Article

FADS: An Intelligent Fatigue and Age Detection System

Mohammad Hijji [1,*,†] , Hikmat Yar [2,†], Fath U Min Ullah [3], Mohammed M. Alwakeel [1] , Rafika Harrabi [1], Fahad Aradah [1], Faouzi Alaya Cheikh [4], Khan Muhammad [5,*] and Muhammad Sajjad [2,4,*]

1 Faculty of Computers and Information Technology, University of Tabuk, Tabuk 47711, Saudi Arabia
2 Digital Image Processing Laboratory, Islamia College Peshawar, Peshawar 25000, Pakistan
3 Department of Software Convergence, Sejong University, Seoul 143-747, Republic of Korea
4 The Software, Data and Digital Ecosystems (SDDE) Research Group, Department of Computer Science, Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway
5 Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, Republic of Korea
* Correspondence: m.hijji@ut.edu.sa (M.H.); khanmuhammad@g.skku.edu (K.M.); muhammad.sajjad@ntnu.no (M.S.)
† These authors contributed equally to this work.

Abstract: Nowadays, the use of public transportation is reducing and people prefer to use private transport because of its low cost, comfortable ride, and personal preferences. However, personal transport causes numerous real-world road accidents due to the conditions of the drivers' state such as drowsiness, stress, tiredness, and age during driving. In such cases, driver fatigue detection is mandatory to avoid road accidents and ensure a comfortable journey. To date, several complex systems have been proposed that have problems due to practicing hand feature engineering tools, causing lower performance and high computation. To tackle these issues, we propose an efficient deep learning-assisted intelligent fatigue and age detection system (FADS) to detect and identify different states of the driver. For this purpose, we investigated several neural computing-based methods and selected the most appropriate model considering its feasibility over edge devices for smart surveillance. Next, we developed a custom convolutional neural network-based system that is efficient for drowsiness detection where the drowsiness information is fused with age information to reach the desired output. The conducted experiments on the custom and publicly available datasets confirm the superiority of the proposed system over state-of-the-art techniques.

Keywords: artificial intelligence; age prediction; deep learning (DL); drowsiness detection; neural computing; smart surveillance

MSC: 68T07

1. Introduction

Modern cities are linked with crossroads and mass communication channels for rapid transportation that facilitate the daily commutes of millions of people [1]. Despite this, road accidents happen, which is one of the highest causes of people's injuries and deaths. The victims of road accidents often have a permanent disability that stays throughout their life. Accidents are roadside injuries that cause an average of 3242 deaths on a daily basis, which is higher than any other single source in the world [2]. Road crashes are very generic worldwide and are annually estimated by the Association for Safe International Road Travel, 2013, showing that the ratio of deaths in road accidents each year is approximately 1.3 million, where 20–50 million people are injured or permanently disabled. Unless urgent actions are taken, roadside injuries are anticipated to become the fifth leading reason for death by 2030 [3]. Every year, around 328,000 crashes occur in the U.S., which has an annual cost to society of millions of dollars [4]. One of the main reasons for road-related accidents

is the inability of drivers due to age. In most of these accidents, the driver is either under- or overage to drive the vehicle. Another reason is that the drivers risk their own life or the lives of the others around them either due to stress, sleepiness, fatigue, drowsiness, or under the influence of alcohol.

Among the above-mentioned reasons, drowsiness is the most common factor. Driver fatigue or drowsiness is a human state where the victim is unaware of its surroundings. Due to the sleep-deprived state, the driver does not know what may be happening in their surroundings, which reduces their attentiveness and leads to road accidents. Millions of people are killed and injured every year due to the driver's state such as sleeping while driving [5,6]. Drowsiness decreases the attentiveness, head, and gaze of the drivers due to which the ratio of road accidents is also increasing. Some studies have revealed that driver drowsiness causes 20% of road accidents, resulting in 50% of serious injuries or death [7–10]. Drivers are usually aware of their drowsiness and can decide to continue or stop driving to rest as most fatal accidents are caused by tired drivers. According to the National Highway Traffic Safety Administration [7], 56,000 crashes occur every year where drowsiness or fatigue was cited by the police as a causal factor that leads to 1550 fatalities and 40,000 nonfatal injuries on an average basis. Similarly, 15% to 44% of crashes take place in the U.S. and Australia [8–10], where 18.6%–30% of heavy vehicle crashes involve fatigue [11,12]. About 30,000 vehicle crashes have caused injuries that were also due to fatigue [13]. Moreover, fatigued or sleepy commercial vehicle drivers have a 21 times greater risk of causing fatal accidents, and safety-related drivers had higher drowsiness levels than other drivers [7]. Along with other issues such as fatigue and drowsiness, most of the accidents involved either underage or overage drivers. According to the report, it is estimated that over 2000 drivers between the ages of 13–19 died in the U.S in 2009 [14].

Several approaches have been suggested by the research community to overcome road accidents. Most of them are working on scalar sensors to detect the driver's heartbeat and temperature while several rely on vision sensors. Most of the existing studies have used complex networks that are costly and difficult to deploy over edge devices. Similarly, the existing methods are limited to detecting the different states of the drivers including age. Thus, to tackle the problems and challenges, we proposed "FADS: An Intelligent Fatigue and Age Detection System" by using several realistic approaches for drowsiness detection and age classification based on facial feature analysis to keep underage and overage people from driving when an alarm is generated when the driver's state is detected as drowsy, angry, or sad. We used a lightweight CNN that is easily deployable over an edge device (e.g., Jetson Nano) to perform real-time processing [7], making it suitable for smart surveillance and the Internet of vehicles. The major contributions of the proposed FADS are summarized as follows:

- We developed a DL-assisted FADS for driver mood detection from an easy-to-deploy resource-constrained vision sensor. Addressing this issue of complex systems, it can overcome high computational costs and ensure the real-time detection of the driver's mood.
- Age is an important factor in avoiding most of the accidents, and for this purpose, the proposed FADS extracts facial features to classify the driver's age. If the classified age is beyond the defined threshold (age <18 and age >60), then an alert is generated to notify the nearby vehicles and the authorized department. Another influencing factor that causes road accidents is drowsiness or driver moods such as anger or sadness. Therefore, their prediction is also performed by the facial features using a lightweight CNN. These factors can avoid most accidents and ensure safe vehicle driving.
- Due to data unavailability, we created a new dataset for FADS as a step toward the smart system, which includes five classes (i.e., active, angry, sad, sleepy, and yawning). Furthermore, a UTKFace dataset was categorized into three classes (i.e., underage (age <18), middle-age (age ≥18 or ≤60), and overage (>60)) for detailed analysis. This categorization further enhances FADS by fusing dual features to reach an optimum outcome, which is needed for smart surveillance.

- Extensive experiments were conducted from different aspects and the results over the baseline CNNs confirm that the proposed FADS achieved state-of-the-art performance on the standard and the new dataset in terms of lower model complexity and good accuracy.

The remainder of the paper is structured as follows. A compact literature review is presented in Section 2. We cover the proposed FADS in Section 3. The experimental results of FADS and its comparisons are given in Section 4. Section 5 concludes our work with some future research directions.

