Deep CNN-based autonomous system for safety measures in logistics transportation

Abdelkarim Rouari, Abdelouahab Moussaoui, Youssef Chahir, Hafiz Tayyab Rauf, Seifedine Kadry

Deep CNN-based autonomous system for safety measures in logistics transportation

Abdelkarim Rouari[1] · Abdelouahab Moussaoui[1] · Youssef Chahir[2] · Hafiz Tayyyab Rauf[3] · Seifedine Kadry[4]

Accepted: 4 June 2021

Abstract
The careless activity of drivers in logistics transportation is a primary reason inside the vehicle during road accidents. This research aims to reduce the number of accidents caused by a failure of the driver in logistics transportation by incorporating an autonomous system. We propose a convolutional neural network -based architecture to recognize and classify different positions which cause road accidents. The proposed system is evaluated with the State Farm Distracted Driver Database, which included examples illustrating ten different driving positions like reaching behind and talking to the passenger, making up, safe driving, talking on the phone, clothing, checking right/left hand, right/left hand, and running the radio. The proposed approach has also been tested against recent algorithms and evaluated. Our model has obtained 98.98% accuracy compared to other types of approaches with different descriptors and classification techniques

Keywords Autonomous system · Logistics transportation · Convolutional neural network · Deep learning · Safety measures

1 Introduction

Hundreds of road accidents occur in the world every day, resulting in many lives, damage to property, and injury (Lu et al. 2019a, c). Consequently, this issue has created significant problems for communities around the world and forced the mobile industry to develop a solution that reduces the occurrence of traffic accidents through innovation and growth and allocates vast resources to put an end to this challenging situation (Hu et al. 2018; Yadav et al. 2019). Over the past decade, the automotive industry and its various products have undergone fascinating progress through the introduction of artificial intelligence in mobile devices, making it easier for cars to drive efficiently and securely and preventing loss of life as a result of a simple mistake of withdrawing attention (Valiente et al. 2019, 2020).

Artificial intelligence (AI) plays an essential role in solving different types of complex real-world problems including autonomous vehicles (Malik et al. 2021), bio-metrics (Younis and Abuhammad 2021), healthcare (Albahli et al. 2019) (Jin et al. 2020; Albahli et al. 2021), industrial applications (Gheisari et al. 2021; Gao et al. 2020a), renewable energy (Gao et al. 2020b, c), and optimization applications (Rauf et al. 2020). Deep learning algorithms, especially CNN (Meraj et al. 2019), have been used in different computer vision-based algorithms (Younis 2021), including transportation patterns of drivers. Driver's behavior is the most important factor leading to serious traffic accidents, like eating, talking on the phone, omissions, using makeup, or anything that distracts the driver while driving (Bichicchi et al. 2020; Xing et al. 2020). The driver's behavior has been

Communicated by Vicente Garcia Diaz.

✉ Hafiz Tayyyab Rauf
h.rauf4@bradford.ac.uk

Abdelkarim Rouari
rouarikarimo@gmail.com

Abdelouahab Moussaoui
moussaoui.abdel@gmail.com

Youssef Chahir
youssef.chahir@unicaen.fr

Seifedine Kadry
skadry@gmail.com

1 Computer Science Department, Faculty of Sciences, Ferhat Abbas University Setif, Setif, Algeria

2 Computer Science Department, University of Caen, Caen, France

3 Department of Computer Science, Faculty of Engineering & Informatics, University of Bradford, Bradford, United Kingdom

4 Faculty of applied computing and technology, Noroff University College, Kristiansand, Norway

addressed to identify and prevent damage in several traditional and modern ways to reduce traffic accidents. However, one of the most important ways that have emerged in the last decade and have been in great demand is deep learning (Shahverdy et al. 2020; Duan et al. 2020).

Deep learning studies these behaviors effectively through approaches that mimic the work and intelligence of the human brain and may outweigh it in some cases, which has shown remarkable results compared to all other methods (Kuutti et al. 2020; Shin et al. 2019). One of the most important deep learning algorithms is CNN which has shown excellent vision computers that outperform all previous methods. The reason is that this method very much mimics the brain's work of learning, recognizing, and classifying things employing a black box, unlike other methods, which makes it the strongest. It receives the image, extracts the essential features automatically, and then works to classify the images according to these features (Huang et al. 2020; Rauf et al. 2019). The thing that controls the accuracy of this method is their architecture ranging from inputs to outputs, and it is the most critical element that determines the success of this method (Khan and Ahmed 2020). The summary table of previous studies that used the Distracted Driver dataset is given in Table 1.

This study's ultimate objective is as follows:

- To propose multilayer CNN-based recolonization system for reducing the road accidents.
- To include different drive poses illustrating ten different driving positions like reaching behind and talking to the passenger, making up, safe driving, talking on the phone, clothing, checking right/left hand, right/left hand, and running the radio.
- To minimize the subsequent deaths, classify, and recognize different places, which constitute significant causes of road accidents.
- To evaluate the proposed recognition system on real-word traffic datasets.

The remaining parts of the manuscript are composed as follows: In Sect. 3, we elaborated on proposed CNN and the different parameters adopted, and the flow architecture of the proposed algorithm. Section 4 talks about the experimentation performed and the parameter setting along with the empirical environment. The results obtained by comparing the proposed architecture with other models are presented in Sect. 5, while Sect. 6 contains the conclusion of the research.

2 Related work

This addressed problem is solved with a great deal of study and experiments. The works can be divided into two groups. The first group involves people who want to study different physical factors outside the vehicle. In the second group, the parameters of the car are studied, like the action of the driver.

Previous studies on visual driving practices primarily concentrate on pupils, facial expressions, or an acceptable combination. Oliver and Pentland (2000) provided a machine-learning system for modeling and recognition of driver's emotions through using graphic models, hidden Markov models (HMMs), and coupled hidden Markov models (CHMMs), with a focus on how the context affects driver performance.

For safety verification, in Ito et al. (2013), two forms of unsupervised neural networks were studied, along with fuzzy adaptive resonance theory (ART) and independent maps for driver body position, where authors addressed categorization of face orientation as well.

Veeraraghavan et al. (2007) introduced an unattended algorithm to classify a probabilistic binary image classification with only two driver actions, i.e., gripping the steering wheel, speaking on a cell phone, and driver activities as a protected class, risky class, or unknown class.

Zhao et al. (2012) developed an efficient feature extraction technique consisting of homomorphic, skin-like segmentation regions for driving postures from a video camera. The proposed Contourlet Transform (CT) feature extractor and the implementation of the Random Forest (RE) grading classifier is used to classify four classes of driving operations: steering wheel grip, shift gear operation, mobile phone feed, and chat.

Zhao et al. (2011a) introduce a new method for the extraction of vehicle drivers' fatigue signals, consisting of the facial recognition algorithm Viola–Jones and Gabor wave line transformation. Using the multilayer perceptron (MLP) classifier, holdout experiments are created with features derived from the fatigue expression dataset generated in Southeast University compared with the naive Bayes classifier, subspace classifier, and k-nearest neighbor (kNN).

