hand activity analysis and distraction alerts using a multi-camera framework," *arXiv preprint arXiv:2301.05838*, 2023.

[64] H. A. Abosaq, M. Ramzan, F. Althobiani, A. Abid, K. M. Aamir, H. Abdushkour, M. Irfan, M. E. Gommosani, S. M. Ghonaim, V. R. Shamji, and S. Rahman, "Unusual driver behavior detection in videos using deep learning models," *Sensors*, vol. 23, no. 1, 2023.

[65] R. Bakshi, "Hand pose classification based on neural networks," *arXiv preprint arXiv:2108.04529*, 2021.

[66] R. Bajpai and D. Joshi, "Movenet: A deep neural network for joint profile prediction across variable walking speeds and slopes," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[67] L. Li, B. Zhong, C. Hutmacher Jr, Y. Liang, W. J. Horrey, and X. Xu, "Detection of driver manual distraction via image-based hand and ear recognition," *Accident Analysis & Prevention*, vol. 137, p. 105432, 2020.

[68] A. Jinda-Apiraksa, W. Pongstiensak, and T. Kondo, "A simple shape-based approach to hand gesture recognition," in *ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 851–855, IEEE, 2010.

[69] A. Jinda-Apiraksa, W. Pongstiensak, and T. Kondo, "Shape-based finger pattern recognition using compactness and radial distance," in *The 3rd International Conference on Embedded Systems and Intelligent Technology (ICESIT 2010), Chiang Mai, Thailand*, pp. –, 2010.

[70] R. Rokade, D. Doye, and M. Kokare, "Hand gesture recognition by thinning method," in *2009 International Conference on Digital Image Processing*, pp. 284–287, IEEE, 2009.

[71] H. Tauseef, M. A. Fahiem, and S. Farhan, "Recognition and translation of hand gestures to urdu alphabets using a geometrical classification," in *2009 Second International Conference in Visualisation*, pp. 213–217, IEEE, 2009.

[72] Y. Liu and P. Zhang, "Vision-based human-computer system using hand gestures," in *2009 International Conference on Computational Intelligence and Security*, vol. 2, pp. 529–532, IEEE, 2009.

[73] N. Yasukochi, A. Mitome, and R. Ishii, "A recognition method of restricted hand shapes in still image and moving image as a man-machine interface," in *2008 Conference on Human System Interactions*, pp. 306–310, IEEE, 2008.

[74] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE transactions on image processing*, vol. 15, no. 7, pp. 1803–1815, 2006.

[75] N. Das, E. Ohn-Bar, and M. M. Trivedi, "On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics," in *2015 IEEE 18th international conference on intelligent transportation systems*, pp. 2953–2958, IEEE, 2015.

[76] R. I. China, "Global and china heavy truck industry report, 2021-2027," January 2022. Report ID: 6228542, Number of Pages: 130, Format: PDF.

[77] N. H. T. S. A. (NHTSA), "Risky driving: Drowsy driving," Access Year. Accessed: 2023-09-02.

[78] M. Zhu, F. Liang, D. Yao, J. Chen, H. Li, L. Han, Y. Liu, and Z. Zhang, "Heavy truck driver's drowsiness detection method using wearable eeg based on convolution neural network," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 195–201, IEEE, 2020.

[79] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[80] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5325–5334, 2015.

[81] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4511–4520, 2015.

[82] D. Bethge, L. F. Coelho, T. Kosch, S. Murugaboopathy, U. v. Zadow, A. Schmidt, and T. Grosse-Puppendahl, "Technical design space analysis for unobtrusive driver emotion assessment using multi-domain context," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–30, 2023.

[83] W. Song, G. Zhang, and Y. Long, "Identification of dangerous driving state based on lightweight deep learning model," *Computers and Electrical Engineering*, vol. 105, p. 108509, 2023.

[84] I. Jahan, K. Uddin, S. A. Murad, M. Miah, T. Z. Khan, M. Masud, S. Aljahdali, and A. K. Bairagi, "4d: a real-time driver drowsiness detector using deep learning," *Electronics*, vol. 12, no. 1, p. 235, 2023.