2. Literature Review

Facial images can be used for the analysis of a driver's behavior based on drowsiness or driver mood detection (i.e., anger or sadness). For instance, the technique named "DriCare" [15] used face landmarks and their key points to detect faces and track them for driver fatigue detection. This includes eye blinking, eye closure, or yawning. Next, Verma et al. [16] enhanced the strategy using two VGG16 CNN parallels to detect the driver expressions. First, the region of interest was detected, which was fed to the VGG16 model as the input while the face landmarks and key points were used as the input in the second VGG16 model. Their combined results were used for fatigue detection. In another approach [17], a dataset called DROZY (ULG Multi-modality Drowsiness dataset) was developed for drowsiness detection. Tsaur et al. [18] proposed a real-time system for driver abnormality detection using edge-fog computing and achieved a promising performance. Furthermore, Xing et al. [19] attempted to detect seven different tasks performed by the drivers such as normal driving, using a mobile phone, checking left and right mirrors, and setting up video devices in a vehicle. They extracted 42 different features using a Kinect camera and used random forests for classification. Yu et al. [20] employed a condition-adaptive representation method for driver drowsiness detection. Their system was evaluated on the NTHU driver drowsiness detection video dataset, which outperformed the state-of-the-art methods based on visual analysis. In the next approach, Dua et al. [21] used four different DL models such as ResNet, FlowImageNet, VGG-FaceNet, and AlextNet for drowsiness detection. However, the time complexity and limited accuracy restricted their system from real-world deployment. Abdelmalik et al. [22] proposed a four-tier approach for driver drowsiness detection consisting of face detection and alignment, pyramid multi-level face representation, face description using multi-level features extraction, and features subset selection. Likewise, in [23], the authors used a DL approach for eye state classification in static facial images, where they fused two deep neural networks for a better decision. Recurrent convolutional neural networks (R-CNN) have also played an important role in detecting the driver's state such as normal blinking or a falling asleep situation from the sequences of the frames [15,24]. Ghoddoosian et al. [25] presented a technique for eye blink detection based on hierarchical multi-scale long short-term memory.

Aside from drowsiness detection, age classification based on facial feature analysis is a trending area due to its wide range of applications such as human–computer interaction, security, and age-oriented commercial advertisement. Several traditional methods [26,27] and DL methods [28–31] have been presented for age classification. For instance, [32] presented a technique in which they used VGG16 CNN architecture for age classification by creating the dataset "IMDB-WIKI". Similarly, Shen et al. [33] presented deep regression forests for end-to-end feature learning for age estimation. The authors in [34] predicted age using a directed acyclic graph CNN. Furthermore, Lou et al. [35] presented an expression invariant age classification method by concurrently learning the age and expression. They studied the correlation between age and expression by deploying a graphical model that adopted a hidden layer. In [36], the researchers presented an ordinal DL mechanism by learning features for both age estimation and face representation.

In the aforementioned techniques, several researchers have individually contributed to the area of driver fatigue detection and age classification. However, none of them could detect the different driver states along with their age. Therefore, we proposed FADS for

both driver fatigue detection and age classification, restricting underage and overage people from driving, and generating an alarm in the case of detecting drivers in a fatigued state.

3. Fatigue and Age Detection System

In this section, the proposed system is explained in detail. First, the face was detected using an improved Faster R-CNN algorithm. Next, different CNN models were used to examine various facial features for age classification and the driver mood or state detection. Finally, we fused both the driver's age and mood information to provide an effective solution. Furthermore, the proposed system was deployable over edge devices where a vision sensor captures the live stream images/videos, and an edge device was mounted on top of the dashboard in a car. The video stream is processed, and the age and overall state of the driver are predicted in real-time. The proposed system was divided into the following steps: face detection, drowsiness detection, age classification, and fusion strategy, as demonstrated in Figure 1.
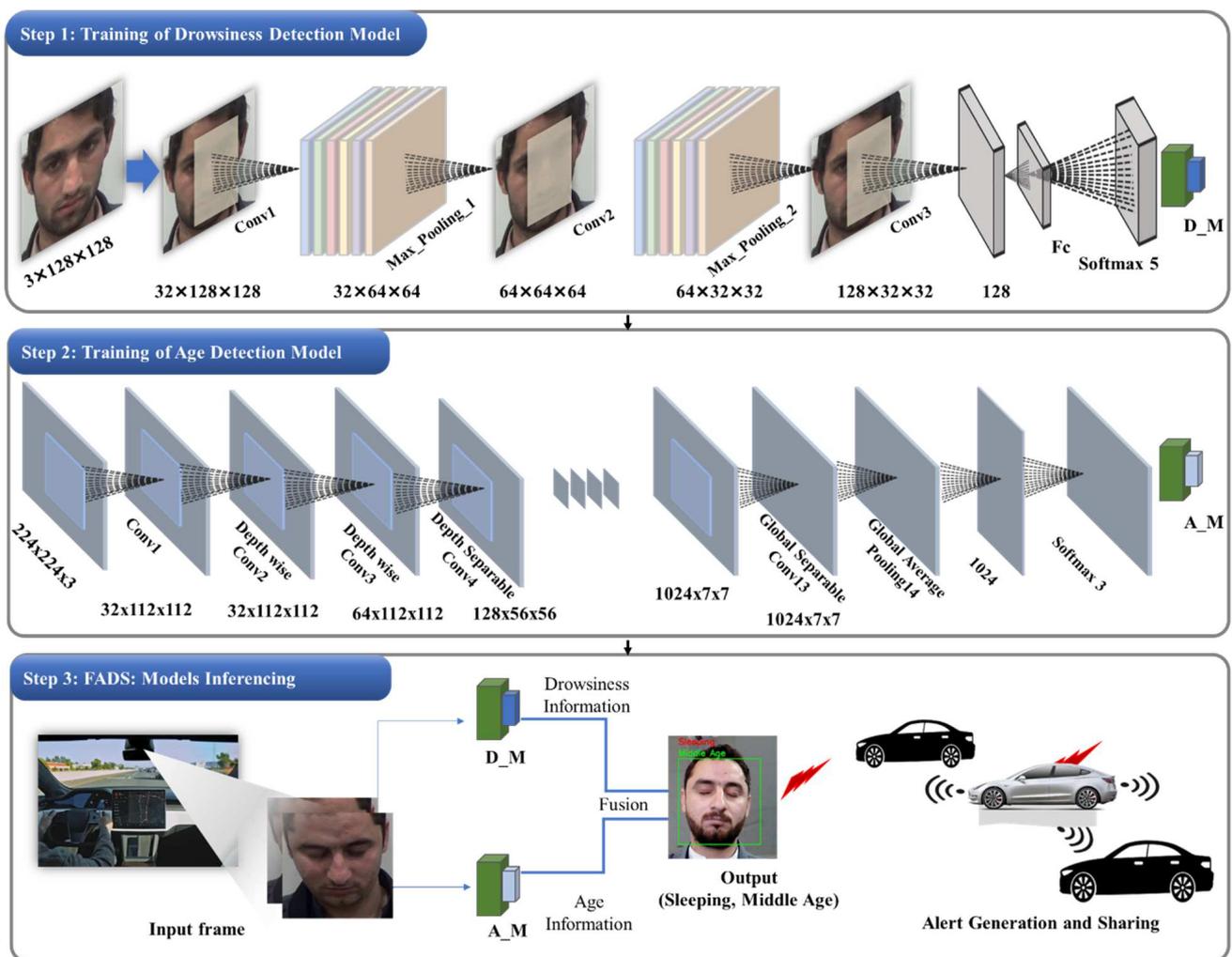
Figure 1. The proposed FADS consists of four different stages: (1) receives an input frame from the edge device, (2) input frames are received where face detection is performed via an efficient algorithm. (3) Two CNN networks are employed for feature extraction and classification purposes and (4) an output label is obtained from the fusion information, which is further sent to the nearest vehicle and authorized authorities for safety concerns.