Mechanized infringement authorization approaches for violating the safety belt, cell phone infringement, and inhabitance infringement identification assignments have been proposed later. Smoking while driving is yet another common form of violation restricted in many nations. The street-side officials conduct physical smoking tests daily. The study (Artan et al. 2019) suggests a robotic approach to perform the location of driver smoking using near-infrared (NIR) recognition camera images. Cigarette tips arrive at 800–900 C during the puff creating a problem area on the NIR images. The proposed strategy aims to identify particular problem areas around the head district of the drivers. First, they use deep learning-based item location to constrain the front windshield and driver 's head area consecutively. Next, they play out a double window (neighborhood) peculiarity locator on the limited district to determine the white problem area, and

then the driver conducts smoking. They have gathered 1472 original NIR images from the world to assess the presence of the solution proposed. The proposed methodology obtained a general accuracy rate of 84 percent and an affectability rate of 70 percent on the test sample.

A new vision-based system for conductor foot behavior analysis was proposed by Tran et al. (2012) and developed using optical flow-based foot tracking and a hidden Markov model (HMM) technique to describe temporal foot behavior. Liu et al. (2002) identified a vision device that monitored the driver's face and used the yaw orientation angles to estimate his face position during driving conditions. Toma et al. (2012) developed an efficient approach for inexperienced drive by using a finite state machine (FSM) and a rule-based system (RBS), using inputs from sensor data fusion. By evaluating the sequence of postures, we can determine whether the maneuvering of the inexperienced drivers is done correctly.

An abnormal driving factor is a crucial cause of actual car accidents which threaten human life and open land all-inclusive. They discuss using a profound learning approach to deal with the ultimately perceived driving behavior (e.g., ordinary driving, hand driving off the wheel, calling, playing cell phone, smoking, and talking to travelers) in a single picture. The allocation of acknowledgment of driving conduct can be viewed as a multi-class order problem. In the study, the author resolves this major issue (Hu et al. 2018) from two perspectives: (1) Using multi-stream CNN to extricate multiscale that includes separating images with open fields of different portion sizes and (2) researching different combination techniques to join multiscale data and producing the ultimate drug choice. The adequacy of their proposed approach is supported by comprehensive analyses conducted on our dataset of self-made re-enacted driving behavior , much like a data collection of actual driving behavior. The test results show that the proposed CNN-based multi-stream solution accomplishes the critical design improvements instead of the best in class.

Kato et al. (2004) established an active capture device for the identification of facial directions of the driver, such as the left, front, and back.

You et al. (2017) choose eight parameters that represent a driving fatigue identification model with a support vector machine algorithm, reflecting the vehicle movement state and the driver's physiological and psychological status.

Yan et al. (2015) proposed a modern vision-based recognition method for driving posture. A side-mounted camera with a view of the left profile of a driver prepared the driving position dataset. After preprocessing, they extracted eight segments of action groups of driving activities, including ordinary drive, mobile phone service, eating, and smoking and applied the directed gradient pyramid histogram (PHOG) for more discriminating characteristics. Four widely used multilayer perceptrons (MLPs) and support vector machines (SVMs) are assessed at each step, including random forest (RF) and k-nearest neighbor (KNN).

In the study (Zhao et al. 2011b), features are derived from a driving posture dataset consisting of gripping the steering wheel, working the shift lever, eating a cake, and talking on a mobile phone, during which vector machines (SVMs) were introduced with five separate kernels.

Diverted driver behavior is the primary driver of street-car accidents, jeopardizing human life protection and open property. Considering the intuition that prompts (like cigarette holding hand) to show what the driver does, a driver operation recognition model is introduced Lu et al. (2019c), which is called Faster R-CNN (DD-RCNN) deformable and expanded. Their methodology uses definite articles to discover movement to arrange driver activities that display incredible intra-class contrasts and class-likeness. The deformable and extended lingering square is built to isolate highlights of overt ROIs operation that are small in size and sporadically fit as a fiddle (for example, cigarettes and cell phones). Consideration modules are added for the reweight highlights in the channel and spatial measurements in the modified ResNet. The District Proposal Advancement Organize is implemented to decrease the quantity of ROIs joining R-CNN and increase model profitability. Furthermore, the deformable one is replaced by the RoI pooling portion, and the rearranged R-CNN without relapse layer is prepared as the final classifier. Results show DD-RCNN on the Kaggle-driving dataset, and the self-manufactured dataset shows the best performance in class.

Dangerous driving activity triggers numerous car accidents, causing actual losses and property misfortunes. The hand with a cigarette can be uncovered from the pieces of information in the image. The new approaches to identifying actions are not suitable for interpreting driving activity with only local contrasts. Throughout the (Lu et al. 2019b) investigation, the author perceives driving actions by defining particular parts of the action and suggests the solution Dilated Light-Head R-CNN (DL-RCNN), which uses broader convolution to guarantee apparent subtleties for the target picture. The specialized curiosity includes: a position-delicate RoI structure to improve the view of small posts and tri-focus nonsense to introduce proximity between intra-class highlights and comparison of highlights in separate classes. We also endorse two procedures, such as site complex model mining and correspondingly modifying the global boundaries. The test results on the comprehensive collection of Kaggle-driving information and the self-manufactured collection of information suggest that DL-RCNN performs cutting-edge perception of driving behavior.

Table 1 Summary table of previous studies used the distracted driver dataset

	References	Methods	Comparison	Results	Evaluation metrics
1	Moslemi et al. (2019)	3D-CNN	VGG, Alex Net, Inception Net	94.40% Accuracy	Accuracy and confusion matrix
2	Baheti et al. (2018)	M-VGG	Alex Net and Inception Net	95.54% Accuracy	Correct predictions, Incorrect predictions, Accuracy (%)
3	Chawan et al. (2018)	Inception V3	Inception V3, Vgg16 and 19	73.00% Accuracy	Loss and Accuracy
4	Chawan et al. (2018)	Vgg 19	Inception V3 and Vgg16	77.00% Accuracy	Loss and Accuracy
5	Tran et al. (2018)	ResNet 152	Vgg 16, AlexNet, and SVM+Handcrafted features	85% Accuracy	Accuracy and confusion matrix
6	Tran et al. (2018)	Vgg 16	ResNet 152, AlexNet, and SVM+ Handcrafted features	82.50% Accuracy	Accuracy and confusion matrix
7	Tran et al. (2018)	AlexNet	Vgg 16, ResNet 152, and SVM+Hand	72.60% Accuracy	Accuracy and confusion matrix
8	Tran et al. (2018)	SVM+Handcrafted features	Vgg 16, AlexNet, and ResNet 152	27.70% Accuracy	Accuracy and confusion matrix
9	Huang et al. (2020)	HCF	Inception V3, Vgg16 and 19	96.74% Accuracy	Accuracy and confusion matrix
10	Dhakate and Dash (2020)	DD-ST-M1	VGG, Alex Net and Inception Net	73.01% Accuracy	Loss, accuracy and confusion matrix
11	Dhakate and Dash (2020)	DD-ST-M2	VGG, Alex Net, Inception Net and ResNet	97.00% Accuracy	Loss, accuracy and confusion matrix

