Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 3, March 2024, pp. 1974~1988
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i3.pp1974-1988

Improving night driving behavior recognition with ResNet50

Muhammad Firdaus Ishak[1], Fadhlan Hafizhelmi Kamaru Zaman[1], Ng Kok Mun[1], Syahrul Afzal Che Abdullah[1], Ahmad Khushairy Makhtar[2]

[1]Vehicle Intelligence and Telematics Lab, School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia.
[2]School of Mechanical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia

Article Info

Article history:

Received Oct 26, 2023
Revised Dec 5, 2023
Accepted Jan 11, 2024

Keywords:

CLAHE
Deep learning
Driving behavior
Infrared images
Low illumination images
Road safety

ABSTRACT

The issue of driving behavior at night poses significant challenges due to reduced visibility and increased risk of accidents. Recent works have leveraged deep learning techniques to enhance night-time driving safety. However, the limited availability of high-quality training data and the lack of robustness in existing models present significant problems. In this work, we propose a novel approach to improve driving behavior recognition at night using ResNet50 with contrast limited adapted histogram equalization (CLAHE). We collected a new dataset and developed a more effective and robust model that can accurately recognize driving behaviors under low-illumination conditions, thereby reducing the likelihood of collisions and improving overall road safety. The experimental results demonstrate significant improvements in the deep learning model's performance compared to conventional methods. Notably, the ResNet50 model delivers the best performance with accuracy rates of 90.73% using NIGHT-VIS-CLAHE data, demonstrating a 16% improvement in accuracy. For benchmark purposes, the InceptionV3, GoogleNet, and MobileNetV2 models also show enhanced accuracy through CLAHE implementation. Furthermore, NIGHT-VIS-CLAHE implementation in ResNet50 achieved 90.29% accuracy, surpassing the best NIGHT-IR InceptionV3 at 89.27%, highlighting the advantage of ResNet50 with CLAHE in low-light conditions even against infra-red sensor.

Corresponding Author:

Fadhlan Hafizhelmi Kamaru Zaman
Vehicle Intelligence and Telematics Lab, School of Electrical Engineering, College of Engineering
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
Email: fadhlan@uitm.edu.my

1. INTRODUCTION

According to WHO's 2018 global report, road traffic crashes cause about 1.35 million deaths annually, with 54% involving vulnerable road users. Although developing nations possess 60% of the world's vehicles, over 90% of road fatalities occur in low and middle-income nations. Road accident is also a significant cause of death for 5 to 29-year-olds [1]. In Malaysia, there are growing numbers of cars on the roads because the population is rising where in 2020 there were 32.3 million cars, up by 1.2 million from the previous year. This is because more people are buying cars and learning to drive [2]. Similarly, in the US the statistic shows more accidents were caused by distracted driving. In 2021, 3,522 people were killed and an estimated 362,415 people were injured in motor vehicle traffic accidents caused by distracted drivers [3]. The most common causes of recorded car accidents were asleep, texting, or talking on the phone while driving.

Drowsy driving is a key accident contributor, and its impact is aggravated by severe consequences. Long journeys or night driving may cause drivers to be tired and sleepy. The urgency of the situation drives the demand for sleepiness detection and warning technology, including acquisition, processing, and alarms. Drowsiness is detected using three basic approaches: vehicle, behavior, and physiology [4]. Face detection and recognition can assist in identifying driving drowsiness behavior.

Additionally, face detection and recognition are critical in both research and practical applications of computer vision. Face recognition, which has been extensively researched and widely applied, provides exceptional versatility in this sector. However, recognizing faces within entire photos necessitates a significant amount of time. To solve this issue, many strategies for extracting facial regions during pre-processing have been investigated to reduce processing time. Haar classifiers have been heavily incorporated in the hardware implementation of face detection algorithms, effectively enhancing face classification efficiency by alleviating computing constraints. This complex interaction between face detection, leveraging haar cascade for computational efficiency, and the broader landscape of face recognition applications emphasizes the continued progress of computer vision technologies, which are poised to improve accuracy and speed in real-world scenarios [5]. Eye detection also plays an important role in driver sleepiness detection [6]. It is critical in effectively detecting and assessing distinct eye characteristics, which are important indicators of attention. While convolutional neural networks (CNNs) excel at object detection and recognition [7], [8], their computing requirements make real-time, cost-effective processing on devices such as central processing units (CPUs) difficult. Furthermore, eyes are unique among facial features, necessitating a particular treatment. As a result, Haar Cascade with three levels of eye detection has been adopted, maximizing both accuracy and efficiency. This synergy between the need for practical, low-cost, real-time eye detection and the uniqueness of eye features highlights the need for novel solutions in driver sleepiness detection, with eyes serving as the main markers of driver attentiveness and road safety [9].

A CNN represents a type of advanced learning system within the field of deep learning. Unlike traditional methods, a CNN can automatically grasp important aspects from input data without requiring human intervention to identify specific features. This innovative approach is especially effective when working with images. Think of it as an upgraded version of conventional networks. The CNN consists of layers that autonomously learn significant patterns from images. This ability makes it exceptionally useful for tasks such as recognizing objects in images, analyzing medical scans, dividing images into sections, and even understanding language. Additionally, CNNs incorporate techniques to avoid becoming overly specialized, preventing them from losing sight of the bigger context. Various versions of CNNs, such as visual geometry. group (VGG), AlexNet, Xception, Inception, and ResNet, are tailored to different applications based on their distinct learning abilities [10].

Chen *et al.* [11], an innovative hybrid framework is introduced for extracting distinctive features essential for recognizing driver distraction. The process commences with the resizing of training images to a standardized dimension, which then serves as input for three prominent models: InceptionV3, Xception, and MobileNet. Following this, a fine-tuning step is applied to these models, involving the selective freezing of specific network layers. The analysis of heat maps and feature visualizations for each individual model reveals the presence of unique feature regions associated with different network architectures. Building on these insights, a fresh deep feature fusion approach is devised to combine the features extracted from these three models. Experimental assessments conducted on the StateFarm database underscore that the proposed method delivers competitive recognition performance when benchmarked against other established methodologies in the domain.

Bahari and Mazalan [4] aims to address the problem of recognizing distracted drivers who are not paying attention to the road by employing a deep learning-based categorization approach. The proposed method entails creating and running the ResNet 50 neural network in Jupyter Notebook and Python. the state farm dataset, which includes ten separate driving habits, is a major resource for the study. To evaluate the model's performance thoroughly, several assessment measures such as confusion matrices, accuracy, precision, recall, and F1-score are used. The results are striking, with the algorithm obtaining an astounding accuracy rate of 94% in spotting distracted drivers looking elsewhere. Importantly, the technology can recognize photos of inattentive drivers and send messages when such occurrences are spotted in video footage.