3.1. Face Detection

Face detection is the most imperative problem that has been intensively surveyed in the last few decades. Early researchers were mainly concerned with hand-crafted feature extraction methods [37,38]. However, there are some limitations in these techniques. They often require experts in the field of image processing to extract effective and useful features where each component is optimized individually, making the entire pipeline of the detection process often sub-optimal. Therefore, Sun et al. [39] proposed a technique to extend the state-of-the-art Faster R-CNN method [40]. Their approach increased the existing Faster R-CNN approach by fusing multiple important schemes, consisting of feature fusion, multi-scale training, and hard negative mining. We employed a similar strategy to Sun et al. [39] to capture the face images. This strategy contained two main steps: a region proposal network to capture the regions of interest and a Fast R-CNN network to classify the region into its corresponding category. Sun et al. [39] trained a Faster R-CNN with the WIDER Face dataset [31]. Furthermore, the targeted dataset was used to test the model to generate hard negatives, which were then fed into the network during training as a second step. By training these hard negative samples, the trained model was capable of generating a lower false positive rate. Moreover, their model was fine-tuned on the FDDB dataset. In the final phase, they employed the multi-scale training process and adopted a feature-fusing strategy to improve the model performance. For the entire training procedure, an end-to-end model training strategy such as Faster R-CNN was used due to its effective performance. Finally, the resulting detection bounding boxes were converted into ellipses as the regions of human faces. Therefore, we employed an improved Faster R-CNN approach for efficient and accurate face detection in the FADS, whose sample results are visualized in Figure 2.

Figure 2. Visual demonstration of the face detection algorithm used in the proposed system.

3.2. Driver Drowsiness Detection

Inspired by the performance of the MobileNet [41–43], GoogleNet [44], SqueezeNet with deep autoencoder [45], Inception [46,47], Darknet [48,49], and ConvLSTM [50] models, we implemented a new custom network consisting of three convolution layers, where a max-pooling layer was used after each convolutional layer and two dense layers, as demonstrated in Figure 3. In the first convolutional layer, the input image size was $128 \times 128 \times 3$ with 32 different kernels. Each kernel's size was 3×3 with a one-pixel stride, on which the pooling operation was applied. The output of the first convolutional layer is the input of the second convolutional layer, and there was a total of 64 kernels having a size of 3×3. The third convolution layer had 128 kernels with the size of 3×3 connected to the output of the previous layer. The fully connected layer had 64 neurons, which fed the output to the softmax classifier to classify the input source image into their corresponding classes. The custom architecture learns 2.2 million parameters compared to all of the above-the mentioned models such as the AlexNet [51], Vgg16 [52], MobileNet [41], GoogleNet [44], Inception [46], and MobileNet [41] models. Custom architecture computes an extremely smaller number of parameters due to the size of the input image and several filters selected during the convolution.
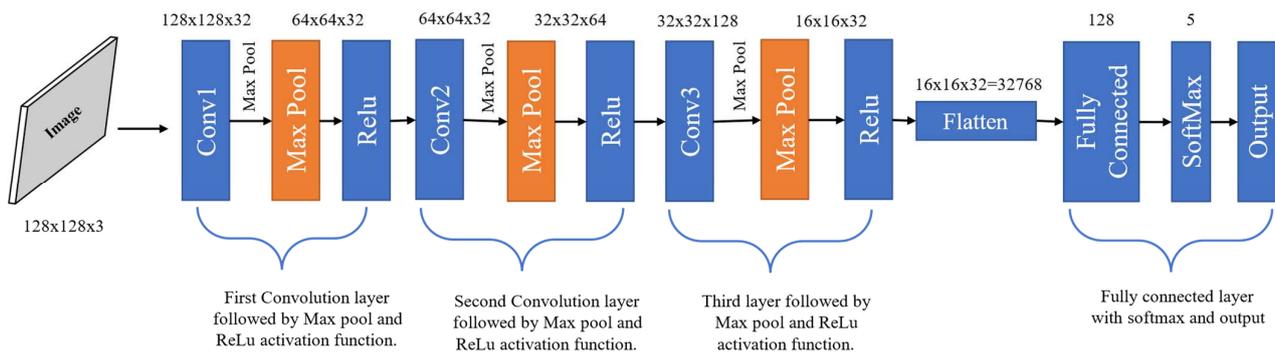
Figure 3. Layer-by-layer architecture of the proposed system.

3.3. Driver Age Classification

For driver age classification in the FADS, we used a fine-tuned lightweight model for prominent facial feature extraction and classification. In the proposed system, we employed MobileNet, a depth-wise separable convolutional neural network, which is a lightweight DCNN and provides an efficient system for embedded vision applications. In this model, the depth-wise separable convolutions are composed of point convolution filters (PCF) and depth-wise convolution filters (DCF). The DCF performs a single convolution on each channel, and PCF combines the output of the DCF linearly with 1×1 convolutions, as shown in Figure 4. The output of the depth-wise separable convolution using RGB images with a 3×3 kernel size and a movement interval of 1 is given in Equations (1) and (2) [53].

$$\hat{O} = \sum_{i=1}^{3} \sum_{j=1}^{3} K_{j,i,c}.F_{x+i,y+j-1,c} \tag{1}$$

$$O = \sum_{c=1}^{3} \check{K}_{c,n} + \hat{O}_{x,y,v} \tag{2}$$

where $\hat{O}$ shows the output result of the depth-wise convolution; K is the kernel; and F is the input. In Equation (2), O represents the output of the pointwise convolution and $\check{K}$ is the kernel of the 1×1 convolution. We employed the above-mentioned strategy of PCF and DCF and modified the MobileNet architecture consisting of 28 convolution layers including deep convolutional layers (point convolution layer) 1×1, batch-normalization, ReLu activation, average pooling, and a softmax layer. In this architecture, the ReLU activation is employed in each convolutional layer to perform a thresholding operation, where each input value less than 0 is set to 0, and positive values remain the same. MobileNet consists of a pooling layer strategy, which summarizes the outputs of neighboring groups of neurons. The pooling layer is used for dimensionality reduction, which influences the duration of the network training, and the output neurons are equal to the number of classes in the dataset recognized by the network. Finally, softmax is used for probabilities to classify a driver's age. This probability is the basis for making the final decision about the classification result. In summary, we employed an efficient MobileNet model for driver age classification into three different categories such as underage, middle age, and overage.