Table 2 Statistics of dataset obtained

Label	Characteristics	Quantity
C0	Safe driving	2489
C1	Texting-right	2267
C2	Talking-right	2317
C3	Texting-left	2346
C4	Talking-left	2326
C5	Operating the radio	2312
C6	Drinking	2325
C7	Reaching behind	2002
C8	Hair and makeup	1911
C9	Talking to passenger	2129
Total	–	22,424

3 Materials and methods

3.1 Driver distraction classification

Driver distraction classification offers a problem identifying if the driver is being distracted while driving using the novel deep learning approach. In this approach, a CNN is used for classifier driver stance during driving. The CNN is trained on images of drivers while driving. The stances of the driver to be predicted are:

- Safe driving
- Texting-right
- Talking on the phone-right
- Texting-left
- Talking on the phone-left
- Operating the radio
- Drinking
- Reaching behind
- Hair and makeup
- Talking to passenger

3.2 Dataset

Below are the statistics of the dataset used for this research (Table 2).

The dataset used included 22424 portraits belonging to all ten groups of people either driving safely or performing one of nine types of interrupted behaviors (texting, eating, talking on the phone, making up, reaching behind, and others). The dataset was divided into 70, 30-ratio training, and validation dataset. The training dataset includes 15702 images, and 6722 images are included in the validation dataset. Both images were resized to 120 * 160 pixels. Images have been rescaled, so their values are between 0 and 1. The training images come with clear labeling, and the challenge is to make the best classifications possible (Figs. 1, 2, 3, 4, 5).

3.3 CNN architecture

CNN or convolutional neural networks have widely used in image-related tasks. CNN is based on the human eye. The main advantage of CNN is that it automatically extracts features from an image. Proposed CNN architecture contains three convolutional, three MaxPooling, two Dense layers, and one output layer. The convolutional layers of the proposed algorithm have 12, 16, 20 kernels and kernel sizes of (5, 5), (5, 5), and (4, 4), respectively. MaxPooling layers have kernel sizes of (2, 2). Our dense layers have 512 and 128 neurons, respectively. All hidden layers have activation of ReLU, and output layers have activation of Softmax.

3.3.1 Convolutional layer

Convolution is basically as kernel or filter which moves over all the image and extract feature map which contains features of an image (Khan and Yong 2017). It helps in detecting features of an image. Kernel is a matrix that moves over the image, and each value of the matrix is multiplied and then summed, and we get the resultant feature map that contains our features. Some famous kernels are edge detection, blurring, and sharpening the image (Hu et al. 2017).

3.3.2 MaxPooling layer

MaxPooling layer helps detect the best or most prominent features of the image matrix (Boureau et al. 2010). It is done basically to reduce the number of computations and find the most prominent and ignore the less prominent feature. It reduces the feature matrices in size and increases the number of essential features in them (Preprint 2014).

3.3.3 Dense layers

Dense layers or fully connected layers are a simple neural network at the end of every CNN that computes the output based on the convolved and max-pooled features, flattened just before the fully connected neural network.

3.3.4 Data augmentation

Data augmentation introduces random zoom, flip, and shear in an image to prevent over-fitting. Data augmentation is used in training CNN so that it does not over-fit the data.

Yan LeCunn introduced CNN based on feed-forward neural networks, which indicates that the information is fed forwarded through the neural network, and after comparing the output of the image with the actual output, the error is

(a) Sample 1 C0. (b) Sample 2 C1. (c) Sample 3 C2.

(d) Sample 4 C3. (e) Sample 5 C4. (f) Sample 6 C5.

(g) Sample 7 C6. (h) Sample 8 C7. (i) Sample 9 C8.

(j) Sample 10 C9.

Fig. 1 Sample of labeled images differentiating different characteristics by each class

back-propagated. CNN uses a backpropagation algorithm for error correction and weight alteration. Backpropagation is an algorithm that determines which neuron is contributing to the overall error. Based on each Neuron's error, the weights are updated accordingly, and the entire process is carried on until the CNN converges.

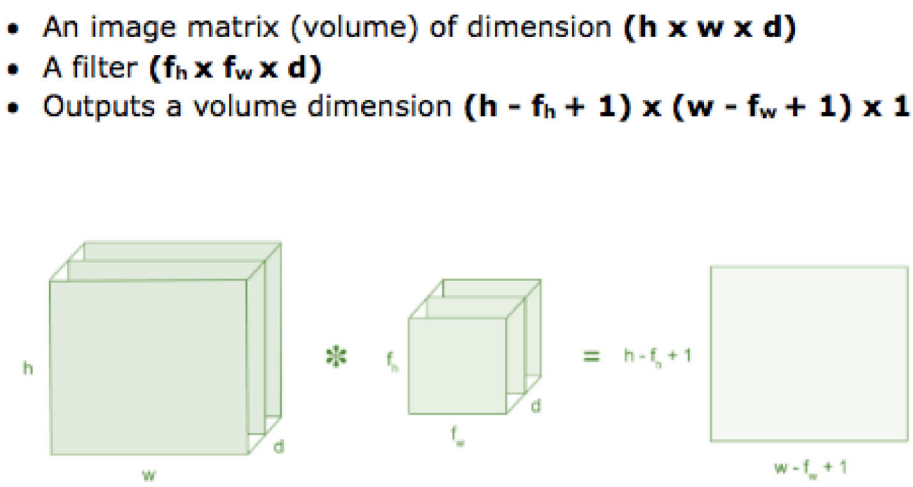

Fig. 2 Image matrix multiplies kernel or filter matrix

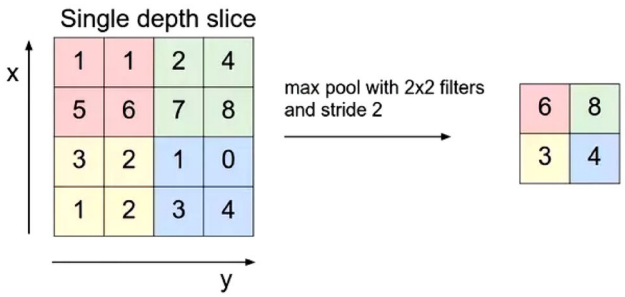

Fig. 3 Single depth slice of max pooling

4 Experimentation

Experiments were conducted with two CNN architectures. One was a larger one containing higher 32, 64, and 64 filters in each convolutional layer, respectively. It also contained a dropout. Dropout is a technique of shutting down random neurons such that noise is introduced in training and the neural network does not over-fit the data. The other CNN was a simple and smaller one described in the previous pages, and it is proposed architecture. The reason why we chose this architecture over the other one is that it was performing better. The other architecture was under-fitting our data. Training loss was higher than validation loss, and training accuracy was lower than validation accuracy. The reason is that simply the problem does not require a more extensive network. Experimentations were conducted with the hyperparameters as Learning_rate = 0.001, Loss_function = Categorical_Crossentropy and Optimizer = Adam.