While dealing with low-illumination images such as images taken at night, several researchers have shown good results with the contrast limited adapted histogram equalization (CLAHE). Chen *et al.* [11], focuses on road sign detection in various illumination circumstances, particularly at night, and is motivated by the success of CLAHE. The research entails significant data collection, including road driving across many Taiwanese cities to create a unique dataset of traffic signs in both day and evening circumstances. Using the YOLO model, the study compares image enhancement approaches such as contrast stretching (CS), histogram equalization (HE), and CLAHE, with CLAHE outperforming the others when integrated into the YOLOv5x model for nighttime traffic sign identification. This study recommends using CLAHE's YOLOv5x model to improve road sign identification in low-light circumstances, leading to enhanced road safety and navigation. Yuan *et al.* [12], focuses on the critical issue of improving the lighting quality in unmanned aerial vehicle (UAV) images, particularly within the realm of power system operation and maintenance. Low-light conditions often result in UAV images with reduced detail, leading to diminished accuracy in detecting objects in overhead power transmission systems. Yet, a notable research gap remains in enhancing low-light images, specifically for UAV-based power system applications. This study introduces an innovative method for enhancing low-light images by utilizing the CLAHE technique. The approach involves separating the luminance and chrominance channels through color space conversion, allowing CLAHE to enhance luminance contrast while preserving chrominance. Furthermore, an optimized contrast limit parameter within CLAHE is adjusted to strike a balance between enhancing contrast and suppressing noise, resulting in a significant improvement in insulator detection accuracy from 58.0% to 82.0%, as confirmed by experimental findings.

CLAHE algorithm improves image contrast by implementing a clip limit and divides the image into non-overlapping contextual sections known as tiles or blocks. These regions are handled independently and driven by two key factors: block size and clip limit. Adjusting the block size broadens the dynamic range, increasing image contrast. Modifying the clip limit, on the other hand, has an impact on image brightness and histogram distribution. CLAHE improves image quality through careful parameter management by increasing contrast and entropy distribution [13].

In this work, we propose to combine ResNet50 with CLAHE to enhance the recognition of driving behavior at night. To do so, we have collected our own driving behavior dataset, which we used to train and validate our models. We also conducted experiments using infrared (IR) cameras to assess performance comparisons between them. In total, we assembled a dataset of 44 subjects to build our proposed model which is greater in numbers than work proposed in [14] where they used 26 subjects to prepare their proposed model. The proposed strategy offers the potential to improve overall driving behavior and reduce collision risks by addressing the difficulties of recognizing behaviors under lower visibility while driving at night.

This paper is structured as follows: Section 2 provides an overview of related works in the study. In section 3 outlines the methodology used to develop the model. In section 4 presents the results and offers a comprehensive discussion. Finally, in section 5 presents our conclusions.

2. PREVIOUS WORKS

Previous research tried to find ways to recognise sleepy driving, but it was impossible to create a model that would function in every circumstance. This issue is resolved by the model proposed in [15], which predicts the outcomes without taking the driver's identity into account and improves the accuracy and dependability of the outcomes. When observing a person's physical changes or potential for distraction, an ECG is a useful and reasonably priced tool. ECG was used in a study involving 8 participants to determine whether or not the driver was distracted by talking to passengers or answering phone calls. The findings demonstrated how well wavelet packet transform and linear discriminant analysis performed together to determine whether or not the driver was distracted [16]. Researchers have previously investigated the impact of conversational distraction on driving performance. The effects of distractions on the driver's brain were observed using EEG data, and the results were classified using a machine learning technique. The EEG data increased the accuracy of the results by 5% and offered helpful information concerning distracted driving. Understanding distracted driving behaviour is greatly influenced by cognitive considerations [17].

The ten most prevalent forms of distracted driving are highlighted using a method known as class activation mapping [18]. The authors employed a pre-trained model that makes use of ResNet50, Inception V3, and Xception to identify driving behaviour. This enhances the driving activity detecting system's accuracy. The hand gestures and patterns associated with driving without distraction can be recognised by the model. A dataset was utilised in [14] to investigate driving when preoccupied. Distractions were classified into eight categories, such as talking on the phone, texting, and looking behind captured from 26 people. Better results were obtained by analysing images and identifying the most distracting activities using a model named EfficientDet. This model was compared with other models, such as faster R-CNN and Yolo-V3 [14]. Meanwhile, earlier research in [19] uses a combination of conventional image analysis and heart rate variability (HRV) to identify driver fatigue and drowsiness. In the initial step, features from InceptionV3 are used to train a long short-term memory network (LSTM) to recognise sequences in video data. After combining the characteristics, the blood volume pulse vector (PBV) technique makes the conclusion after the LSTM removes static distortions to offer precise recognition. The method's efficacy in identifying driver fatigue is assessed [19].

A deep-learning method was proposed by Tran *et al.* [20] to identify various distracted driving behaviours. A two-camera synchronised image recognition system was created. The driver's facial and body movements are captured by the cameras. After that, two CNNs are fed the collected images concurrently to enhance classification performance. In order to create a driving experience that is almost realistic, the suggested distraction detection technique was tested in a lab-based assisted driving environment in this work [21]. Images of safe and distracted driving are included in the dataset utilised for the study. A voice-alert system was also created to warn inattentive drivers to pay attention to the road. For this method, several networks including

VGG-16, ResNet, and MobileNet-v2 were assessed. The two-camera system with VGG-16 networks demonstrated a 96.7% recognition accuracy at a calculation speed of 8 frames per second, according to the results.

The goal of the study Wei *et al.* [22] is to monitor driving behaviour and pinpoint any potential distraction factors. Ten categories are used by the model to categorise distractions, including using the dashboard, reaching behind, chatting on the phone with either hand, and communicating with other passengers. Various CNN architectures, such as Inception-v3 and AlexNet, have demonstrated encouraging performance. A collection of dashboard photos that are used to monitor the driver for distractions make up the model's input. Furthermore, Tran *et al.* [20] looks into the detection of driver distractions using ResNet50 neural networks. The ResNet50 network's performance is examined, and its applicability in distraction detection is assessed. According to the study, the network performs better at categorising drivers who are distracted than it does at determining the exact kind of distraction.

Rahman *et al.* [23] focuses on face recognition, a vital feature of biometric identification systems in which image quality is critical, particularly in challenging low-light circumstances. This paper introduces a specialized face recognition system designed to address the specific challenges posed by low-light conditions, based on prior research utilizing CLAHE techniques for enhancing image quality in low-light scenarios. The suggested method employs CLAHE to improve image contrast, resulting in higher image quality. Notably, experimental results show that this strategy is effective, with an amazing accuracy rate of 76.92% achieved even in extremely low-light settings with a brightness level as low as -80. Lashkon *et al.* [24], presents an innovative approach for enhancing nighttime image contrast to improve the detection of vehicles in low-light scenarios. The technique combines CLAHE with image dehazing to augment contrast in images without over-amplifying it. The practical application involves employing a camera-based internet of things (IoT) edge computing system for traffic and road surveillance, where the proposed method outperforms existing CLAHE-based approaches across a range of image enhancement quality criteria. To address the challenges associated with detecting vehicles in low-light conditions, a deep neural network based on YOLOv5 is designed and trained using a custom dataset.

Low-illumination images frequently offer difficulties when performing complicated tasks such as object detection and semantic segmentation. This complexity presents challenges for autonomous vehicles operating at night. As a result, the quality of these low-light photographs must be improved. Strategies for improving low-light images, like many other techniques in computer vision, can be divided into two types: traditional methods and ones that rely on deep learning. There are approaches that rely on histogram equalization within the scope of conventional procedures [25].

3. METHOD

The following section discusses data collection, data processing, and data augmentation used in this work. Subsequently, the implementation of CLAHE and ResNet50 are also elaborated in detail. Then, we explain the experimental setup used in this work. Several sample images collected for this work are also illustrated in this section.

