IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 11, No. 2, June 2022, pp. 494~503
ISSN: 2252-8938, DOI: 10.11591/ijai.v11.i2.pp494-503

DistractNet: a deep convolutional neural network architecture for distracted driver classification

Ismail Nasri[1], Mohammed Karrouchi[1], Hajar Snoussi[1], Kamal Kassmi[1], Abdelhafid Messaoudi[2]
[1]Electrical Engineering and Maintenance Laboratory, High School of Technology, Mohammed First University, Oujda, Morocco
[2]Energy, Embedded Systems and Information Processing Laboratory, National School of Applied Sciences, Mohammed First University, Oujda, Morocco

Article Info

Article history:

Received Aug 22, 2021
Revised Dec 16, 2021
Accepted Jan 3, 2022

Keywords:

Convolutional neural network
Deep learning
Distracted driver
Safe driving
Transfer learning

ABSTRACT

Distracted driving has been considered one of the reasons for traffic accidents. The american national highway traffic safety administration (NHTSA) defines distracted driving as any activity that takes attention away from driving, such as doing makeup, texting, calling, and reaching behind. Most deaths, physical injuries, and economic losses could have been prevented if the distracted driver is alerted on time. This paper has proposed a new convolutional neural network (CNN) called DistractNet to detect drivers' distractions. The proposed model was trained and tested by state farm distracted driver detection image datasets available at Kaggle that contains images of drivers in the most common activities performed, which lead to distraction while driving divided into ten classes. Also, we have studied the performances of the proposed CNN model based on accuracy, training time, and model size. The performance of the proposed model was compared with four pre-trained networks such as ResNet-50, GoogLeNet, InceptionV3, and AlexNet using transfer learning techniques. The obtained experimental results show that the developed model-based CNN can achieve an overage accuracy of more than 99.32% with 93 min of training time and 7.99 MB of size. The extracted model can classify driver states into ten different classes with the predicted label and probability % for each class.

Corresponding Author:

Ismail Nasri
Electrical Engineering and Maintenance Laboratory, High School of Technology, Mohammed First University
BP. 473, Oujda 60000, Morocco
Email: i.nasri@ump.ac.ma

1. INTRODUCTION

In the United States, every day about eight people are killed in crashes due to a distracted driving [1]. According to the American national highway traffic safety administration (NHTSA), in 2018, distracted driving was responsible for 2,841 deaths and 400,000 injuries in a motor vehicle crash [2]. Distraction can be divided into three main types: visual, manual, and cognitive distraction [3]. The first type of distraction is defined as all tasks requiring the driver to look away from the roadway, such as texting, and reaching behind. Manual distraction represents all tasks that require the driver to take a hand off the steering wheel, likes doing makeup, operating the radio, and drinking. In cognitive distraction, the driver thinking about something other than the driving task, likes talking on the phone, and talking to the passenger. In recent years, distracted driver detection has received attention from the research community. Many solutions and approaches have been proposed to prevent road accidents. Also, the assistance systems and driving monitoring provide an efficient solution to reduce road accidents caused by distracted driving [4].

The first previously used method is based on image processing using deep learning techniques for features extraction that can be used to interpret the state of driver. The second class of techniques based on driving signals, such as acceleration, speed, gravity, and revolutions per minute. These driving signals can be accessed through the standard on-board diagnostics (OBD-II) port or using external physical sensors like accelerometer, and gyroscope. In this paper, we have designed a new convolutional neural network (CNN) [5], [6] model-based distracted driver detection, study and evaluate the performances of the developed model in terms of accuracy, and compare the performances of developed CNN with pretrained models using transfer learning techniques. The remainder of this paper is organized as follows: the rest of this section introduces some related works already published about distracted driver detection. In section 2, we describe the materials and methods used to develop the proposed CNN model. Section 3 provides the experimental result and comparison with related works. Section 4 provides a summary of the main findings of the work and suggestions for future research.