The trial after-effects of the proposed model and the other state-of-the-art models are as per the following, under differing parameters, i.e., epochs and verbose per epochs, training loss, training accuracy, testing loss, and testing accuracy of proposed CNN, in Table 3.

5 Results

It is observed that the training accuracy (referred to Fig. 6) after eight epochs starts converging. The neural network is not stopped early because the CNN still needs to attain the maximum accuracy it could, and after 20 epoch, it reached 98.98%.

The training loss (Fig. 7) started converging after eight epochs, and training loss was almost reduced to around 0.0316 in the 20th. After that, the loss might not have reduced much, and CNN converged.

The validation accuracy (referred to Fig. 8) showed similar behavior to the training accuracy, which shows that the proposed CNN was not over-fitting the data nor was under-fitting the data. In initial epochs, the validation accuracy was more significant than our training accuracy; the reason is that the proposed CNN was under-fitting which is expected as per bias–variance curve. After a few more epochs, the CNN started showing best-fit results.

Moreover, the validation loss (referred to Fig. 9) showed similar results to the training loss, determining the CNN is not over-fitted, and the loss is also shallow (Table 4).

The confusion matrix shows that our model was not biased toward any class. Furthermore, we have brought the comparative analysis of the proposed approach with recent state-of-the-art algorithms based on the same parameter setting, the same number of classes with a standard experimental environment. Performance comparison of the proposed approach with state-of-the-art variants on an exact number of classes is presented in Table 6. The bar chart accuracy comparison of proposed vs. others is illustrated in Fig. 10.

5.1 Proposed Algorithm

The proposed deep learning approach has been tested on the State Farm Driver distraction dataset, demonstrating the