3.1. Data collection

The data collection was conducted at the vehicle intelligence and telematics lab at UiTM Shah Alam. It involved 44 participants, where in total, 1,104,714 images were collected. Each participant is allocated approximately one hour, including a 15-minute briefing on the 12 tasks related to 6 behaviors namely normal behavior, and distracting behavior including yawning, nodding, talking, texting, and calling. Data were collected during a single nighttime session to capture these 6 driver behaviors. The task involved recording 12 variations of 6 different behaviors, each lasting for 1 minute, with a preparation time of 25-30 seconds and a 5-second countdown before recording commenced. The total duration of the recorded videos was approximately 19 minutes, and each participant completed a single low-illumination session referred to as the NIGHT-VIS session. We also recorded similar data for day driving under normal lighting conditions referred to as DAY-VIS. Additionally, we used an IR camera to record driving behavior in low illumination referred to as NIGHT-IR. Following the data collection process, the videos underwent processing to ensure synchronization and were then transformed into images. To model the 6 behaviors, we used ResNet50 and compared them against several CNN models such as InceptionV3, GoogleNet, and MobileNetV2. Figure 1 illustrates the hardware utilized for collecting model data, which comprises a 55-inch TV mounted on a racing simulator rig stand, driving equipment, and 2 cameras (1 webcam and 1 infrared camera).

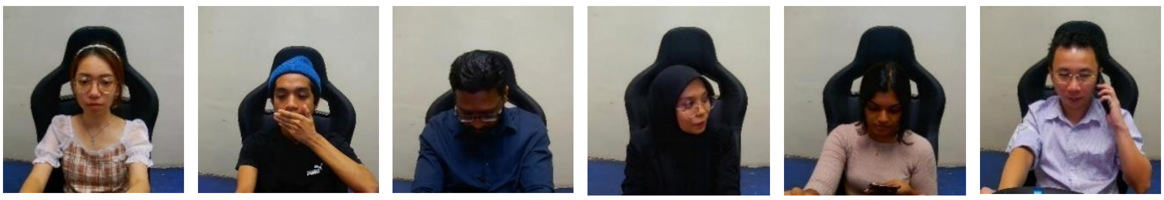
The following are the standard protocols for the driving behavior simulation: during the usual driving session, participants were asked to position their left hand on the steering wheel at 11 o'clock and their right hand at 2 o'clock.

Figure 1. Driving simulator

Yawning was defined by either covering the mouth with the left or right hand or not covering it at all. There were four types of nodding-off behavior: nodding to the left, centre, right, or behind. Normal driving, yawning, and nodding-off were also conducted when subjects wore sunglasses. Texting on a phone was carried out with the phone at the same level as the driving wheel and near the chest. Participants were also asked to use their left hand for the left ear and their right hand for the right ear when making a phone conversation. Furthermore, conversing with the left or right passenger required participants to look to the left or right while they were talking to accurately mimic the anticipated actions. The total duration of the videos including the preparation time is 20 minutes and 45 seconds. Finally, the recorded videos are processed, synced, labelled, and extracted into still images, and they are organized based on the participant's identity and behaviors.

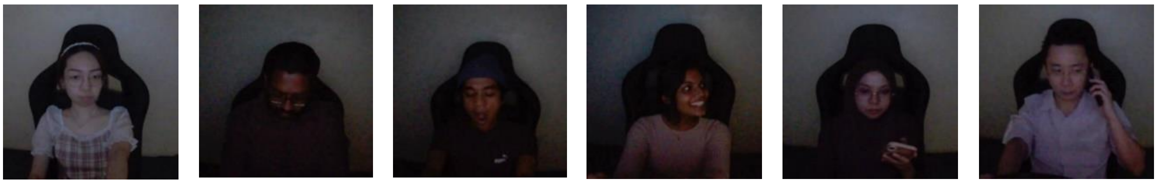
In summary, we implemented different modes of capturing and processing visual information in the context of driving scenarios. The resulting dataset consists of four variants of data namely DAY-VIS, NIGHT-VIS, NIGHT-IR, and NIGHT-VIS-CLAHE. DAY-VIS represents daytime driving conditions, where a standard visual camera is used to capture images and videos of the road and surroundings. NIGHT-VIS, on the other hand, pertains to nighttime driving using a visual camera, which can be challenging due to low light conditions. NIGHT-IR involves nighttime driving but with the use of an IR camera, which is sensitive to heat radiation and can capture images even in total darkness. Our proposed data variant, NIGHT-VIS-CLAHE denotes nighttime driving using a visual camera but with an added image enhancement technique called CLAHE to improve visibility and contrast in low-light situations. These terms help categorize and understand the specific visual conditions under which driving scenarios are observed and monitored, each with its own set of challenges and technological solutions.

Figures 2-4 shows the images from the driving behaviors captured in DAY-VIS conditions, which include normal driving as well as behaviors like yawning, nodding, talking, texting, and calling. The variants of data collected in this work are shown in Figures 2(a) to 2(f), Figures 3(a) to 3(f), and Figures 4(a) to 4(f) respectively. Similarly, Figure 3 illustrates these behaviors but under NIGHT-VIS conditions, providing insights into how they manifest in low-illumination settings. Figure 4, on the other hand, showcases the same behaviors but in NIGHT-IR conditions, where infrared imaging is used. The comparison of these behaviors across different lighting conditions and imaging techniques suggests that night driving poses more challenges, as indicated by the need for specialized techniques to enhance visibility before the behavior can be accurately recognized. Additionally, this highlights the prevalence of risky behaviors like yawning, nodding, and phone usage during both daytime and night-time driving, underscoring the importance of understanding these behaviors for road safety initiatives.



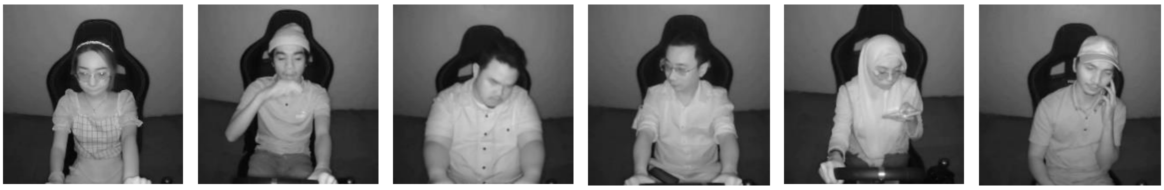
(a) (b) (c) (d) (e) (f)

Figure 2. Example of driving behaviors for DAY-VIS images which includes: (a) normal, (b) yawning, (c) nodding, (d) talking, (e) texting, and (f) calling



(a) (b) (c) (d) (e) (f)

Figure 3. Example of driving behaviors for NIGHT-VIS images which includes: (a) normal, (b) yawning, (c) nodding, (d) talking, (e) texting, and (f) calling



(a) (b) (c) (d) (e) (f)

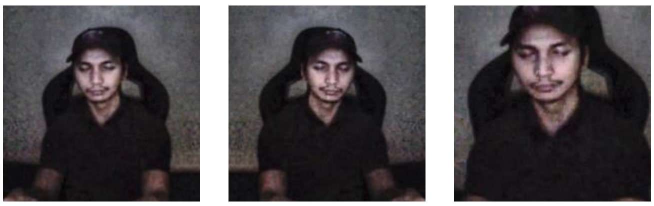
Figure 4. Example of driving behaviors for NIGHT-IR images which include: (a) normal, (b) yawning, (c) nodding, (d) talking, (e) texting, and (f) calling

3.2. Data augmentation

We also employ data augmentation to ensure the training data are balanced across all classes. Data augmentation is a vital technique used to enhance the performance of machine learning models by artificially expanding the size and diversity of the training dataset [26]. Here data augmentation is used to stabilize accuracy and mitigate bias caused by imbalanced data. This process involves generating additional training samples from the existing ones, introducing variations such as flipping, zooming, and other transformations to diversify the dataset. For instance, yawning behavior was augmented to a total of 10,000 samples as shown in Figure 5, Figures 5(a) to 5(c), nodding behavior is augmented to 11,000 samples as shown in Figure 6, Figures 6(a) to 6(c), and talking behavior is augmented to 10,000 samples in Figure 7, Figures 7(a) to 7(c).