In related work [7], [8], the authors have proposed a portable system for monitoring and controlling driver behavior. The system was designed to acquire data from the vehicle using OBD-II. Similarly, in the paper, Ahmed *et al.* [9] have used the accelerometer and gyroscope sensors implemented in smartphones to detect distracted driving. To collect data from sensors, the authors have experimented with 16 subjects instructed to driving in most activities that led to distracting driving. The collected data have been considered as input of the machine learning classifier. The obtained system can reach accuracy in detecting un-distracted driving of 98.76%. In the paper, Shahverdy *et al.* [10], the authors aimed to classify the driving styles into five classes, including distracted, normal, aggressive, drowsy, and drunk driving using driving signals like acceleration, gravity, and throttle. In the paper, Craye and Karray [11], the authors proposed a module to detect distraction driving and recognize the type of distraction based on features extraction, including facial expressions, eye behavior, head movement, and arm position using computer vision techniques. The proposed module can achieve an accuracy score of 90% for distraction detection and 85% for recognition of the type of distraction. In the paper, Kutila *et al.* [12] have introduced a camera vision system to detect visual and cognitive distraction driving based on support vector machine (SVM) classification methods [13], [14]. This method achieved an accuracy of 80% for visual distraction and 86% for cognitive distraction. In the paper, Mbouna *et al.* [15], the authors have analyzed the eye state and head pose to monitor driver alertness. To extract the most critical information the authors have used visual features like pupil activity, eye index, and head pose. These features will be passed through a SVM to classify the driver's state. The obtained system can reach an accuracy of 91%.

2. MATERIALS AND METHOD

This section describes the materials and method: dataset description and method used to develop the CNN model architecture to detect and classify distracted drivers in teen class. For the training and test processing platform, we have used a machine with the following hardware characteristics: 3.6 GHz Intel (R) CPU Core i5-8350U. We have also used Matlab R2018b software for the development and execution of the proposed algorithm.

2.1. Dataset description

Several datasets are proposed for detection and classification of distracted driving. The American University in Cairo (AUC) distracted driver's dataset [16] consists of 17308 RGB images with a size of 1080×1920 obtained from a video of 11 participants from different ethnicities. The video input was acquired using a phone camera with a resolution of up to 1080×1920 pixels, placed on top of the passenger's seat to capture the driver's state in different vehicles and driving conditions.

The state farm distracted drivers dataset [17] available at Kaggle contains images of drivers in the most common activities performed, leading to distracted driving divided into ten classes. Figure 1 as shows the image samples of each class from the statefarm's dataset containing 22,424 RGB images with a size of 640×480. In Figure 1 there are pictures of safe driving shown in Figure 1(a), texting (right hand) as shown in Figure 1(b), talking (right hand) as shown in Figure 1(c), texting (left hand) as shown in Figure 1(d), talking (left hand) as shown in Figure 1(e), operating the radio as shown in Figure 1(f), drinking as shown in Figure 1(g), reaching behind as shown in Figure 1(h), hair and makeup as shown in Figure 1(i), and talking to passenger(s) as shown in Figure 1(j).

2.2. Proposed method

The proposed method is composed of five steps:

– Data preprocessing: statefarm's dataset [17] contains images with a size of 640×480 pixels. According to the input layer of proposed CNNs, we have resized all images dataset to the size of 224×224 pixels,

227×227, and 299×299 pixels, using the “imresize” function available in Image Processing Toolbox in Matlab software.
- Split data: processed images from the dataset have been divided into two categories to prepare the proposed CNN model, 70% (15,697 Images) of the dataset for training and 30% (6,727 Images) for testing. Table 1 shows the total samples for each class.
- Create and configure the CNN model: In this step, we have defined the CNN layers, including convolutional, pooling, and fully connected layer, each layer having a filter at different resolutions defined by width, depth, and height. The proposed CNN DistractNet was designed and trained using MATLAB with deep learning toolbox.
- Train CNN: the commonly used training options are: max epochs and learning rate value. The first one defines the elapsed time for training using the entire dataset, and the second controls the training speed.
- Evaluate CNN’s performances: we have evaluated the performances of the proposed CNNs based on classification accuracy, training time, execution time, and model size.