3.4. Fusion Strategy in FADS

This subsection explains the fusion strategy of the proposed system to achieve the desired output. The CNN-based architecture was used for face detection, drowsiness detection, and driver age classification. Our system consisted of three steps. (1) The input frames were acquired from the vision sensor mounted with Jetson Nano, where the face is detected through an improved Faster R-CNN. The detected face was cropped from the entire image and fed into CNN for drowsiness detection and driver age classification. (2) The detected face was processed by our new custom CNN architecture that performed

the drowsiness detection of the individual driver present in the entire frame. (3) The driver age was computed and classified using a customized version of the MobileNet architecture. Finally, we fused these models during the inference time to achieve the desired output, as shown in Figure 1.
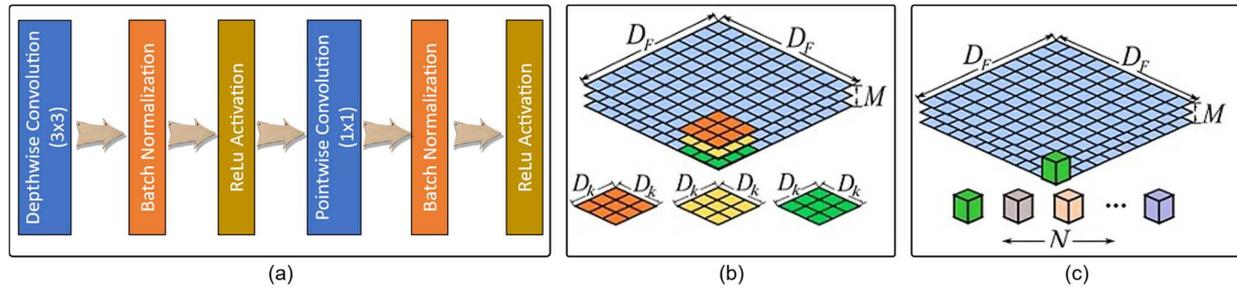
Figure 4. MobileNetV2 architecture where (**a**) represents the depth-wise and pointwise layers followed by batch-normalization and the ReLU activation function, (**b**) depth-wise convolutional layer, and (**c**) pointwise convolutional layer [54].

4. Results and Discussion

This section provides a detailed explanation of the hardware configurations, the datasets used for the driver age classification and drowsiness detection, and the training and testing process in the evaluation. Furthermore, quantitative, and qualitative assessments were performed with the state-of-the-art for both driver age and drowsiness detection. For the training process, we categorized both datasets into three subclasses (i.e., training, testing, and validation) with the proportion of 70%, 10%, and 20%, respectively.

4.1. System Configuration and Evaluation

The proposed system was trained using NVidia GPU GTX 1070 GPU, which has 8 GB of RAM and a 2.9 GHz processor. The operating system, programming language, and libraries used in our work are listed in Table 1.

Table 1. Software specification and libraries used for the proposed system.

Name	Configuration
OS	Window 10
Programming language and IDE	Jupyter Notebook, Python 3.7.2
Libraries	TensorFlow, PyLab, Numpy, Keras, Matplotlib
Imaging libraries	OpenCV 4.0, Scikit-Image, Scikit-Learn

In the computer vision domain, the trained CNN is mostly assessed by conducting quantitative analysis via commonly used different evaluation parameters including accuracy, F1-measure, precision, and recall (sensitivity). These evaluation metrics can be easily calculated from the confusion matrix by forwarding the predicted and actual labels. The mathematical expression of accuracy, precision, recall, and F1-measure are given in Equations (3)–(6), respectively. The accuracy is considered on the major evaluation matrix to evaluate the overall performance of the system.

$$Accuracy = \left(\frac{TPV + TNV}{TPV + TNV + FPV + FNV} \right) \tag{3}$$

$$Precision = \left(\frac{TPV}{TPV + FPV} \right) \tag{4}$$

$$Recall = \left(\frac{TPV}{TPV + FNV} \right) \tag{5}$$

$$F1 - measure = 2 * \left(\frac{P * R}{P + R} \right) \tag{6}$$

4.2. Dataset Explanation

In the proposed system, we used two datasets: a custom dataset for drowsiness detection and a UTKFace [55] dataset for age classification. The custom dataset had images that were collected from different sources. We used UTKFace [55] for he acaquisition of better results. Each dataset was properly cleaned and labeled. These datasets are described as follows.

The UTKFace is a large-scale publicly available dataset for facial feature analysis to predict age, ranging from 0 to 116 years. It consists of 23,708 RGB facial images, having a resolution of 200 × 200 pixels with .jpg extensions, and annotations of age, ethnicity, and gender. The images cover large variations in pose, occlusion, illumination, facial expression, and resolution. This dataset can be used in a variety of vision-related tasks such as age regression, age estimation, face detection, and landmark localization. In this research, we converted all images to one standard JPG format and made three classes from the dataset, as demonstrated in Figure 5. Their corresponding age factors are given in Table 2 such as underaged (6–18), middle-aged (18–60), and overaged (60+).


Figure 5. Sample images of the modified datasets for age classification (**a**) underage, (**b**) middle age, and (**c**) overage.

Table 2. The information about the ages (in years) of different people.

Class	Age Group
Underage	6–16
Middle age	18–60
Overage	60+

Furthermore, we collected a custom dataset for drowsiness detection. It contains three classes (i.e., active, sleeping, and yawning). Each class had two thousand images, having a resolution of 124 * 124 with three channels Red (G), Green (G), and Blue (B). In the creation of this dataset, a total of 40 university students participated, whose ages were in the range of 16–35 years. Furthermore, the angry and sad classes were taken from the publicly available KDEF [56] dataset. Figure 6 shows sample images of our dataset.
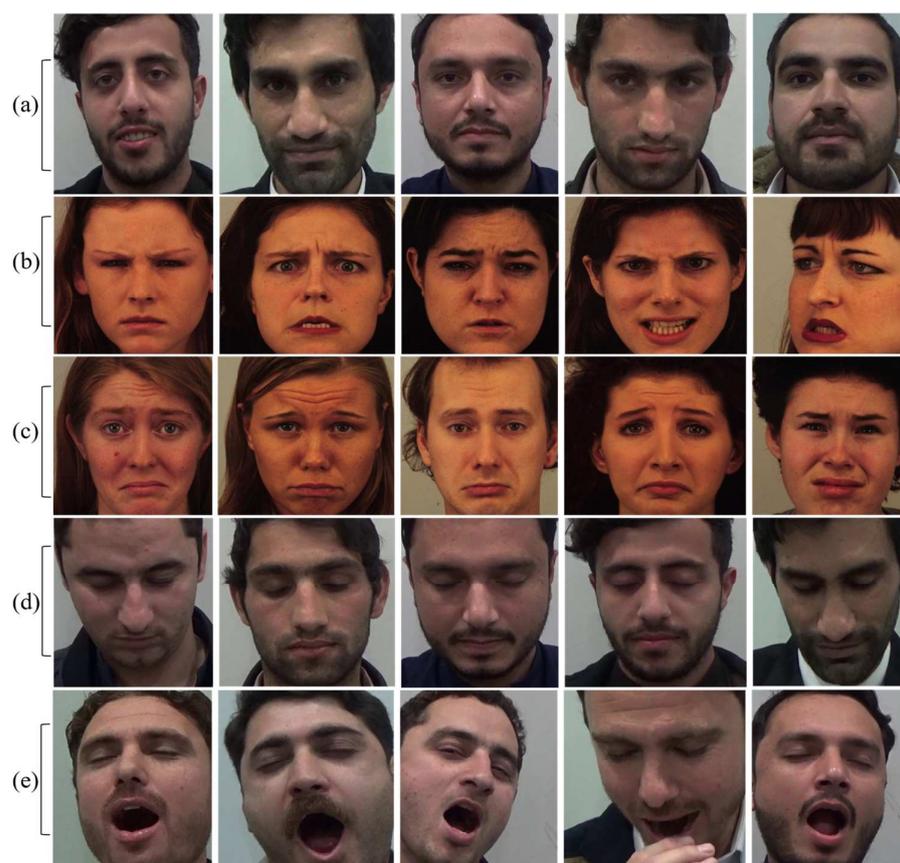
Figure 6. Sample images of the custom dataset. (**a**) Sample images of the active class, (**b**) angry, (**c**) sad, (**d**) sleeping class, (**e**) yawning class.