(a) (b) (c)

Figure 5. Example of image augmentation for yawning behavior which includes: (a) original image, (b) flipped image, and (c) zoomed image



(a) (b) (c)

Figure 6. Example of image augmentation for nodding behavior which includes: (a) original image, (b) flipped image, and (c) zoomed image

(a) (b) (c)

Figure 7. Example of image augmentation for talking behavior which includes: (a) original image, (b) flipped image, and (c) zoomed image

During training, an additional stage of imageaugmentation is also applied to reduce overfitting, a common concern in machine learning. This step ensured that the model encounters a broader range of data variations during training. By doing so, the image augmentation helps guard against overfitting, ensuring that the model learns meaningful patterns in the data. Table 1 provides a comparison of the distribution of training data by behavior class before and after data augmentation in a classification task related to behavior detection.

Table 1 is divided into two sections, before data augmentation and after data augmentation. Each section shows the number of data samples for different behaviors, categorized as normal, yawning, nodding, talking, texting, and calling, along with the respective ratios within each class. Before data augmentation, there are imbalanced class ratios, with varying numbers of samples for each behavior. For example, the yawning class goes from being underrepresented before data augmentation to having a more balanced ratio after augmentation, as indicated in Table 1.

Table 1. The ratio of training data by behavior class before and after data balancing

Behavior	Before data balancing			After data balancing		
	DAY-VIS (Ratio)	NIGHT-VIS, NIGHT-VIS-CLAHE (Ratio)	NIGHT-IR (Ratio)	DAY-VIS (Ratio)	NIGHT-VIS, NIGHT-VIS-CLAHE (Ratio)	NIGHT-IR (Ratio)
Normal	12,279 (0.248)	11,149 (0.229)	10,492 (0.266)	12,279 (0.156)	11,149 (0.138)	10,492 (0.149)
Yawning	2,496 (0.049)	2,637 (0.542)	1,877 (0.048)	10,496 (0.134)	11,637 (0.144)	11,877 (0.168)
Nodding	919 (0.017)	1,042 (0.021)	901 (0.023)	9,919 (0.075)	11,042 (0.137)	11,901 (0.169)
Talking	2,350 (0.047)	2,407 (0.049)	2,261 (0.057)	10,350 (0.132)	11,407 (0.141)	12,261 (0.174)
Texting	12,214 (0.246)	12,28 (0.252)	12,276 (0.311)	12,214 (0.155)	12,287 (0.152)	12,276 (0.174)
Calling	11,891 (0.239)	11,846 (0.243)	11,703 (0.296)	11,891 (0.151)	11,846 (0.146)	11,703 (0.166)

3.3. CLAHE

The CLAHE technique was employed for image preprocessing and enhancement to improve the low-illuminated images collected. CLAHE is a well-established method in computer vision and image processing that enhances image contrast while mitigating issues such as noise amplification. A CLAHE object was configured using the OpenCV library with specific parameters. The 'clipLimit' parameter was set to 3.0 to control the extent of contrast enhancement, with higher values resulting in more enhancement but requiring careful adjustment to prevent over-amplification. Additionally, a 'tileGridSize' of (8,8) was defined, dividing the image into 8×8 blocks for localized adaptive contrast adjustment. This division enabled CLAHE to adaptively equalize the histogram within each block, effectively enhancing visibility in regions with diverse lighting conditions. These parameter settings in CLAHE were crucial for improving image quality, facilitating robust feature extraction, and supporting subsequent analysis in the research. It is adapted using (1) [13].

$$f(x) = c_1 e^{-\lambda_1 x} + c_2 e^{-\lambda_2 x} \quad (1)$$

3.4. Proposed ResNet50 model

The ResNet50 architecture [27] as shown in Table 2 is a deep CNN architecture renowned for its exceptional performance in image classification tasks. It comprises of 50 layers, making it considerably deep, and is characterized by a unique residual learning approach. This architecture introduces skip connections or shortcuts that enable the network to skip over a certain number of layers during training. These skip connections facilitate the flow of gradient information, mitigating the vanishing gradient problem associated with training very deep networks. ResNet50 consists of five convolutional stages, each containing a varying number of residual blocks. These blocks contain multiple convolutional layers with batch normalization and rectified linear unit (ReLU) activations. Additionally, ResNet50 incorporates global average pooling and a fully connected layer at the end to produce classification predictions. Its architectural design, with skip connections and residual blocks, allows it to achieve remarkable accuracy in image recognition tasks while effectively addressing the challenges of deep network training. ResNet50's design enables it to learn meaningful features from input data and achieve cutting-edge accuracy on numerous benchmark datasets [28].

For benchmarking purposes, several pre-trained CNN models are used to benchmark the performance of our proposed model. The benchmarked models include MobileNetV2, InceptionV3, ResNet50, and GoogleNet. The pre-trained models are fine-tuned to learn new classes of driving behaviors from our dataset. Fine-tuning is the process of adapting a previously trained machine-learning model to a new task or dataset. The pre-trained models were trained on a large, broad dataset, but fine-tuning allows it to learn specific features from a smaller, specialized dataset. This is accomplished by allowing some layers of the model to be retrained using the fresh data. The purpose of fine-tuning is to make the model better at the new task by applying what it learned from the prior assignment. MobileNetV2 is a CNN architecture designed specifically for mobile devices. It learns features from images using a variety of approaches such as residual connections, lightweight depth-wise convolutions, and batch normalization [29]. The InceptionV3 model has 42 layers, including numerous inception modules. To extract features at various scales, these modules employ a combination of filters of diverse sizes and pooling layers. The features are then concatenated and input into a succession of fully connected layers to create the final prediction [30]. GoogleNet, also known as InceptionV1, was introduced in 2014 as a collaborative effort between Google and various universities. It achieved remarkable success by winning the ILSVRC 2014 image classification challenge, outperforming previous champions like AlexNet and ZF-Net. GoogleNet's key innovation lies in its efficient architecture, featuring 22 layers designed for optimal computational performance even on devices with limited resources. This architecture incorporates techniques like 1×1 convolutions and global average pooling. Notably, it includes two auxiliary classifier layers connected to specific Inception layers, enhancing its classification capabilities. These auxiliary classifiers employ techniques like average pooling, dimension reduction via 1×1 convolutions, ReLU activation, dropout regularization, and a Softmax classifier with 1,000 output classes, similar to the main Softmax classifier.