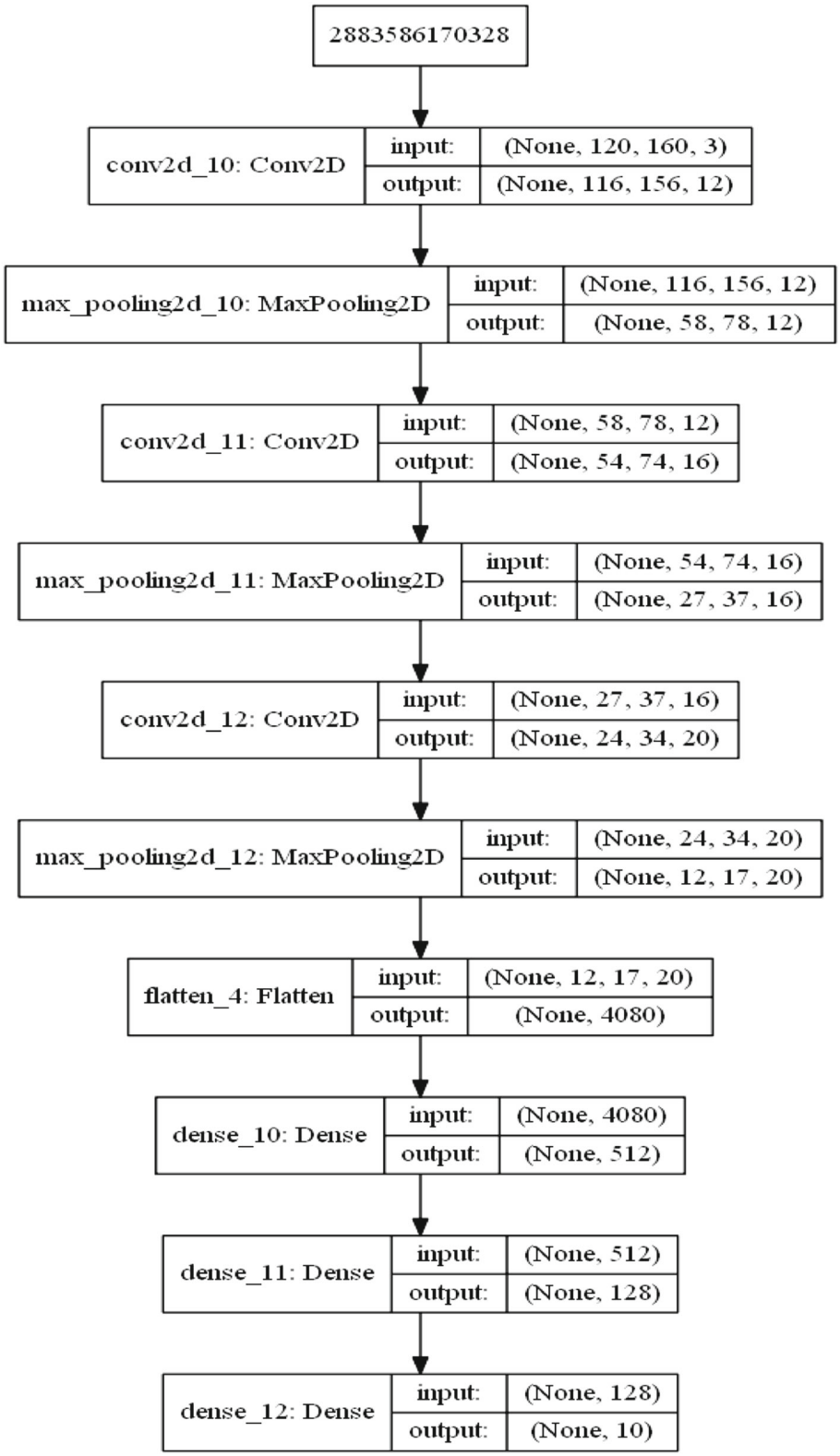

Fig. 4 Model architecture with input–output parameters

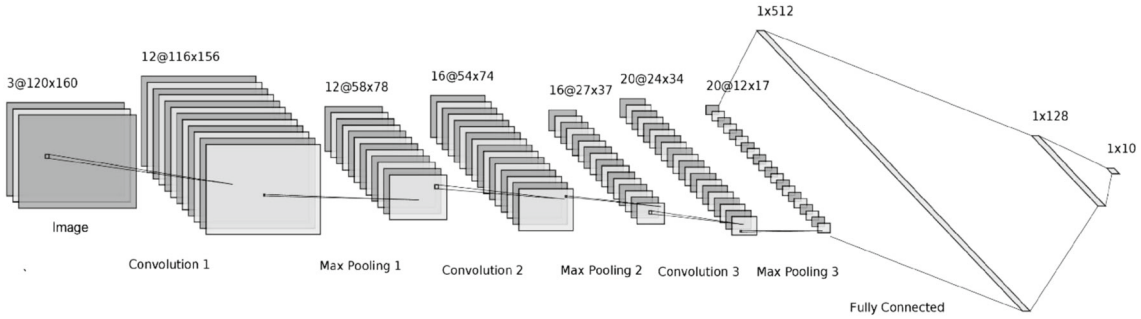

Fig. 5 Proposed CNN architecture

Table 3 Training results obtained on different parameters

Epoch #	Training loss	Training accuracy	Validation loss	Validation accuracy
1	1.2986	0.5397	0.6280	0.7871
2	0.3822	0.8768	0.2605	0.9177
3	0.2301	0.9288	0.1886	0.9424
4	0.1568	0.9513	0.1448	0.9578
5	0.1230	0.9616	0.1591	0.9511
6	0.1060	0.9671	0.1251	0.9655
7	0.0899	0.9726	0.1099	0.9676
8	0.0942	0.9708	0.1065	0.9680
9	0.0712	0.9783	0.1011	0.9714
10	0.0682	0.9797	0.1316	0.9612
11	0.0615	0.9806	0.0898	0.9774
12	0.0623	0.9796	0.1147	0.9664
13	0.0590	0.9798	0.0959	0.9749
14	0.0542	0.9828	0.0863	0.9762
15	0.0477	0.9860	0.0786	0.9793
16	0.0514	0.9831	0.0987	0.9737
17	0.0500	0.9855	0.0793	0.9784
18	0.0481	0.9846	0.0938	0.9750
19	0.0449	0.9857	0.0796	0.9807
20	0.0316	0.9898	0.0786	0.9805

effectiveness in the generalization performance in natural driving environments. The CNN model shows better results in the contrast experiment than a handcrafted approach and ensemble learning system. The end-to-end learning-based model can learn from raw images using temporal signs and discrimination ratio. This section examines and contrasts the findings with some previous models.

Among the six pre-prepared (referred to Table 5) models with the proposed CNN architectures, i.e., Alex Net, Inception V3, Best-Rf, R BF- SVM, Majority voting ensemble, and GA-weighted ensemble, the best-Rf accomplishes the highest accuracy yet additionally the most significant loss. The proposed CNN has the most reduced accuracy and second-lowest loss. Inception V3 accomplishes the third-lowest loss, yet its accuracy is lower than GA-weighted ensemble and majority voting ensemble. The proposed CNN architecture accomplishes an accuracy of 98.96% on the validation set and 98.98% on the training set for distracted behavior detection.

For GA-weighted ensemble and majority voting ensemble, these two accuracy rates are 95.98% and 95.77%, separately, and for Alex Net, they are 93.66% with 0.39% loss, individually. In this way, the over six pre-trained models can perceive distracted driving behaviors in the State Farm dataset. Notwithstanding, these pre-trained models likewise all have enormous losses; this way, they are unequipped for recognizing data instances of occupied CNN drivers without this particular dataset. The testing and training curves for the proposed CNN have appeared in figures. It tends to be seen that the proposed CNN improves the testing accuracy and loss and improves the validation accuracy and loss.

Similarly, referred to Table 6, we compare the proposed model with other recent CNN models having a standard