4.3. Performance Comparison of Different Edge Devices

A single-board-computer or System-on-Chip (SoC) is becoming popular among the research community because of its versatility in different video streaming and machine learning-based applications [57,58]. SoC consists of input and output ports along with enough memory and disk space to run certain applications smoothly. However, the major issue with these devices is that they are usually incapable of using DL models as they do not have enough neural computation capabilities needed for DL model implementation in real-time. For this purpose, we conducted a survey and came up with several different options, among which we selected Nvidia's Jetson Nano as a prime candidate for this application. A list of available options is given in Table 3.

Table 3. Comparison of the different prototyping platforms along with their specifications.

Board	Chip	RAM	OS
Udoo [59]	ARM Cortex A9	1 GB	Debian, Android
Phidgets [60]	SBC	64 MB	Linux
Beagle Bone [61]	ARM AM335 @ 1 Ghz	512 MB	Linux Angstrom
Raspberry Pi 4 [62,63]	Broadcom BCM2711 Processor	2 GB, 4 GB, 8 GB	Raspbian
Jetson Nano [64]	1.43 Ghz Quad Core Cortex A57	4 GB	All Linux Distro

The Jetson Nano has most of the required capabilities compared to other platforms. It is a small-size high-performance computer that can run modern AI applications at low cost and low power. Recently, the AI community is leaning more toward using Jetson Nano as a computational platform for real-time applications because it can run different

AI-based systems for applications (i.e., image segmentation, object detection, and image classification). The Jetson Nano can be powered by micro-USB and it comes with wide-ranging I/O interfaces including general purpose input/output (GPIO) pins that provide an ease of implementation for different sensors to provide an easy-to-use interface for different sensors, as explained in [64].

4.4. Results of Drowsiness Detection

In this subsection, we discuss the experimental results of the drowsiness detection in terms of the confusion matrix, accuracy, and loss graphs, as shown in Figures 7 and 8. During testing, the proposed system was evaluated for each class, and we found that the accuracy of active, sad, and yawning was far better than the angry and sleeping classes, which was 97% and 98%, respectively. We performed various experiments with the help of the above-mentioned dataset. The experiments were performed with different parameters such as a different number of training epochs for the purpose of achieving high accuracy. In Figure 8, we can see that the training accuracy started from 62% and the validation accuracy started from 55% in the first epoch. After each epoch, the accuracy of training and validation improved. In the third epoch, the training accuracy intercepts the line of validation accuracy. Finally, after 30 epochs, the level of training accuracy reached 98% and the validation accuracy reached 97%, as shown in Figure 8.
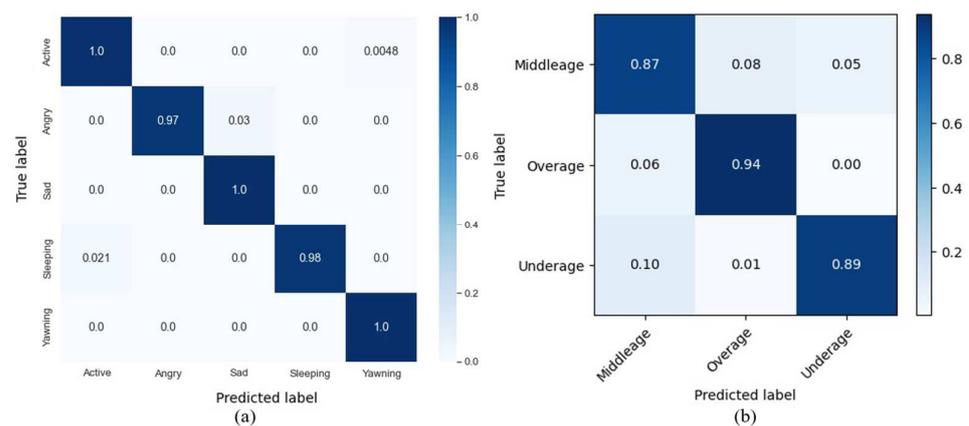

Figure 7. Confusion matrix of our system to validate the class-wise performance. (**a**) Confusion matrix of the drowsiness dataset (**b**) Confusion matrix for the age dataset.
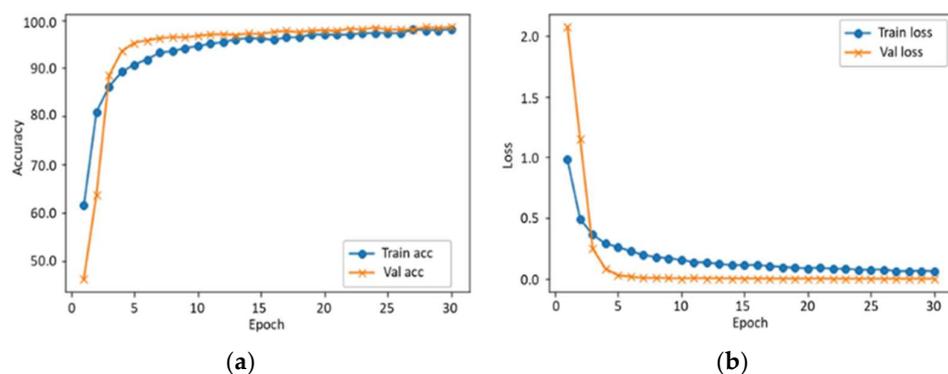

Figure 8. Training/validation accuracy and loss, where (**a**) represents the accuracy and (**b**) represents a loss of drowsiness detection.

We used different evaluation metrics (i.e., recall, precision, and F1-measure) for the performance validation. The results obtained by the proposed system using the drowsiness dataset are given in Table 4. The comparative analysis of the proposed system is given in Table 5, where it was compared with four state-of-the-art systems. The proposed system

reached an accuracy of 98%, where the accuracy of AlexNet, VGG16, ResNet50, and MobileNet was 94.0%, 98.3%, 88.0%, and 93.5%, respectively.

Table 4. Results of the drowsiness detection in terms of the precision, recall, and F1-score.

Driver State	Precision	Recall	F1-Measure
Active	0.98	1	0.99
Angry	1	0.97	0.98
Sad	0.97	1	0.98
Sleeping	1	0.98	0.99
Yawning	0.99	1	0.99

Table 5. Comparison of different DL architectures on the custom drowsiness detection dataset in terms of model size, parameters, and accuracy.

Technique	Model Size (MB)	Parameters (Million)	Accuracy (%)
AlexNet [51]	233	60	94.0
VGG16 [52]	528	138	98.3
ResNet50 [65]	98	20	88.0
MobileNet [41]	13	4.2	93.5
The proposed system	15	2.2	98.0

4.5. Results of Age Classification

A detailed explanation of age classification using different DL architectures is given in this section. In Figure 9, we demonstrated the accuracy of MobileNet, which had the highest accuracy in our experiments. Figure 9 represents the training accuracy, which started from 64% in the first epoch, whereas the validation accuracy started from 56%. After each epoch, the accuracy of training and validation showed a certain fluctuation.