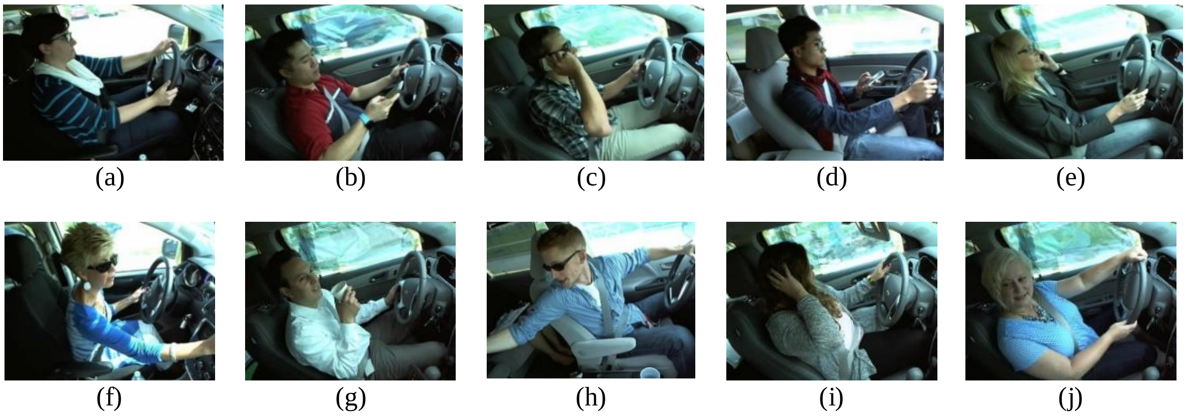


Figure 1. Image samples from statefarm’s dataset, (a) safe driving, (b) texting (right hand), (c) talking (right hand), (d) texting (left hand) (e) talking (left hand), (f) operating the radio, (g) drinking, (h) reaching behind, (i) hair and makeup, and (j) talking to passenger(s)

Table 1. Total samples from statefarm’s dataset for each class

Class	Samples	Training (70%)	Test (30%)
A-Safe driving	2489	1742	747
B-Texting (right hand)	2267	1587	680
C-Talking on the phone (right hand)	2317	1622	695
D-Texting (left hand)	2346	1642	704
E-Talking on the phone (left hand)	2326	1628	698
F-Operating the radio	2312	1618	694
G-Drinking	2325	1628	698
H-Reaching behind	2002	1401	601
I-Hair and makeup	1911	1338	573
J-Talking to passenger(s)	2129	1490	639
Total Samples	22424	15697	6729

2.3. Proposed CNN DistractNet

The proposed CNN DistractNet contains many layers, such as the input layer, output layer, and seven hidden layers, as shown in Figure 2. These layers are divided into two categories: features detection layers and classification layers. The feature detection layers consist of convolutional layers, pooling layers. The first layer consists of an input image with dimensions of 227×227×3 followed by a convolutional layer with 16 filters of size 3×3 resulting in dimension of 227×227×16. We have used three convolutional layers to extract the various features such as edges, color, and gradient orientation. The third layer is max poling with filter size 2×2 and stride of 2, resulting in an image dimension of 113×113×16. The pooling layer simplifies the output by reducing the number of parameters that the network needs to learn about. In the same way, the fourth layer is convolutional with 32 filters of size 3×3 followed by max pooling layer with a filter size of 2×2 and stride of 2, resulting in reduced image dimension 56×56×32. The sixth layer is convolutional with a 64-filter size of 3×3 and stride of 1 to have an image dimension of 56×56×32.

Finally, after several convolutional and pooling layers, the CNN shifts to classification. For that, we have used the fully connected layer (FC). The final layer of the DistractNet architecture is softmax, with 10 possible classes to provide the classification output and to get the probabilities for each class of the input classified image. Table 2 shows the training DistractNet accuracy and loss with varying epochs and iterations.

2.4. Transfer learning

Transfer learning [18]–[20] is currently the most and popular technique used in deep learning when the target dataset is relatively small or insufficient. The purpose of transfer learning is to reuse of a pre-trained CNN architecture for a similar or new related task [21]. In this study, we have used four pre-trained networks, such as ResNet-50 [22], GoogLeNet [23], inceptionV3 [24], and AlexNet [25]. Table 3 introduces an overview of the pre-trained CNNs used in this work.