Table 2. ResNet50 architecture [27]

Layer name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
Conv1	112×112	7×7, 64, stride 2				
Conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix}3\times3, & 64\\3\times3, & 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 64\\3\times3, & 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$
Conv3_x	28×112	$\begin{bmatrix}3\times3, & 128\\3\times3, & 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 128\\3\times3, & 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times8$
Conv4_x	112×112	$\begin{bmatrix}3\times3, & 256\\3\times3, & 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 256\\3\times3, & 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times36$
Conv5_x	112×112	$\begin{bmatrix}3\times3, & 512\\3\times3, & 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 512\\3\times3, & 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$
	1×1	Average pool, 1,000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

3.4. Experimental setup

The datasets were divided based on driver identity, utilizing 35 drivers for the training set and 9 drivers for the testing set. The selection of test datasets considered factors such as ethnicity, skin color, hijab-wearing, and participant age. To train the model, transfer learning was employed by replacing the existing CNN model with new trainable layers. This adaptation was necessary because the original network had been trained on the ImageNet dataset, which had 1,000 output categories. Table 3 provides details on the training hyperparameters. Following training, the CNN models were saved with the updated layers, facilitating validation testing on the test dataset. Additionally, each model's performance was assessed by its behavior detection capability and the confusion matrix of behaviors, offering insights into behavior distribution, and aiding in the selection of the CNN model with the highest accuracy and minimal prediction errors for each behavior.

Figure 8 summarizes this work which aims to enhance the detection and understanding of driving behavior, particularly in low-light conditions, with a specific emphasis on night-time situations. The methodology utilized involves applying CLAHE to improve the quality of visual data captured during dark conditions. Then, the performance gained by the enhanced data is compared with data from infra-red cameras (NIGHT-IR), which are often used for better visuals at night. By applying CLAHE to the visual data, this work contributes to the enhancement of the visual data and creates the NIGHT-VIS-CLAHE dataset. In order to detect a range of driving behaviors, including but not limited to normal driving, yawning, nodding off, talking, messaging, and calling, this work suggests the implementation of ResNet 50 models at night. The comparative analysis of these datasets and the performance evaluation of the ResNet 50 models are presumably depicted in Figure 8. This provides valuable insights into the methodology's effectiveness in enhancing the precision of driving behavior identification in low-light conditions.

Table 3. Training hyperparameters

Batch size	Learning rate	Epoch	Pixel range augmentation	Scale range augmentation
100	0.0003	10	-30 to 30	-1.1

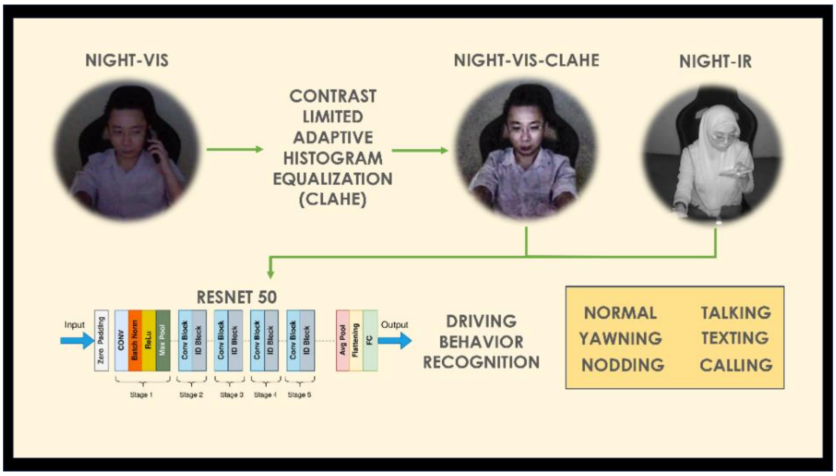


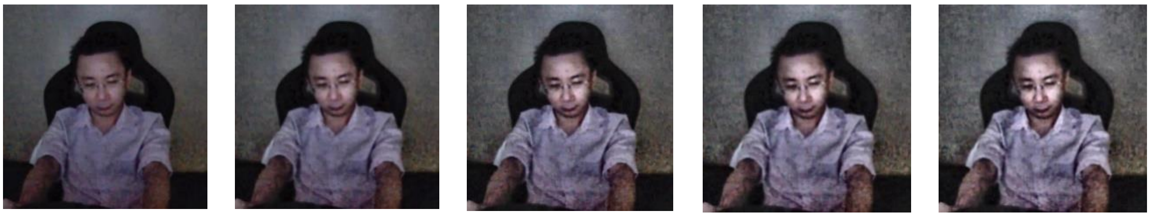
Figure 8. Comprehensive NIGHT-VIS, NIGHT-VIS-CLAHE, and NIGHT-IR analysis: using ResNet 50 for nighttime driving behavior recognition

4. RESULT AND DISCUSSIONS

The following section outlines the results of CLAHE implementation on the collected dataset and the recognition performance of ResNet50 and the benchmarked deep networks. In this section we show that ResNet50 with CLAHE delivered the best performance for night driving behavior recognition, surpassing other benchmarked networks. ResNet50 with CLAHE also gave better performance than the behvaior recognition from infra-red images.

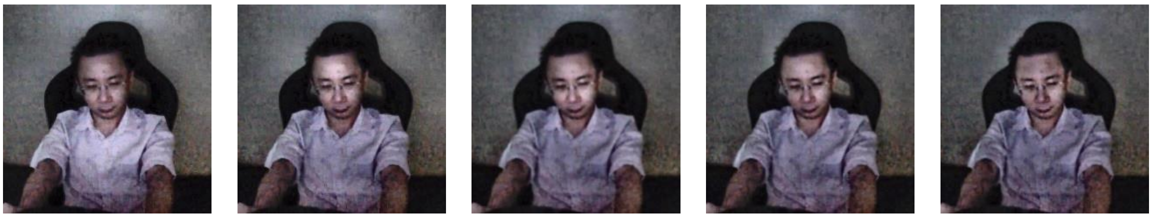
4.1. CLAHE implementation

CLAHE is a technique used to improve the contrast and visibility of details within an image while preventing over-amplification of noise. This method divides the image into small, non-overlapping tiles and equalizes the histogram of each tile separately. The key parameters in CLAHE are the clip limit and the tile grid size. In the experiment, as depicted in Figure 9, Figures 9(a) to 9(e), different clip limits (1, 2, 3, 4, and 5) were tested to determine the most suitable one. Ultimately, a clip limit of 3 was chosen, which means that the histogram equalization process is limited to prevent excessive amplification of local contrast. In Figure 10, Figures 10(a) to 10(e), a similar experiment was conducted to select the optimal tile grid size (3x3, 5x5, 8x8, 10x10, and 15x15), and a grid size of 8x8 was found to be the most effective. These parameter selections are crucial in achieving the desired balance between enhancing image contrast and avoiding unwanted artifacts in the final enhanced image. Figure 11, Figures 11(a) to 11(f) finally presents these driving behaviors in NIGHT-VIS CLAHE conditions using the proposed parameters, which involve contrast-enhanced low-illumination images.



(a) (b) (c) (d) (e)

Figure 9. The effect of different CLAHE clip limits on the images with the values of clip limit is set to: (a) clip limit = 1, (b) clip limit = 2, (c) clip limit = 3, (d) clip limit = 4, and (e) clip limit = 5



(a) (b) (c) (d) (e)

Figure 10. The effect of different CLAHE tile grid sizes on the images with the values of tile grid size is set to: (a) grid size = 3x3, (b) grid size = 5x5, (c) grid size = 8x8, (d) grid size = 10x10, and (e) grid size = 15x15

(a) (b) (c) (d) (e) (f)

Figure 11. Example of driving behaviors for NIGHT-VIS CLAHE images which includes several examples showing: (a) normal, (b) yawning, (c) nodding, (d) talking, (e) texting, and (f) calling behaviors.