(**a**)   (**b**)
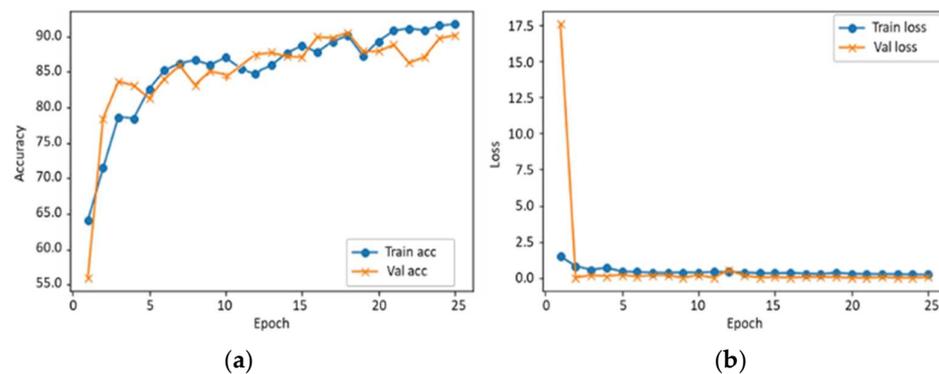
Figure 9. The proposed system's training and validation accuracy and loss of age classification where (**a**) is the accuracy and (**b**) is the loss.

Finally, after the 25th epoch, the level of training accuracy reached 91%, the validation accuracy reached 89%, and the loss of training and validation was nearly 0%, as shown in Figure 9.

The reports generated in Table 6 demonstrate the result of the age estimation using metrics such as the F1-measure, recall, and precision. The experimental evaluation based on accuracy is given in Table 7, where the proposed system obtained an average accuracy of 90%, which surpassed AlexNet, VGG16, and ResNet50 by achieving the higher value of 13%, 9%, and 6%, respectively.

Table 6. Performance of the proposed system over the modified UTKFace dataset in terms of the precision, recall, and f1-score.

Age Classes	Precision	Recall	F1-Measure
Middle age	0.88	0.84	0.86
Overage	0.90	0.97	0.93
Underage	0.92	0.88	0.90

Table 7. Comparison of different DL architectures on the modified UTKFace dataset.

| Technique	Accuracy (%)
AlexNet [51]	77.0
VGG16 [52]	81.0
ResNet50 [65]	84.0
The proposed system	90.0

4.6. Time Complexity Analysis

In this section, we discuss the time complexity analysis of the proposed system and compare it with various versions of deep CNNs as given in Table 8. We show the frame per second (FPS) of four different fused methods such as (AlexNet and MobileNet), (VGG16 and MobileNet), (ResNet50 and MobileNet), and the proposed (MobileNet and custom CNN) over CPU, GPU, and Jetson Nano. The CPU system used for running time analysis was an Intel(R) Core (TM) i3-4010u CPU @ 1.70 GHz with 4 GB RAM. To validate the system performance, we calculated the FPS using Equation (7). We verified that the proposed system was significantly faster than other CNN architectures over CPU, GPU, and Jetson Nano. In Table 8, it can be seen that a lower FPS is associated with VGG16 + MobileNet (i.e., 5.73 FPS on CPU and 33.07 FPS on GPU). The FPS of AlexNet + MobileNet using CPU, GPU, and Jetson Nano was 6.37, 39.87, and 8.01, respectively. The ResNet50 + MobileNet achieved a higher FPS compared with AlexNet + MobileNet and VGG16 + MobileNet. However, our model outperformed the ResNet50 + MobileNet by achieving a higher FPS of 4.98, 12.53, and 5.31, respectively. The better FPS shows that our system is easily deployable over resource-constrained devices.

Table 8. Comparison of different DL architectures on the custom drowsiness detection dataset.

Method Fusion	CPU	GPU	Jetson Nano
AlexNet + MobileNet	6.37	39.87	8.01
VGG16 + MobileNet	5.73	33.07	6.78
ResNet50 + MobileNet	8.90	42.50	13.12
The proposed system	13.88	55.03	18.43

4.7. Qualitative Analysis of the Proposed System

In this section, we demonstrate the visual results of the proposed system, as shown in Figure 10. First, the proposed system was able to accurately detect the face in the entire image. Next, the system could classify the age and state of the driver. In Figure 10a, a set of sample images are shown from the Internet and Figure 10b shows a few sample images taken from a real-time scenario of a camera. In the last row of Figure 10a,b, there was a wrong classification due to the visual similarity of each class. These results show the efficiency and effectiveness of the proposed model and can be deployed for age and various states of driver detection, as evidenced by the quantitative and qualitative analysis.
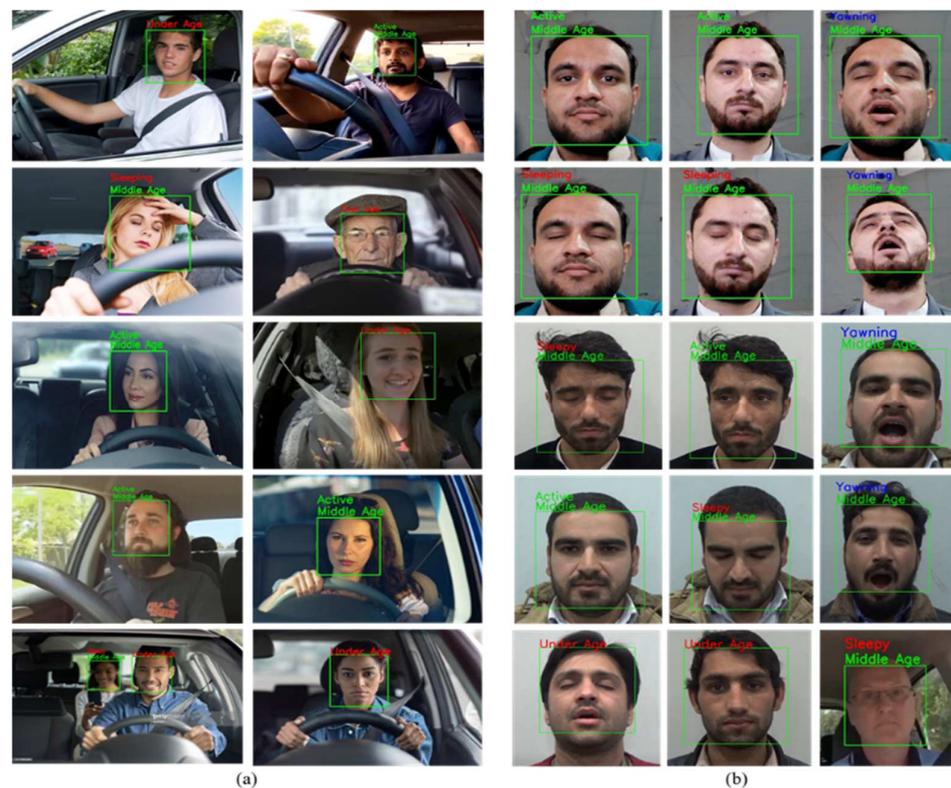
Figure 10. The visual results analysis of the proposed system over images taken from the Internet and from a real scenario. (**a**) Images were taken from the Internet to check the different states and ages of the drivers. (**b**) Images were taken from a real-time scenario for drowsiness detection and age classification.