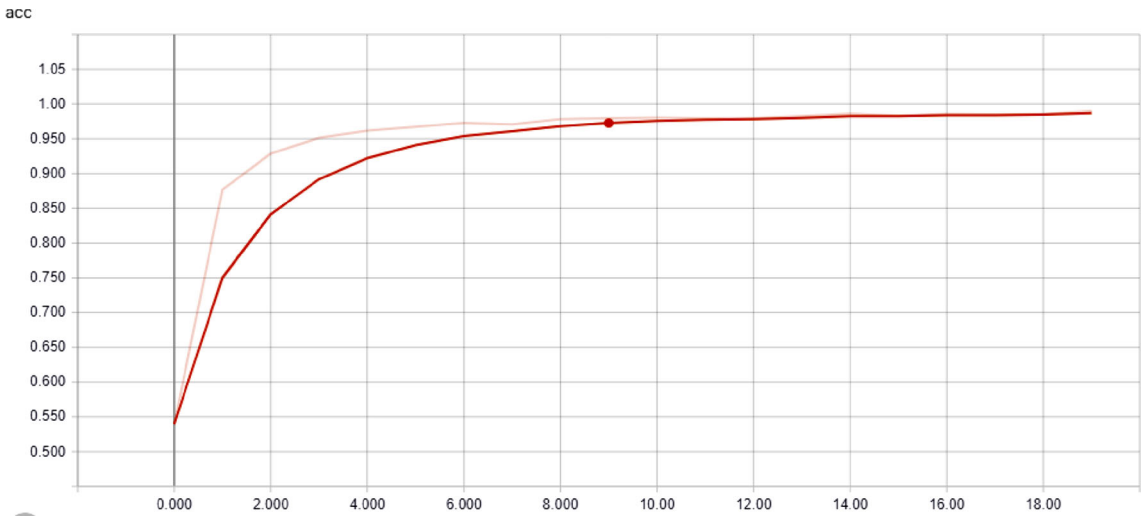

Fig. 6 Training accuracy obtained on different parameters

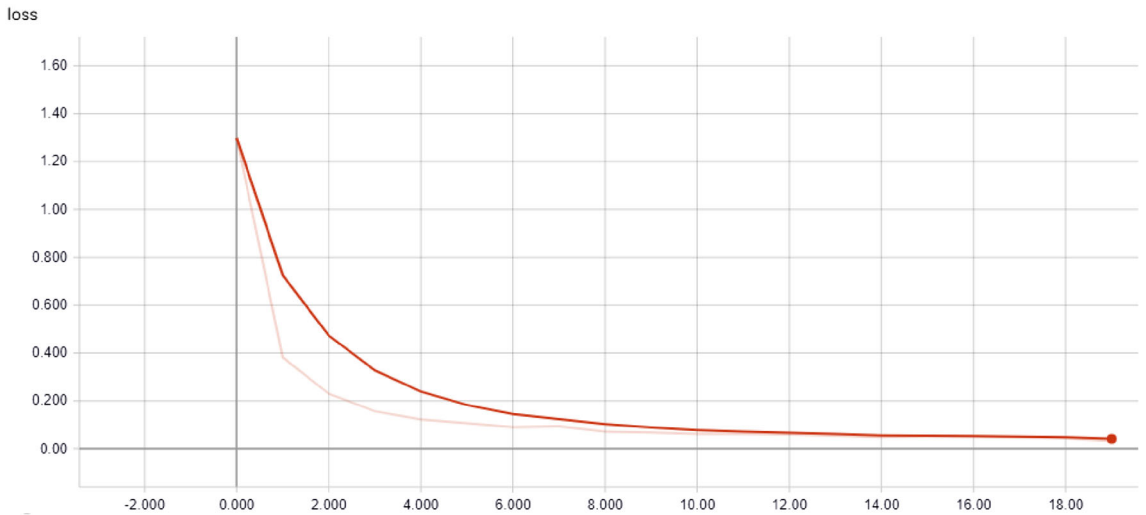

Fig. 7 Training loss obtained on different parameters

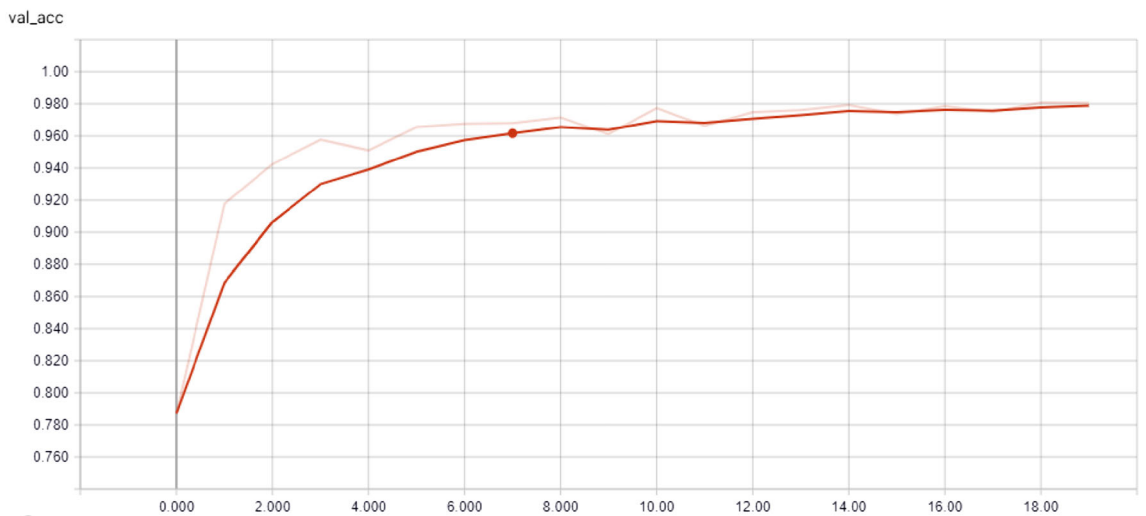

Fig. 8 Validation accuracy obtained on different parameters

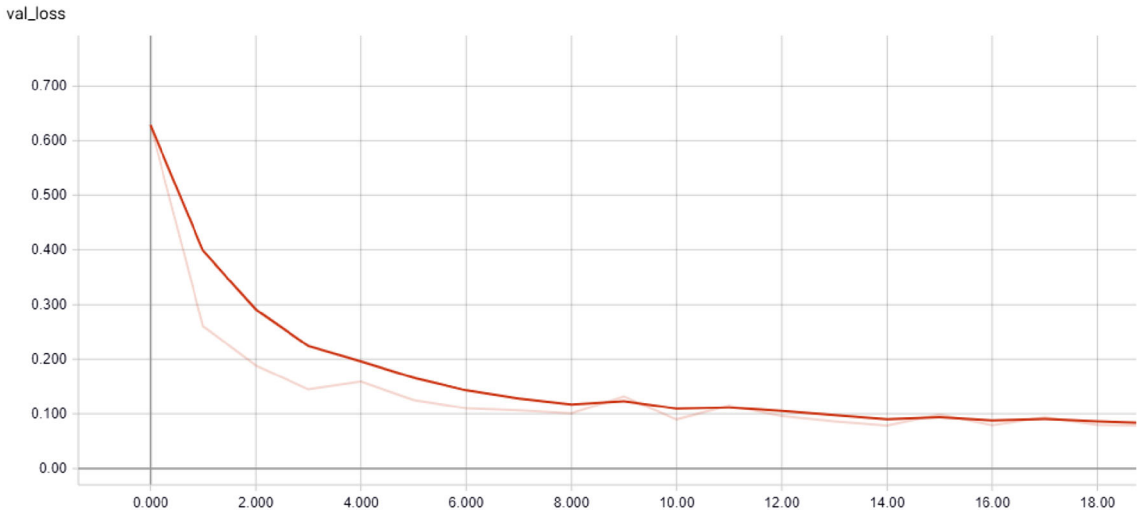

Fig. 9 Validation loss obtained on different parameters

Table 4 Confusion matrix of validation data obtained using CNN

Class	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
C0	717	1	0	2	5	15	1	1	2	2
C1	2	675	0	1	0	0	2	0	0	0
C2	0	0	691	0	1	0	0	1	1	1
C3	2	0	0	695	3	0	0	0	1	2
C4	0	0	1	0	694	1	0	0	1	0
C5	2	0	0	0	1	688	0	0	2	0
C6	0	0	8	0	1	0	680	0	6	2
C7	0	0	0	0	0	0	0	594	5	1
C8	3	0	1	0	4	1	2	3	555	4
C9	5	1	0	1	1	2	0	1	12	615

Table 5 Comparison between some methods using accuracy and loss metrics

Method	Training accuracy %	Testing accuracy %	Loss %
Our CNN	**98.98**	**98.05**	0.3
AlexNet	–	93.65	0.3909
InceptionV3	–	94.57	0.2937
Best-RF	88.47	80.6	0.4
R BF- SVM	92.81	92.45	–
Majority voting ensemble	–	95.77	0.1661
GA-weighted ensemble	–	95.98	**0.1575**

Improved results are shown in bold

testing benchmark. We can observe that the combination of handcrafted features and the external classifier did not work in the literature and obtained a poor accuracy score of 27%.

Table 6 Performance comparison of proposed approach with state-of-the-art variants on same number of classes

Network	Accuracy (%)
3D-CNN (Moslemi et al. 2019)	94.40
M-VGG (Baheti et al. 2018)	95.54
Inception V3 (Chawan et al. 2018)	73.00
Vgg 19 (Chawan et al. 2018)	77
ResNet 152 (Tran et al. 2018)	85
Vgg 16 (Tran et al. 2018)	82.50
AlexNet (Tran et al. 2018)	72.60
SVM+Handcrafted features (Tran et al. 2018)	27.70
HCF (Huang et al. 2020)	96.74
DD-ST-M1 (Dhakate and Dash 2020)	73.01
DD-ST-M2 (Dhakate and Dash 2020)	97.00
Proposed method	98.05

On the other hand, 3D CNN got a better accuracy score than 94.40% instead of the modified Inception V3 with 73% accuracy. The proposed approach obtained 98.05% accuracy and stood first in the comparison list, while the second modified variant of DD-ST reached 97% and got second place in the evaluation comparison in terms of superior performance.