4.2. CNN model recognition performance

The experiments were conducted on a computer equipped with an AMD EPYC 24 Core CPU and an NVIDIA RTX A5000 GPU, with TensorFlow as the core framework. During training, various key parameters were set to optimize the model's performance. The batch size was configured at 64, and the Adam optimizer utilized a learning rate of 0.001. Input images were then resized to 100x100 pixels. To enhance diversity and prevent overfitting, data augmentation techniques were applied to the training dataset. These techniques included rotation (up to 5 degrees), zooming (up to 10%), horizontal shifting (up to 20% of the image width), vertical shifting (up to 5% of the image height), shearing (up to 10 degrees), and horizontal flipping. These augmentations improved the model's robustness and generalization capabilities by exposing it to a wider range of variations in the training data. This comprehensive methodology ensured the effective training of the ResNet50 model for behavior detection.

Table 4 provides a comprehensive overview of the performance of ResNet50 and several benchmarked CNN models on different data variations. In this context, the models include InceptionV3, MobileNetV2, and GoogleNet, while the data types consist of DAY-VIS, NIGHT-VIS, NIGHT-IR, and NIGHT-VIS-CLAHE. The percentages represent the accuracy of these models in classifying or processing the respective data types. According to Table 4, InceptionV3 performs well on DAY-VIS and NIGHT-IR variant, with accuracies of 87.99% and 89.27%, respectively. Meanwhile, ResNet50 excels in NIGHT-VIS-CLAHE variant with an accuracy of 90.29%. Compared to NIGHT-IR images, the results suggest that generally the IR images give better accuracy, which will be discussed in detail in the next section. Besides, NIGHT-VIS-CLAHE gives higher accuracy, suggesting its suitability for the given task. Consequently, it is recommended to use the ResNet50 model with NIGHT-VIS-CLAHE data for the most reliable results in the analyzed context.

4.3. Behavior recognition performance of CLAHE implementation for low-illumination images

Figure 12 shows the performance of ResNet50, benchmarked against three models, InceptionV3, MobileNetV2, and ResNet50, in two separate sessions, NIGHT-VIS and NIGHT-VIS-CLAHE. In the NIGHT-VIS session, InceptionV3 scored highest with 85.47%, followed by ResNet50 with 74.68%, and GoogleNet with 71.4%. MobileNetV2 clearly struggled with low-illumination images and scored worst with 35.06%. According to Figure 12, all models improved in accuracy under NIGHT-VIS-CLAHE session. The highest improvement was observed in MobileNetV2 with 30% increase, and 16% accuracy improvement was observed in ResNet50, with 16% increase. Both GoogleNet and InceptionV3 were just slightly improved. Overall, the best performance was obtained by ResNet50 with 90.3% recognition accuracy on NIGHT-VIS-CLAHE. Even when compared to DAY-VIS, ResNet50 NIGHT-VIS-CLAHE still outperformed the best DAY-VIS performance by InceptionV3 which is at 87.99%. These findings illustrate the positive influence of CLAHE on model correctness, which is especially obvious for MobileNetV2 and ResNet50.

Table 4. Behavior recognition performance for ResNet50 compared against other CNN models using different image data

Image session	Accuracy (%)			
	MobileNetV2	GoogleNet	InceptionV3	ResNet50
DAY-VIS	67.57	60.92	87.99	86.01
NIGHT-VIS	35.06	71.38	85.47	74.68
NIGHT-IR	85.97	88.86	89.27	80.88
NIGHT-VIS-CLAHE	65.41	76.97	86.3	90.29

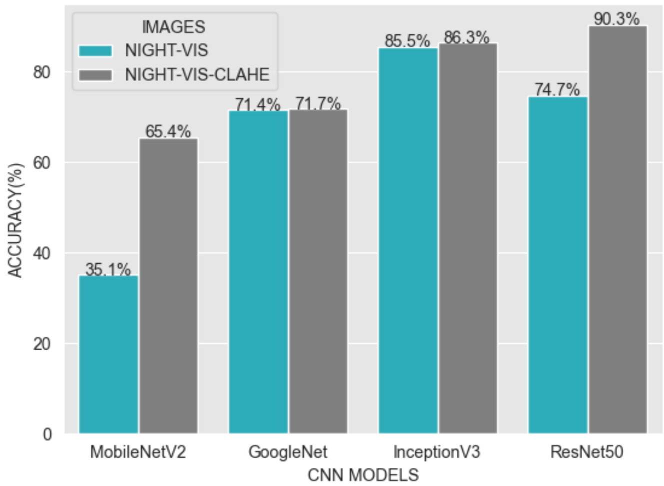


Figure 12. Performance of behavior recognition based on CLAHE implementation for night driving

4.3.1. Performance comparison between NIGHT-IR and NIGHT-VIS-CLAHE

Figure 13 provides a comparative analysis of the performance of ResNet50, InceptionV3, MobileNetV2, and GoogleNet, across two specific sessions: NIGHT-IR and NIGHT-VIS-CLAHE. The objective of this comparison is to analyze the performance difference between using normal camera sensor with CLAHE implementation vs. using infrared sensors. According to Figure 13, in the NIGHT-IR session, InceptionV3 achieved the highest accuracy of 89.27%, followed by GoogleNet with 88.9%, MobileNetV2 with 86% and ResNet50 with 80.9%. For MobileNetV2, GoogleNet, and InceptionV3, NIGHT-IR captured with an infra-red sensor showed better performance compared to NIGHT-VIS-CLAHE. MobileNetV2 showed about a 20% recognition performance difference between normal and infra-red sensors, whereas GoogleNet and InceptionV3 were 11% and 3% better respectively when using infra-red images. However, for both NIGHT-IR and NIGHT-VIS-CLAHE sessions, ResNet50 emerges as the top performer with an accuracy of 90.29%. The results could serve as a valuable reference for selecting the most suitable CNN model depending on the specific dataset or session, highlighting the variations in their performance across different scenarios. For example, InceptionV3 is the best-performing model if the system were to use infrared sensors to capture driving behaviors at night.

4.4. Class recognition performance for ResNet50 NIGHT-VIS-CLAHE implementation

Here we used a confusion matrix to analyze the class performance for our proposed ResNet50 NIGHT-VIS-CLAHE method. A confusion matrix is a critical tool in the evaluation of a classification model's performance. It presents a comprehensive breakdown of how the model's predictions compare to the actual labels across different classes. In confusion matrix, each row represents the actual class, while each column represents the predicted class.

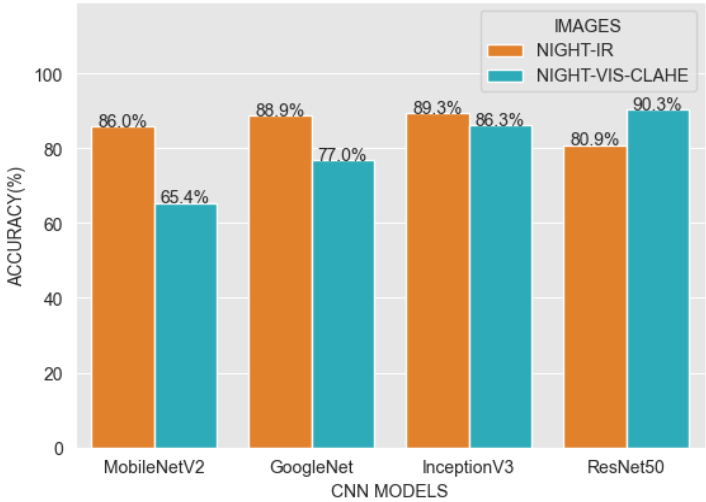


Figure 13. Performance comparison between NIGHT-IR and NIGHT-VIS-CLAHE

For instance, according to Table 5, in the row labelled calling, the model correctly predicted 3073 instances as calling, while it incorrectly predicted 25 instances of calling as yawning. Similarly, in the nodding row, there were 54 instances of nodding that were correctly predicted as such, but 122 instances of nodding were incorrectly classified as texting. From the confusion matrix, we can derive the precision and recall performance for each class. Precision values provide insights into the accuracy of the model's positive predictions for each class. For calling, the precision is 99.19%, indicating that when the model predicts calling, it is correct about 99.19% of the time. Recall values, on the other hand, reveal how well the model captures all the instances of each class. In the calling class, the recall is 98.40%, indicating that the model correctly identifies 98.40% of all the actual instances of calling. Overall, calling behavior has the highest precision, followed by talking, normal, and yawning. Nodding has the lowest precision with 17.09% indicating that the model struggles to correctly recognize this behavior. Most of the times, nodding was misclassified as normal or texting.