5. Conclusions and Future Work

In this study, an intelligent fatigue and age detection system (FADS) was proposed for the safety of drivers, helping to prevent plenty of human losses and increasing the intelligence level of vehicles for smart surveillance. The proposed FADS was tested on different platforms for comparison and real-time applicability. A custom CNN model was suitable for low-power hardware, which was deployable over Nvidia's Jetson Nano to achieve portability and a relatively good inference time. After extensive experimentation, we chose two different CNN architectures for driver drowsiness detection and driver age classification based on face feature analysis. The driver's drowsiness detection was achieved using a custom dataset and for age classification, we modified the UTKFace dataset. For experimental evaluation, we evaluated different DL architectures such as AlexNet, VGG16, ResNet50, MobileNetV2, and a 3-layer custom CNN for driver drowsiness detection. The Custom CNN model provided better results and reached an accuracy of 98% for drowsiness detection, whereas MobileNetV2 provided good results in terms of 90% accuracy using the UTKFace dataset. Finally, the results of both models were fused during inference time to ease the deployment for real-time assistance. The developed system is helpful for aged people to prevent them from vehicle accidents. In addition, the results shown in Section 4.6 exhibit the deployability of the proposed method over resource-constrained devices to reduce heavy computation and consumption.

Future work of this study can consider different scenarios such as optimizing a single end-to-end network for its usage in an embedded system to reduce the computational and financial costs without affecting the performance. Next, a federated learning mechanism can be designed for the development of an online model to improve the edge learning capability of FADS. The current dataset can be extended by adding the fatigue levels of people of different ages.

References

1. Lin, C.; Han, G.; Du, J.; Xu, T.; Peng, Y. Adaptive traffic engineering based on active network measurement towards software defined internet of vehicles. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3697–3706. [CrossRef]

2. Peden, M.; Scurfield, R.; Sleet, D.; Mohan, D.; Hyder, A.A.; Jarawan, E.; Mathers, C. *World Report on Road Traffic Injury Prevention*; World Health Organization: Geneva, Switzerland, 2004.

3. World Health Organization. *Association for Safe International Road Travel. Faces behind Igures: Voices of Road Trafic Crash Victims and Their Families*; OMS: Genebra, Switzerland, 2007.

4. National Safety Council. Drivers are Falling Asleep Behind the Wheel. 2020. Available online: https://www.nsc.org/road/safety-topics/fatigued-driver (accessed on 1 January 2023).

5. Vennelle, M.; Engleman, H.M.; Douglas, N.J. Sleepiness and sleep-related accidents in commercial bus drivers. *Sleep Breath.* **2010**, *14*, 39–42. [CrossRef] [PubMed]

6. de Castro, J.R.; Gallo, J.; Loureiro, H. Tiredness and sleepiness in bus drivers and road accidents in Peru: A quantitative study. *Rev. Panam. Salud Publica (Pan Am. J. Public Health)* **2004**, *16*, 11–18.

7. Lenné, M.G.; Jacobs, E.E. Predicting drowsiness-related driving events: A review of recent research methods and future opportunities. *Theor. Issues Ergon. Sci.* **2016**, *17*, 533–553. [CrossRef]

8. Tefft, B.C. Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999–2008. *Accid. Anal. Prev.* **2012**, *45*, 180–186. [CrossRef]

9. Armstrong, K.; Filtness, A.J.; Watling, C.N.; Barraclough, P.; Haworth, N. Efficacy of proxy definitions for identification of fatigue/sleep-related crashes: An Australian evaluation. *Transp. Res. Part F Traffic Psychol. Behav.* **2013**, *21*, 242–252. [CrossRef]

10. Centers for Disease Control and Prevention Drowsy driving-19 states and the District of Columbia, 2009–2010. *MMWR Morb. Mortal. Wkly. Rep.* **2013**, *61*, 1033–1037.

11. Williamson, A.; Friswell, R. The effect of external non-driving factors, payment type and waiting and queuing on fatigue in long distance trucking. *Accid. Anal. Prev.* **2013**, *58*, 26–34. [CrossRef]

12. Hassall, K. Do 'safe rates'actually produce safety outcomes? A decade of experience from Australia. In *HVTT14: International Symposium on Heavy Vehicle Transport Technology*, 14th ed.; HVTT Forum: Rotorua, New Zealand, 2016.

13. Kalra, N. *Challenges and Approaches to Realizing Autonomous Vehicle Safety*; RAND: Santa Monica, CA, USA, 2017.

14. Ballesteros, M.F.; Webb, K.; McClure, R.J. A review of CDC's Web-based Injury Statistics Query and Reporting System (WISQARS™): Planning for the future of injury surveillance. *J. Saf. Res.* **2017**, *61*, 211–215. [CrossRef]

15. Deng, W.; Wu, R. Real-time driver-drowsiness detection system using facial features. *IEEE Access* **2019**, *7*, 118727–118738. [CrossRef]

16. Zhao, L.; Wang, Z.; Wang, X.; Liu, Q. Driver drowsiness detection using facial dynamic fusion information and a DBN. *IET Intell. Transp. Syst.* **2017**, *12*, 127–133. [CrossRef]

17. Massoz, Q.; Langohr, T.; François, C.; Verly, J.G. The ULg multimodality drowsiness database (called DROZY) and examples of use. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–7.

18. Tsaur, W.J.; Yeh, L.Y. DANS: A Secure and Efficient Driver-Abnormal Notification Scheme with I oT Devices Over I o V. *IEEE Syst. J.* **2018**, *13*, 1628–1639. [CrossRef]

19. Xing, Y.; Lv, C.; Zhang, Z.; Wang, H.; Na, X.; Cao, D.; Velenis, E.; Wang, F.-Y. Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition. *IEEE Trans. Comput. Soc. Syst.* **2017**, *5*, 95–108. [CrossRef]

20. Yu, J.; Park, S.; Lee, S.; Jeon, M. Driver drowsiness detection using condition-adaptive representation learning framework. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 4206–4218. [CrossRef]

21. Dua, M.; Singla, R.; Raj, S.; Jangra, A. Deep CNN models-based ensemble approach to driver drowsiness detection. *Neural Comput. Appl.* **2021**, *33*, 3155–3168. [CrossRef]

22. Moujahid, A.; Dornaika, F.; Arganda-Carreras, I.; Reta, J. Efficient and compact face descriptor for driver drowsiness detection. *Expert Syst. Appl.* **2021**, *168*, 114334. [CrossRef]

23. Karuna, Y.; Reddy, G.R. Broadband subspace decomposition of convoluted speech data using polynomial EVD algorithms. *Multimed. Tools Appl.* **2020**, *79*, 5281–5299. [CrossRef]

24. Ji, Y.; Wang, S.; Zhao, Y.; Wei, J.; Lu, Y. Fatigue state detection based on multi-index fusion and state recognition network. *IEEE Access* **2019**, *7*, 64136–64147. [CrossRef]

25. Ghoddoosian, R.; Galib, M.; Athitsos, V. A realistic dataset and baseline temporal model for early drowsiness detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; IEEE: Piscataway, NJ, USA, 2019.

26. Sai, P.-K.; Wang, J.-G.; Teoh, E.-K. Facial age range estimation with extreme learning machines. *Neurocomputing* **2015**, *149*, 364–372. [CrossRef]

27. Lu, J.; Liong, V.E.; Zhou, J. Cost-sensitive local binary feature learning for facial age estimation. *IEEE Trans. Image Process.* **2015**, *24*, 5356–5368. [CrossRef]

28. Huerta, I.; Fernández, C.; Segura, C.; Hernando, J.; Prati, A. A deep analysis on age estimation. *Pattern Recognit. Lett.* **2015**, *68*, 239–249. [CrossRef]

29. Ranjan, R.; Zhou, S.; Chen, J.C.; Kumar, A.; Alavi, A.; Patel, V.M.; Chellappa, R. Unconstrained age estimation with deep convolutional neural networks. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 109–117.