6 Conclusion

This article discusses the minimization of traffic accidents using an intelligent system based on the CNN method by distinguishing between dangerous and safe driving situations. The proposed CNN architecture implicitly extracts the fea-

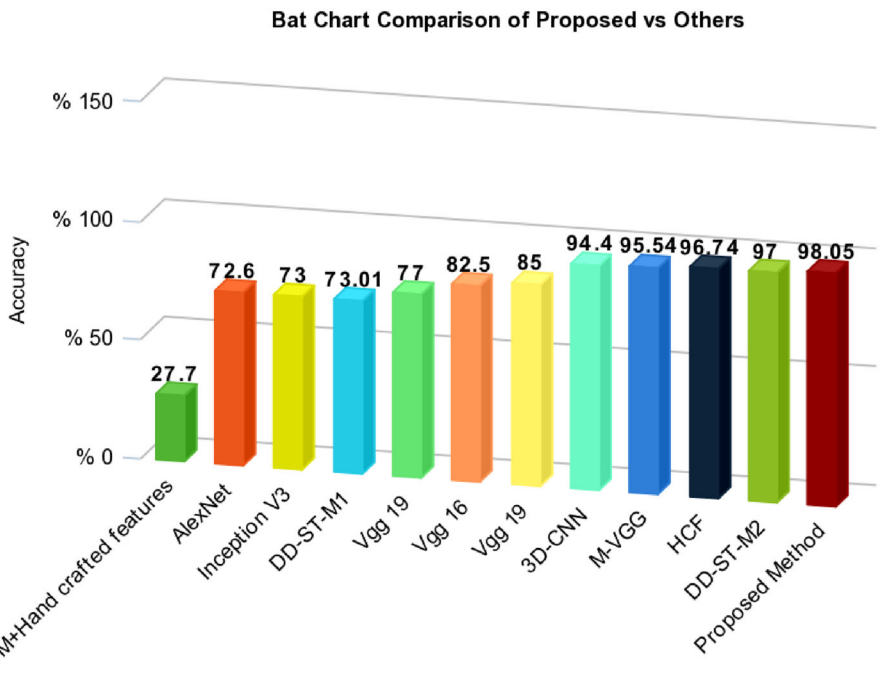

Fig. 10 Bar chart accuracy comparison of proposed versus others

tures of each network and is more robust than the existing methods. The proposed system is evaluated with the State Farm Distracted Driver Database, which included examples illustrating ten different driving positions like reaching behind and talking to the passenger, making up, safe driving, talking on the phone, hair, checking right/left hand, right/left hand and running the radio. Compared to other common approaches with different image descriptors and classification methods, we defined the selection of proposed CNN architecture parameters and the different layers and presented the results for the different epoch configurations. The proposed model has achieved the best performance with an accuracy of 98.98%.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by [AR, AM, YC, HTR, and SK]. The first draft of the manuscript was written by [AR], and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funds, grants, or other support were received.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Albahli S, Rauf HT, Arif M, Nafis MT, Algosaibi A (2019) Identification of thoracic diseases by exploiting deep neural networks. Neural Networks 5:6

Albahli S, Rauf HT, Algosaibi A, Balas VE (2021) Ai-driven deep cnn approach for multi-label pathology classification using chest x-rays. PeerJ Computer Science 7:e495

Artan Y, Balcı B, Elihoş A, Alkan B (2019) Vision based driver smoking behavior detection using surveillance camera images. In: Lecture notes in computer science, Springer International Publishing, pp 468–476. https://doi.org/10.1007/978-3-030-30645-8_43

Baheti B, Gajre S, Talbar S (2018) Detection of distracted driver using convolutional neural network. In: 2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), IEEE. https://doi.org/10.1109/cvprw.2018.00150

Bichicchi A, Belaroussi R, Simone A, Vignali V, Lantieri C, Li X (2020) Analysis of road-user interaction by extraction of driver behavior features using deep learning. IEEE Access 8:19638–19645. doi: 10.1109/access.2020.2965940

Boureau YL, Bach F, LeCun Y, Ponce J (2010) Learning mid-level features for recognition. In: 2010 IEEE computer society conference on computer vision and pattern recognition, IEEE. https://doi.org/10.1109/cvpr.2010.5539963

Chawan PM, Satardekar S, Shah D, Badugu R, Pawar A (2018) Distracted driver detection and classification. International Journal of Engineering Research and Applications 4:7. https://doi.org/10.1109/cloudfront.net/56914644

Dhakate KR, Dash R (2020) Distracted driver detection using stacking ensemble. In: 2020 IEEE international students conference on electrical, electronics and computer science (SCEECS), IEEE. https://doi.org/10.1109/sceecs48394.2020.184

Duan J, Li SE, Guan Y, Sun Q, Cheng B (2020) Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data. IET Intelligent Transport Systems 14(5):297–305. https://doi.org/10.1049/iet-its.2019.0317

Gao J, Wang H, Shen H (2020a) Machine learning based workload prediction in cloud computing. In: 2020 29th international conference

on computer communications and networks (ICCCN), IEEE, pp 1–9

Gao J, Wang H, Shen H (2020b) Smartly handling renewable energy instability in supporting a cloud datacenter. In: 2020 IEEE international parallel and distributed processing symposium (IPDPS), IEEE, pp 769–778

Gao J, Wang H, Shen H (2020c) Task failure prediction in cloud data centers using deep learning. IEEE Trans Serv Comput

Gheisari M, Najafabadi HE, Alzubi JA, Gao J, Wang G, Abbasi AA, Castiglione A (2021) Obpp: an ontology-based framework for privacy-preserving in iot-based smart city. Future Gen Comput Syst

Hu H, Liu B, Zhang P (2017) Several models and applications for deep learning. In: 2017 3rd IEEE international conference on computer and communications (ICCC), IEEE. https://doi.org/10.1109/compcomm.2017.8322601

Hu Y, Lu M, Lu X (2018) Driving behaviour recognition from still images by using multi-stream fusion CNN. Machine Vision and Applications 30(5):851–865. https://doi.org/10.1007/s00138-018-0994-z

Huang C, Wang X, Cao J, Wang S, Zhang Y (2020) HCF: A hybrid CNN framework for behavior detection of distracted drivers. IEEE Access 8:109335–109349. https://doi.org/10.1109/access.2020.3001159

Ito M, Fukumi M, Sato K (2013) Analysis of safety verification behavior and classification of driver's head posture. In: 2013 IEEE international conference on mechatronics and automation, IEEE, pp 884–889. https://doi.org/10.1109/icma.2013.6618032

Jin B, Cruz L, Goncalves N (2020) Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. IEEE Access 8:123649–123661

Kato T, Fujii T, Tanimoto M (2004) Detection of driver's posture in the car by using far infrared camera. In: IEEE intelligent vehicles symposium, 2004, IEEE, pp 339–344. https://doi.org/10.1109/ivs.2004.1336406

Khan MN, Ahmed MM (2020) Trajectory-level fog detection based on in-vehicle video camera with TensorFlow deep learning utilizing SHRP2 naturalistic driving data. Accident Analysis & Prevention 142:105521. https://doi.org/10.1016/j.aap.2020.105521

Khan S, Yong SP (2017) A deep learning architecture for classifying medical images of anatomy object. In: Pacific signal and information processing association annual summit and conference (APSIPA ASC), IEEE. https://doi.org/10.1109/apsipa.2017.8282299