[85] B. Akrout and S. Fakhfakh, "How to prevent drivers before their sleepiness using deep learning-based approach," *Electronics*, vol. 12, no. 4, p. 965, 2023.

[86] Q. Abbas, M. E. Ibrahim, S. Khan, and A. R. Baig, "Hypo-driver: a multiview driver fatigue and distraction level detection system," *CMC-computers Mater Contin*, vol. 71, no. 1, pp. 1999–2017, 2022.

[87] D. Patil, V. Lokhande, P. Patil, P. Patil, and S. Gaikwad, "Real-time driver behaviour monitoring system invehicles using image processing," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 4, no. 5, pp. 1890–1894, 2022.

[88] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5325–5334, 2015.

[89] B. Esmaeili, A. AkhavanPour, and A. Bosaghzadeh, "An ensemble model for human posture recognition," in *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pp. 1–7, IEEE, 2020.

[90] M. H. Z. M. Fodli, F. H. K. Zaman, N. K. Mun, and L. Mazalan, "Driving behavior recognition using multiple deep learning models," in *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, pp. 138–143, IEEE, 2022.

[91] N. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, 2000.

[92] T. Quettier, F. Gambarota, N. Tsuchiya, and P. Sessa, "Blocking facial mimicry during binocular rivalry modulates visual awareness of faces with a neutral expression," *Scientific Reports*, vol. 11, no. 1, p. 9972, 2021.

[93] L. Karthik, G. Kumar, T. Keswani, A. Bhattacharyya, S. S. Chandar, and K. Bhaskara Rao, "Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound," *PloS one*, vol. 9, no. 3, p. e90972, 2014.

[94] L. Alam and M. M. Hoque, "Real-time distraction detection based on driver's visual features," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6, IEEE, 2019.

[95] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I–I, Ieee, 2001.

[96] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE transactions on intelligent transportation systems*, vol. 8, no. 2, pp. 340–350, 2007.

[97] N. Li and C. Busso, "Analysis of facial features of drivers under cognitive and visual distractions," in *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2013.

[98] L. B. Neto, F. Grijalva, V. R. M. L. Maike, L. C. Martini, D. Florencio, M. C. C. Baranauskas, A. Rocha, and S. Goldenstein, "A kinect-based wearable face recognition system to aid visually impaired users," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 1, pp. 52–64, 2016.

[99] B. Esmaeili, A. Akhavanpour, and A. Bosaghzadeh, "An ensemble model for human posture recognition," *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pp. 1–7, 2020.

[100] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F.-Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5379–5390, 2019.

[101] M.-F. R. Lee, Y.-C. Chen, and C.-Y. Tsai, "Deep learning-based human body posture recognition and tracking for unmanned aerial vehicles," *Processes*, vol. 10, no. 11, 2022.

[102] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE international conference on computer vision*, pp. 2938–2946, 2015.

[103] R. Bajpai and D. Joshi, "Movenet: A deep neural network for joint profile prediction across variable walking speeds and slopes," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[104] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "D3-guard: Acoustic-based drowsy driving detection using smartphones," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 1225–1233, 2019.

[105] W. Yang, C. Tan, Y. Chen, H. Xia, X. Tang, Y. Cao, W. Zhou, L. Lin, and G. Dai, "Birswint: Bilinear full-scale residual swin-transformer for fine-grained driver behavior recognition," *Journal of the Franklin Institute*, vol. 360, no. 2, pp. 1166–1183, 2023.

[106] A. A. Aljohani, "Real-time driver distraction recognition: A hybrid genetic deep network based approach," *Alexandria Engineering Journal*, vol. 66, pp. 377–389, 2023.

[107] C. Fan, S. Huang, S. Lin, D. Xu, Y. Peng, and S. Yi, "Types, risk factors, consequences, and detection methods of train driver fatigue and distraction," *Computational Intelligence and Neuroscience*, vol. 2022, p. 8328077, Mar 2022. PMID: 35371223; PMCID: PMC8970922.