30. Han, H.; Jain, A.K.; Wang, F.; Shan, S.; Chen, X. Heterogeneous face attribute estimation: A deep multi-task learning approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 2597–2609. [CrossRef] [PubMed]

31. Dornaika, F.; Arganda-Carreras, I.; Belver, C. Age estimation in facial images through transfer learning. *Mach. Vis. Appl.* **2019**, *30*, 177–187. [CrossRef]

32. Rothe, R.; Timofte, R.; van Gool, L. Deep expectation of real and apparent age from a single image without facial landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [CrossRef]

33. Shen, W.; Guo, Y.; Wang, Y.; Zhao, K.; Wang, B.; Yuille, A.L. Deep regression forests for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2304–2313.

34. Taheri, S.; Toygar, Ö. On the use of DAG-CNN architecture for age estimation with multi-stage features fusion. *Neurocomputing* **2019**, *329*, 300–310. [CrossRef]

35. Lou, Z.; Alnajar, F.; Alvarez, J.M.; Hu, N.; Gevers, T. Expression-invariant age estimation using structured learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 365–375. [CrossRef]

36. Liu, H.; Lu, J.; Feng, J.; Zhou, J. Group-aware deep feature learning for facial age estimation. *Pattern Recognit.* **2017**, *66*, 82–94. [CrossRef]

37. Ullah, F.U.M.; Obaidat, M.S.; Ullah, A.; Muhammad, K.; Hijji, M.; Baik, S.W. A Comprehensive Review on Vision-based Violence Detection in Surveillance Videos. *ACM Comput. Surv.* **2022**, *55*, 1–44. [CrossRef]

38. Sajjad, M.; Nasir, M.; Ullah, F.U.M.; Muhammad, K.; Sangaiah, A.K.; Baik, S.W. Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services. *Inf. Sci.* **2019**, *479*, 416–431. [CrossRef]

39. Sun, X.; Wu, P.; Hoi, S.C. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **2018**, *299*, 42–50. [CrossRef]

40. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2015; pp. 91–99.

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

42. Ullah, W.; Ullah, A.; Hussain, T.; Khan, Z.A.; Baik, S.W. An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos. *Sensors* **2021**, *21*, 2811. [CrossRef]

43. Yar, H.; Hussain, T.; Khan, Z.A.; Koundal, D.; Lee, M.Y.; Baik, S.W. Vision Sensor-Based Real-Time Fire Detection in Resource-Constrained IoT Environments. *Comput. Intell. Neurosci.* **2021**, *2021*, 5195508. [CrossRef]

44. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

45. Khan, S.U.; Hussain, T.; Ullah, A.; Baik, S.W. Deep-ReID: Deep features and autoencoder assisted image patching strategy for person re-identification in smart cities surveillance. *Multimed. Tools Appl.* **2021**, 1–22. [CrossRef]

46. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

47. Yar, H.; Hussain, T.; Agarwal, M.; Khan, Z.A.; Gupta, S.K.; Baik, S.W. Optimized Dual Fire Attention Network and Medium-Scale Fire Classification Benchmark. *IEEE Trans. Image Process.* **2022**, *31*, 6331–6343. [CrossRef]

48. Redmon, J.; Farhadi, A. Darknet: Open Source Neural Networks in C. 2013. Available online: https://pjreddie.com/darknet/ (accessed on 1 January 2023).

49. Ullah, F.U.M.; Obaidat, M.S.; Muhammad, K.; Ullah, A.; Baik, S.W.; Cuzzolin, F.; Rodrigues, J.J.P.C.; de Albuquerque, V.H.C. An intelligent system for complex violence pattern analysis and detection. *Int. J. Intell. Syst.* **2021**, *37*, 10400–10422. [CrossRef]

50. Ullah, F.U.M.; Muhammad, K.; Haq, I.U.; Khan, N.; Heidari, A.A.; Baik, S.W.; de Albuquerque, V.H.C. AI assisted Edge Vision for Violence Detection in IoT based Industrial Surveillance Networks. *IEEE Trans. Ind. Inform.* **2021**, *18*, 5359–5370. [CrossRef]

51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

52. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

53. Wang, H.; Lu, F.; Tong, X.; Gao, X.; Wang, L.; Liao, Z.J.E.R. A model for detecting safety hazards in key electrical sites based on hybrid attention mechanisms and lightweight Mobilenet. *Energy Rep.* **2021**, *7*, 716–724. [CrossRef]

54. Bi, C.; Wang, J.; Duan, Y.; Fu, B.; Kang, J.-R.; Shi, Y. MobileNet based apple leaf diseases identification. *Mob. Netw. Appl.* **2020**, *27*, 172–180. [CrossRef]

55. Rothe, R.; Timofte, R.; van Gool, L. Dex: Deep expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 10–15.

56. Sajjad, M.; Zahir, S.; Ullah, A.; Akhtar, Z.; Muhammad, K. Human behavior understanding in big multimedia data using CNN based facial expression recognition. *Mob. Netw. Appl.* **2020**, *25*, 1611–1621. [CrossRef]

57. Zhang, T.; Han, G.; Yan, L.; Peng, Y. Low-Complexity Effective Sound Velocity Algorithm for Acoustic Ranging of Small Underwater Mobile Vehicles in Deep-Sea Internet of Underwater Things. *IEEE Internet Things J.* **2022**, *10*, 563–574. [CrossRef]

58. Sun, F.; Zhang, Z.; Zeadally, S.; Han, G.; Tong, S. Edge Computing-Enabled Internet of Vehicles: Towards Federated Learning Empowered Scheduling. *IEEE Trans. Veh. Technol.* **2022**, *71*, 10088–10103. [CrossRef]

59. Rizzo, A.; Burresi, G.; Montefoschi, F.; Caporali, M.; Giorgi, R. Making IoT with UDOO. *IxD&A* **2016**, *30*, 95–112.

60. Nasir, M.; Muhammad, K.; Ullah, A.; Ahmad, J.; Baik, S.W.; Sajjad, M. Enabling automation and edge intelligence over resource constraint IoT devices for smart home. *Neurocomputing* **2022**, *491*, 494–506. [CrossRef]

61. Nayyar, A.; Puri, V.A.; Puri, V. A review of Beaglebone Smart Board's-A Linux/Android powered low cost development platform based on ARM technology. 9th International Conference on Future Generation Communication and Networking (FGCN), Jeju, Republic of Korea, 25–28 November 2015; IEEE: Genebra, Switzerland; pp. 55–63.

62. Yar, H.; Imran, A.S.; Khan, Z.A.; Sajjad, M.; Kastrati, Z. Towards smart home automation using IoT-enabled edge-computing paradigm. *Sensors* **2021**, *21*, 4932. [CrossRef]

63. Jan, H.; Yar, H.; Iqbal, J.; Farman, H.; Khan, Z.; Koubaa, A. Raspberry pi assisted safety system for elderly people: An application of smart home. 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 3–5 November 2020; IEEE: Genebra, Switzerland, 2020; pp. 155–160.

64. Cass, S. Nvidia makes it easy to embed AI: The Jetson nano packs a lot of machine-learning power into DIY projects-[Hands on]. *IEEE Spectr.* **2020**, *57*, 14–16. [CrossRef]

65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.