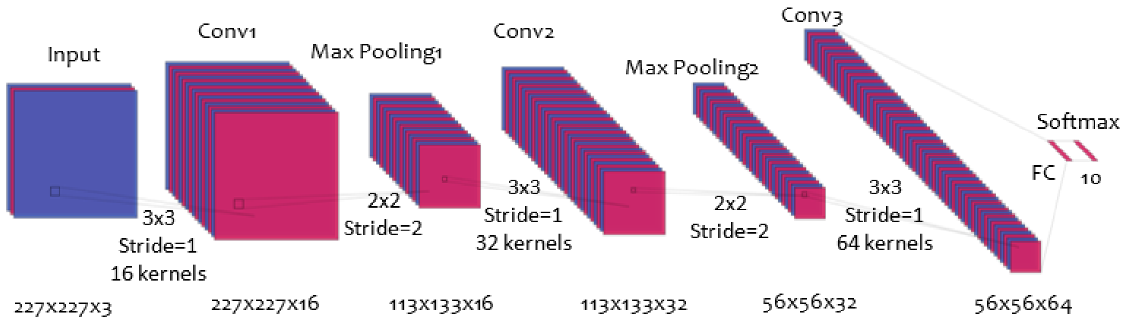


Figure 2. Proposed DistractNet architecture

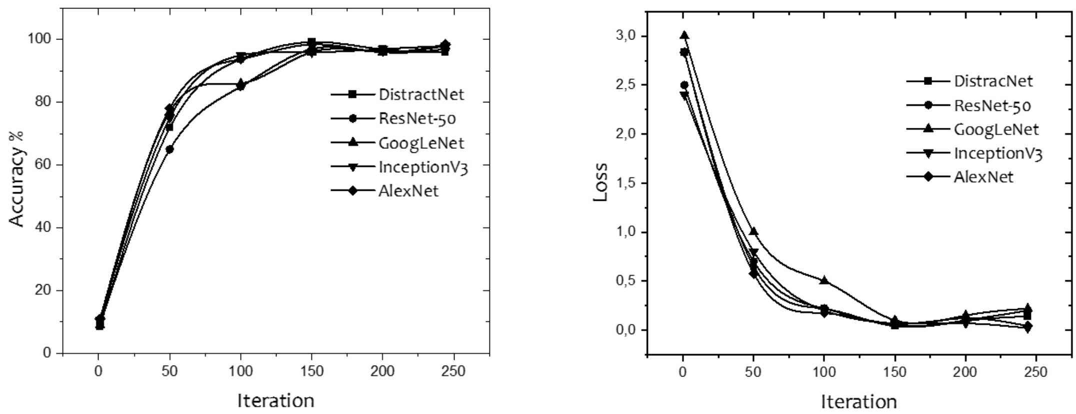
Table 2. Training and loss process of DistractNet

Epoch	Iteration	Accuracy	Loss	Learning rate
1	1	8.59%	2.8339	0.001
1	50	71.88%	0.6480	0.001
1	100	93.75%	0.2187	0.001
2	150	99.22%	0.0397	0.001
2	200	96.88%	0.0938	0.001
2	244	96.09%	0.1417	0.001

Table 3. Comparison of different CNNs model

Network	Year	Layers	Size input	Parameters
ResNet-50	2015	50	224×224×3	23 M
GoogLeNet	2014	22	224×224×3	7 M
InceptionV3	2015	48	299×299×3	23.6 M
AlexNet	2012	8	227×227×3	61 M

ResNet-50 [22] consists of 50 layers, 48 convolutional layers, and 1 maxpooll and 1 average pool layer. This CNN can be used on computer vision problems, including image classification, and object detection. The image input size of ResNet-50 is 224×224×3 with 23 million parameters. GoogLeNet [23] is a CNN that contains 22 layers. This network has two versions; one trained by ImageNet datasets [26], and can classify images into 1,000 object classes. The other version trained by places365 [27] datasets but classify images into 365 different place categories. The pretrained versions both have an image input size of 224×224×3. inceptionV3 [24], developed in 2015, has 48 deep layers trained and tested on more than a million images from the ImageNet datasets [26]. Inceptionv3 primarily aims to consume less computing power using less than 25 million parameters by modifying previous Inception architectures. The network has an image input size of 299×299×3. AlexNet [25] is a CNN that contains eight layers; three fully connected layers and five convolutional layers. The AlexNet can predect images into 1000 object categories. This CNN was trained on large image datasets such as the ImageNet [26], which contains 1.2 million of images divided into 1000 different classes. For this network, the size of the input image is 227×227×3. Figure 3 shows the training progress and loss plot with varying epochs and iterations for all CNNs used in this work, as shown in Figure 3(a) and Figure 3(b).