Kuutti S, Bowden R, Jin Y, Barber P, Fallah S (2020) A survey of deep learning applications to autonomous vehicle control. IEEE Trans Intell Transp Syst, doi: 10.1109/tits.2019.2962338

Liu X, Zhu Y, Fujimura K (2002) Real-time pose classification for driver monitoring. In: Proceedings. The IEEE 5th international conference on intelligent transportation systems, IEEE. https://doi.org/10.1109/itsc.2002.1041209

Lu M, Hu Y, Lu X (2019a) Dilated light-head r-CNN using tri-center loss for driving behavior recognition. Image and Vision Computing 90:103800. https://doi.org/10.1016/j.imavis.2019.08.004

Lu M, Hu Y, Lu X (2019b) Dilated light-head r-CNN using tri-center loss for driving behavior recognition. Image and Vision Computing 90:103800. https://doi.org/10.1016/j.imavis.2019.08.004

Lu M, Hu Y, Lu X (2019c) Driver action recognition using deformable and dilated faster r-CNN with optimized region proposals. Applied Intelligence 50(4):1100–1111. https://doi.org/10.1007/s10489-019-01603-4

Malik S, Khattak HA, Ameer Z, Shoaib U, Rauf HT, Song H (2021) Proactive scheduling and resource management for connected autonomous vehicles: a data science perspective. IEEE Sens J

Meraj T, Rauf HT, Zahoor S, Hassan A, Lali MI, Ali L, Bukhari SAC, Shoaib U (2019) Lung nodules detection using semantic segmentation and classification with optimal features. Neural Comput Appl:1–14

Moslemi N, Azmi R, Soryani M (2019) Driver distraction recognition using 3d convolutional neural networks. In: 2019 4th international conference on pattern recognition and image analysis (IPRIA), IEEE. https://doi.org/10.1109/pria.2019.8786012

Oliver N, Pentland A (2000) Graphical models for driver behavior recognition in a SmartCar. In: Proceedings of the IEEE intelligent vehicles symposium 2000 (Cat. No.00TH8511), IEEE, pp 7–12. https://doi.org/10.1109/ivs.2000.898310

Preprint repository arXiv achieves milestone million uploads (2014) Physics Today https://doi.org/10.1063/pt.5.028530

Rauf HT, Lali MIU, Zahoor S, Shah SZH, Rehman AU, Bukhari SAC (2019) Visual features based automated identification of fish species using deep convolutional neural networks. Computers and Electronics in Agriculture 167:105075. https://doi.org/10.1016/j.compag.2019.105075

Rauf HT, Malik S, Shoaib U, Irfan MN, Lali MI (2020) Adaptive inertia weight bat algorithm with sugeno-function fuzzy search. Applied Soft Computing 90:106159

Shahverdy M, Fathy M, Berangi R, Sabokrou M (2020) Driver behavior detection and classification using deep convolutional neural networks. Expert Systems with Applications 149:113240. https://doi.org/10.1016/j.eswa.2020.113240

Shin D, Geun Kim H, MoonPark K, Yi K (2019) Development of deep learning based human-centered threat assessment for application to automated driving vehicle. Applied Sciences 10(1):253. https://doi.org/10.3390/app10010253

Toma MI, Rothkrantz LJ, Antonya C (2012) Car driver skills assessment based on driving postures recognition. In: 2012 IEEE 3rd international conference on cognitive infocommunications (CogInfoCom), IEEE. https://doi.org/10.1109/coginfocom.2012.6422021

Tran C, Doshi A, Trivedi MM (2012) Modeling and prediction of driver behavior by foot gesture analysis. Computer Vision and Image Understanding 116(3):435–445. https://doi.org/10.1016/j.cviu.2011.09.008

Tran D, Do HM, Sheng W, Bai H, Chowdhary G (2018) Real-time detection of distracted driving based on deep learning. IET Intelligent Transport Systems 12(10):1210–1219. https://doi.org/10.1049/iet-its.2018.5172

Valiente R, Zaman M, Ozer S, Fallah YP (2019) Controlling steering angle for cooperative self-driving vehicles utilizing CNN and LSTM-based deep networks. In: IEEE intelligent vehicles symposium (IV), IEEE. https://doi.org/10.1109/ivs.2019.8814260

Valiente R, Zaman M, Fallah YP, Ozer S (2020) Connected and autonomous vehicles in the deep learning era: a case study on computer-guided steering. In: Handbook of pattern recognition and computer vision, World Scientific, pp 365–384. https://doi.org/10.1142/9789811211072_0019

Veeraraghavan H, Bird N, Atev S, Papanikolopoulos N (2007) Classifiers for driver activity monitoring. Transportation Research Part C: Emerging Technologies 15(1):51–67. https://doi.org/10.1016/j.trc.2007.01.001

Xing Y, Lv C, Wang H, Cao D, Velenis E (2020) An ensemble deep learning approach for driver lane change intention inference. Transportation Research Part C: Emerging Technologies 115:102615. https://doi.org/10.1016/j.trc.2020.102615

Yadav S, Patwa A, Rane S, Narvekar C (2019) Indian traffic signboard recognition and driver alert system using machine learning. International Journal of Applied Sciences and Smart Technologies 1(1):1–10. https://doi.org/10.24071/ijasst.v1i1.1843

Yan C, Zhang B, Coenen F (2015) Driving posture recognition by convolutional neural networks. In: 2015 11th international conference on natural computation (ICNC), IEEE. https://doi.org/10.1109/icnc.2015.7378072

You Z, Gao Y, Zhang J, Zhang H, Zhou M, Wu C (2017) A study on driver fatigue recognition based on SVM method. In: 2017 4th international conference on transportation information and safety (ICTIS), IEEE. https://doi.org/10.1109/ictis.2017.8047842

Younis MC (2021) Evaluation of deep learning approaches for identification of different corona-virus species and time series prediction. Comput Med Imaging Graphics:101921

Younis MC, Abuhammad H (2021) A hybrid fusion framework to multimodal bio metric identification. Multimed Tools Appl:1–24

Zhao C, He J, Zhu T, Lian J, Shen J, Zhang H (2011a) Recognition of driver's fatigue expressions by gabor wavelet transform and multilayer perceptron classifier. In: Proceedings 2011 international conference on transportation, mechanical, and electrical engineering (TMEE), IEEE, pp 617–620. https://doi.org/10.1109/tmee.2011.6199279

Zhao C, Zhang B, Lian J, He J, Lin T, Zhang X (2011b) Classification of driving postures by support vector machines. In: 2011 sixth international conference on image and graphics, IEEE. https://doi.org/10.1109/icig.2011.184

Zhao C, Zhang B, He J, Lian J (2012) Recognition of driving postures by contourlet transform and random forests. IET Intelligent Transport Systems 6(2):161. https://doi.org/10.1049/iet-its.2011.0116