[108] Z. Zheng, S. Dai, Y. Liang, and X. Xie, "Driver fatigue analysis based on upper body posture and dbn-bpnn model," in *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 1, pp. 574–581, 2019.

[109] A. Kondyli, V. P. Sisiopiku, L. Zhao, and A. Barmpoutis, "Computer assisted analysis of drivers' body activity using a range camera," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 3, pp. 18–28, 2015.

[110] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-d posture data," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 586–597, 2014.

[111] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291–7299, 2017.

[112] M. Rezaei and R. Klette, "Look at the driver, look at the road: No distraction! no accident!," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 129–136, 2014.

[113] E. Blythe, L. Garrido, and M. R. Longo, "Emotion is perceived accurately from isolated body parts, especially hands," *Available at SSRN*, 2023.

[114] A. Ezzouhri, Z. Charouh, M. Ghogho, and Z. Guennoun, "Robust deep learning-based driver distraction detection and classification," *IEEE Access*, vol. 9, pp. 168080–168092, 2021.

[115] A. Heitmann, R. Guttkuhn, A. Aguirre, U. Trutschel, and M. Moore-Ede, "Technologies for the monitoring and prevention of driver fatigue," in *Driving Assessment Conference*, 1, pp. 81–86, University of Iowa, 2001.

[116] T. Billah, S. M. Rahman, M. O. Ahmad, and M. Swamy, "Recognizing distractions for assistive driving by tracking body parts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1048–1062, 2018.

[117] S. M. Rahman, T. Howlader, and D. Hatzinakos, "On the selection of 2d krawtchouk moments for face recognition," *Pattern Recognition*, vol. 54, pp. 83–93, 2016.

[118] M. Panwar and P. S. Mehra, "Hand gesture recognition for human computer interaction," in *2011 International Conference on Image Information Processing*, pp. 1–7, IEEE, 2011.

[119] P. Weyers, D. Schiebener, and A. Kummert, "Action and object interaction recognition for driver activity classification," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 4336–4341, 2019.

[120] Y. Xing, C. Lv, Z. Zhang, H. Wang, X. Na, D. Cao, E. Velenis, and F.-Y. Wang, "Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 95–108, 2018.

[121] M. Gjoreski, M. Ž. Gams, M. Luštrek, P. Genc, J.-U. Garbas, and T. Hassan, "Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals," *IEEE access*, vol. 8, pp. 70590–70603, 2020.

[122] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *2014 22nd international conference on pattern recognition*, pp. 660–665, IEEE, 2014.

[123] D. Bethge, C. Patsch, P. Hallgarten, and T. Kosch, "Interpretable time-dependent convolutional emotion recognition with contextual data streams," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2023.

[124] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F.-Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5379–5390, 2019.

[125] D. Tran, H. Manh Do, W. Sheng, H. Bai, and G. Chowdhary, "Real-time detection of distracted driving based on deep learning," *IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1210–1219, 2018.

[126] S. Mühlbacher-Karrer, A. H. Mosa, L.-M. Faller, M. Ali, R. Hamid, H. Zangl, and K. Kyamakya, "A driver state detection system—combining a capacitive hand detection sensor with physiological sensors," *IEEE transactions on instrumentation and measurement*, vol. 66, no. 4, pp. 624–636, 2017.

[127] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "3d convolutional neural networks for efficient and robust hand pose estimation from single depth images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1991–2000, 2017.

[128] A. Dutta and A. Zisserman, "The via annotation software for images, audio and video," in *Proceedings of the 27th ACM international conference on multimedia*, pp. 2276–2279, 2019.

[129] B. Benfold and I. Reid, "Guiding visual surveillance by tracking human attention.," in *BMVC*, vol. 2, p. 7, 2009.

[130] N. K. Vaegae, K. K. Pulluri, K. Bagadi, and O. O. Oyerinde, "Design of an efficient distracted driver detection system: Deep learning approaches," *IEEE Access*, vol. 10, pp. 116087–116097, 2022.

[131] S. Sharma and V. Kumar, "Distracted driver detection using learning representations," *Multimedia Tools and Applications*, pp. 1–18, 2023.

[132] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.