Table 5. Confusion matrix of NIGHT-VIS-CLAHE ResNet50

Calling	3073	0	0	0	0	25	99.19%
Nodding	3	54	94	35	122	8	17.09%
Normal	6	0	2390	2	220	59	89.28%
Talking	3	0	5	362	15	0	94.03%
Texting	20	52	183	6	2793	1	91.42
Yawning	18	0	58	10	7	649	87.47%
	Calling	Nodding	Normal	Talking	Texting	Yawning	
	98.40%	50.94%	87.55%	87.22%	88.47%	87.47	

5. CONCLUSION

In conclusion, this work addresses the challenges of night-time driving behavior under low-illumination condition by proposing an innovative approach using ResNet50 with CLAHE implementation. The integration of CLAHE into the model's training process enhances its robustness, allowing accurate recognition of distracted driving behaviors in low-light conditions. We also showed the performance of behavior recognition using images captured by infra-red sensors and compared against CLAHE. Experimental results demonstrated significant improvements in model performance for NIGHT-VIS-CLAHE. Notably, the ResNet50 model delivered the best accuracy rates of 90.73% when tested with NIGHT-VIS-CLAHE with a

16% improvement over NIGHT-VIS images. The InceptionV3, GoogleNet, and MobileNetV2 models also show improved recognition accuracy through CLAHE. Additionally, MobileNetV2 showed about 20% better performance using infra-red sensors for low-illumination conditions, whereas GoogleNet and InceptionV3 were 11% and 3% better respectively. ResNet50 NIGHT-VIS-CLAHE still performs better compared to the best performance of day-driving conditions by InceptionV3 at 87.99% accuracy. In summary, the study's findings emphasize the potential of deep learning and CLAHE in improving night-time driving behavior recognition and advancing road safety, with implications for integration into advanced driver assistance systems and autonomous vehicles.

ACKNOWLEDGEMENTS

The authors extend their gratitude to the Ministry of Science, Technology, and Innovation for their support through the Technology Development Fund 1 (TDF04211376, 600-RMC/MOSTI-TeD1/5/3 (008/2021)) and to the College of Engineering at Universiti Teknologi MARA for their assistance in this project.

REFERENCES

[1] A. H. Azhari, F. A. Md Zaidi, M. H. Ahmad Anuar, and J. Othman, "A statistical analysis of road accident fatalities in Malaysia," *International Journal of Academic Research in Business and Social Sciences*, vol. 12, no. 5, May 2022, doi: 10.6007/IJARBSS/v12-i5/13058.

[2] M. F. Musa, S. A. Hassan, and N. Mashros, "The impact of roadway conditions towards accident severity on federal roads in Malaysia," *PLOS ONE*, vol. 15, no. 7, Jul. 2020, doi: 10.1371/journal.pone.0235564.

[3] L. Blincoe *et al.*, "The economic and societal impact of motor vehicle crashes, 2019 (Revised)," National Highway Traffic Safety Administration, 2023.

[4] M. S. H. S. Bahari and L. Mazalan, "Distracted driver detection using deep learning," in *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, May 2022, pp. 198–203. doi: 10.1109/CSPA55076.2022.9781938.

[5] J. Hyun, J. Kim, C.-H. Choi, and B. Moon, "Hardware architecture of a haar classifier based face detection system using a skip scheme," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2021, pp. 1–4. doi: 10.1109/ISCAS51556.2021.9401114.

[6] M. Dwisnanto Putro, D.-L. Nguyen, and K.-H. Jo, "Fast eye detector using CPU based lightweight convolutional neural network," in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, Oct. 2020, pp. 12–16. doi: 10.23919/ICCAS50221.2020.9268234.

[7] R. N. Yousef, A. T. Khalil, A. S. Samra, and M. M. Ata, "Accurate metaheuristic deep convolutional structure for a robust human gait recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 7005–7015, Dec. 2023, doi: 10.11591/ijece.v13i6.pp7005-7015.

[8] N. Tran, T. Nguyen, M. Nguyen, K. Luong, and T. Lam, "Global-local attention with triplet loss and label smoothed crossentropy for person re-identification," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 4, pp. 1883–1891, Dec. 2023, doi: 10.11591/ijai.v12.i4.pp1883-1891.

[9] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, Nov. 2021, doi: 10.1007/s42979-021-00815-1.

[10] M. Chen, Y. Lin, Z. Huang, and X. Zeng, "Driver distraction recognition based on transfer learning and feature fusion," in *2021 5th International Conference on Communication and Information Systems (ICCIS)*, Oct. 2021, pp. 160–164. doi: 10.1109/ICCIS53528.2021.9646003.

[11] R.-C. Chen, C. Dewi, Y.-C. Zhuang, and J.-K. Chen, "Contrast limited adaptive histogram equalization for recognizing road marking at night based on Yolo models," *IEEE Access*, vol. 11, pp. 92926–92942, 2023, doi: 10.1109/ACCESS.2023.3309410.

[12] Z. Yuan *et al.*, "CLAHE-based low-light image enhancement for robust object detection in overhead power transmission system," *IEEE Transactions on Power Delivery*, vol. 38, no. 3, pp. 2240–2243, Jun. 2023, doi: 10.1109/TPWRD.2023.3269206.

[13] B. S. Min, D. K. Lim, S. J. Kim, and J. H. Lee, "A novel method of determining parameters of CLAHE based on image entropy," *International Journal of Software Engineering and Its Applications*, vol. 7, no. 5, pp. 113–120, Sep. 2013, doi: 10.14257/ijseia.2013.7.5.11.

[14] F. Sajid, A. R. Javed, A. Basharat, N. Kryvinska, A. Afzal, and M. Rizwan, "An efficient deep learning framework for distracted driver detection," *IEEE Access*, vol. 9, pp. 169270–169280, 2021, doi: 10.1109/ACCESS.2021.3138137.

[15] S. Hwang, P. Lee, S. Park, and H. Byun, "Learning subject-independent representation for EEG-based Drowsy driving detection," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, Feb. 2021, pp. 1–3. doi: 10.1109/BCI51272.2021.9385364.

[16] S. V. Deshmukh and O. Dehzangi, "ECG-based driver distraction identification using wavelet packet transform and discriminative kernel-based features," in *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, May 2017, pp. 1–7. doi: 10.1109/SMARTCOMP.2017.7947003.

[17] S. Bhattacharya and S. Bernadin, "Correlation studies on the cognitive responses of distracted drivers during conversational tasks," in *2020 SoutheastCon*, Mar. 2020, pp. 1–5. doi: 10.1109/SoutheastCon44009.2020.9249678.