(a) (b)

Figure 3. Training and loss process of CNN models, (a) accuracy rates for training sets, and (b) loss values for training sets

3. RESULTS AND DISCUSSION

3.1. Result and analysis

Once the proposed DistractNet is trained from scratch, we can then evaluate the performances and demonstrate the efficiency of our proposed system. For that, we have used the confusion matrix [28] as shown in Figure 4 and calculated the accuracy for each class. 6729 of samples dataset contains images of drivers of various ethnicities in different scenarios, and conditions that have been used for testing. The overall accuracy represents the total number of correct predictions among the total number of the dataset (correct predictions + incorrect predictions) as shown in (1):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where, true positive (TP) is the number of positive class records correctly classified, true negative (TN) is is the number of negative class records correctly classified, false positive (FP) is the number of negative class records incorrectly classified, and false negative (FN) is the number of positive class records incorrectly classified.

According to the results obtained from the confusion matrix for each pretrained network as shown in Figure 5, we have observed that “Reaching behind” and “Talking to passenger(s)” classes are confused and misclassified with each other. The reason is that the position of the head is the same in those classes. Similarly, “Talking (right hand)” is confused with “Hair and makeup”.

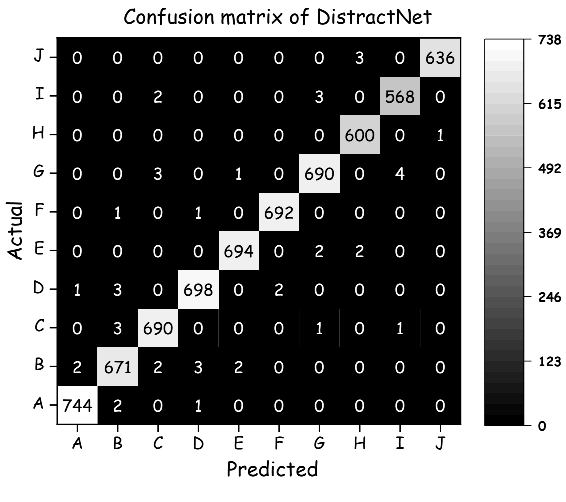
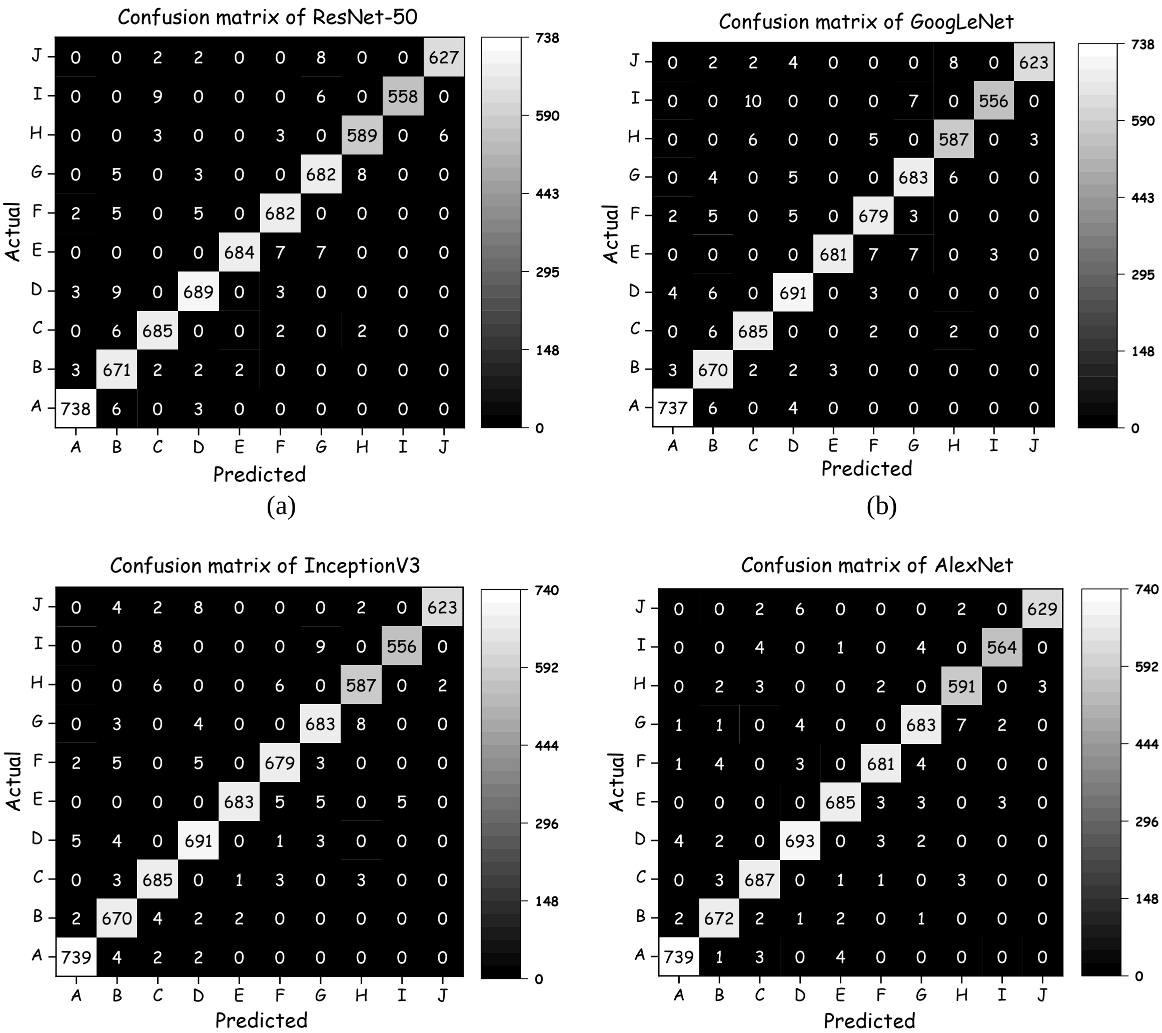