[18] C. Huang, X. Wang, J. Cao, S. Wang, and Y. Zhang, "HCF: A hybrid CNN framework for behavior detection of distracted drivers," *IEEE Access*, vol. 8, pp. 109335–109349, 2020, doi: 10.1109/ACCESS.2020.3001159.

[19] Y. Zhao, K. Xie, Z. Zou, and J.-B. He, "Intelligent recognition of fatigue and sleepiness based on inceptionV3-LSTM via multi-feature fusion," *IEEE Access*, vol. 8, pp. 144205–144217, 2020, doi: 10.1109/ACCESS.2020.3014508.

[20] D. Tran, H. M. Do, J. Lu, and W. Sheng, "Real-time detection of distracted driving using dual cameras," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, pp. 2014–2019. doi: 10.1109/IROS45743.2020.9340921.

[21] V. Unadkat, P. Sayani, H. Kapadia, P. Shah, and H. Dalvi, "Automated system for detecting distracted driver," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Dec. 2018, pp. 1–4. doi: 10.1109/CCAA.2018.8777709.
[22] C. Wei, C. Liu, and S. Chi, "A driver distraction detection method based on convolutional neural network," in *2022 7th International Conference on Computer and Communication Systems (ICCCS)*, Apr. 2022, pp. 272–277. doi: 10.1109/ICCCS55155.2022.9846062.
[23] M. F. Rahman, F. Sthevanie, and K. N. Ramadhani, "Face recognition in low lighting conditions using fisherface method and CLAHE techniques," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Jun. 2020, pp. 1–6. doi: 10.1109/ICoICT49345.2020.9166317.
[24] I. Lashkov, R. Yuan, and G. Zhang, "Edge-computing-facilitated nighttime vehicle detection investigations with CLAHE-enhanced images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 13370–13383, Nov. 2023, doi: 10.1109/TITS.2023.3255202.
[25] C. Nie, S. Zhou, H. Zhang, and Z. Sun, "Monocular vision based perception system for nighttime driving," in *2022 8th International Conference on Control, Automation and Robotics (ICCAR)*, Apr. 2022, pp. 258–263. doi: 10.1109/ICCAR55106.2022.9782645.
[26] R. Poojary, R. Raina, and A. K. Mondal, "Effect of data-augmentation on fine-tuned CNN model performance," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 84–92, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp84-92.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[28] "Understanding ResNet50 architecture." https://iq.opengenus.org/resnet50-architecture/ (accessed Sep. 22, 2023).
[29] "MobileNetV2 explained | papers with cod." https://paperswithcode.com/method/mobilenetv2 (accessed Sep. 22, 2023).
[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

BIOGRAPHIES OF AUTHORS

Muhammad Firdaus Ishak is an aspiring graduate student currently pursuing a Master of Science in Electrical Engineering (Research) at Universiti Teknologi MARA, Malaysia. Dedicated to the field of Vehicle Intelligence and Telematics, he actively contributes as a member of the esteemed VITAL team at the College of Engineering, Universiti Teknologi MARA. Focused on the implementation of deep learning networks, his research primarily centers around detecting distracted driving behaviors in night driving scenarios, emphasizing the critical domain of image processing. With a solid academic background, including a Bachelor of Electrical Engineering (Hons) in Electronics Engineering and a Diploma of Electrical Engineering (Hons) in Electronics Engineering, he recently showcased his expertise as a presenter at the 11th International Conference on Information and Communication Technology. For further communication, he can be contacted at email: 2021655474@student.uitm.edu.my.

Fadhlan Hafizhelmi Kamaru Zaman received the B.Sc. (Hons.) and Ph.D. degrees from International Islamic University Malaysia in 2008 and 2015, respectively. He is currently an Associate Professor and the Head of Vehicle Intelligence and Telematics Lab (VITAL), College of Engineering, Universiti Teknologi MARA, Malaysia. His research interests are in pattern recognition, signal and image processing, artificial intelligence, and computer vision. Fadhlan is a Professional Engineer with Board of Engineer Malaysia (BEM) member of IEEE, Professional Technologist with Malaysian Board of Technologist (MBOT), and a Chartered Engineer from the Institution of Engineering and Technology, UK. He has published more than 50 research articles on pattern recognition, signal and image processing, artificial intelligence, and computer vision. He is actively involved in research and currently supervising more than 40 students, including Ph.D., Master and Final Year students. He has experience in many consultation projects with various private companies, mainly focusing on implementing artificial intelligence in industries. He has also been awarded several national and international grants for research in artificial intelligence. He can be contacted at email: fadhlan@uitm.edu.my.

Ng Kok Mun received the B.Eng. and M.Eng. degrees in systems engineering from the Universiti Tun Hussein Onn, Malaysia, in 2002 and 2006 respectively. He obtained the Ph.D. degree in Systems Engineering from Universiti Kebangsaan Malaysia, Bangi, Malaysia in 2016. From 2002 to 2003, he worked as a system engineer at Matromatics Handling Systems, Malaysia. He was with Stamford College, Malaysia from 2006 to 2008 as a lecturer before joining the School of Electrical Engineering, Universiti Teknologi MARA, where he works as a senior lecturer at the Systems Engineering department. His research interests include system modeling and optimization, artificial intelligence, intelligent transportation systems, mapping and localization. He is a sub-editor of the Journal of Electrical and Electronic Systems Research (JEESR) and a research fellow at the university vehicle telematics industrial research lab. He can be contacted at email: kokmun77@yahoo.com.sg.

Syahrul Afzal Che Abdullah obtained BEng degree in Electronic Engineering from University of Southampton, UK, MSc in Real Time Software Engineering from Universiti Teknologi Malaysia (UTM), MY, and Ph.D. in Software Engineering from Universiti Sains Malaysia (USM), MY. Currently, he is a senior lecturer at the Centre for Computer Engineering Studies, School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), MY. His research interests include Artificial Intelligence, Autonomous Vehicle, Internet of Things, Software Engineering and Software Testing. He is recognized as a Professional Engineer by the Board of Engineers Malaysia, a corporate member of the Institution of Engineers Malaysia (IEM), and a senior member of Institute of Electrical and Electronics Engineers (IEEE), US. He is recognized as Graduate Technologist and Professional Technologist by the Malaysia Board of Technologists (MBOT). He can be contacted at email: bekabox181343@uitm.edu.my.

Ahmad Khushairy Makhtar received the Bachelor's Degree in Mechanical Engineering from Tokyo University of Science, Master's Degree in Risk Engineering from the University of Tsukuba, and Doctor of Philosophy in Risk Engineering from the University of Tsukuba. He is currently a senior lecturer at the School of Mechanical Engineering, College of Engineering, Universiti Teknologi MARA. He is a Profesional Engineer with Board of Engineer Malaysia (BEM), vice president 1 Human Factors and Ergonomics Society Malaysia (HFEM), Profesional Technologist from the Malaysia Board of Technologies (MBOT) and a Chartered Engineer from the Institution of Mechanical Engineers (IMeche). He also a certified Ergonomics Trained Person endorsed by DOSH. He had supervised a total of 29 undergraduate students. He had completed many research projects sponsored by the ministry of Science and Innovation and Ministry of Higher Education. His current research focus is on Human Factors and Ergonomics with engineering approaches. He published numbers of article in the international level conferences and journals. He can be contacted at email: ahmadkhushairy@uitm.edu.my.