Figure 4. Confusion matrix of proposed CNN DistractNet

Table 4 as shows the performances of all CNN models used in this work. The accuracy of classification obtained using ResNet-50 as shown in Figure 5(a), GoogLeNet as shown in Figure 5(b), InceptionV3 as shown in Figure 5(c) and AlexNet as shown in Figure 5(d) were 98.16%, 97.96%, 98.02%, and 98.44% respectively for 22,424 images, two epochs of training, and 0.001 learning rate value. The obtained experimental results also demonstrate that the proposed model DistractNet is more accurate than all pre-trained networks with the smallest size (7.99 Mbit). The size of ResNet-50, GoogLeNet, InceptionV3 and AlexNet was 91.4 MB, 94.4 MB, 92 MB, and 629 MB. The model size is proportional to CNN architecture including, layers number, and parameters. We have also compared the execution time (speed) in seconds that a classification would take for all the proposed CNN models on central processing unit (CPU).



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

Figure 5. Confusion's matrix of pretrained networks, (a) ResNet-50, (b) GoogLeNet, (c) InceptionV3, and (d) AlexNet

The proposed CNN DistractNet has 2 million parameters in total, which is the number of weight and bias of convolution (Conv) and FC. The training time for transfer learning is small compared to training from scratch because, in transfer learning, the pre-trained model has already learned the weights based on previous learnings. This technique can reduce training time and computing resources. Also, the training time depends on many factors, such as the number of image datasets, networks architecture, and processing platform performances like processor, RAM, and Graphics. Figure 6 provides a comparison of the CNNs used in this work. The vertical axis shows the accuracy of classification. The horizontal axis shows the training time in minutes for 2 epochs. Circle size represents the number of parameters for each network.

Table 4. Performences comparison between DistractNet and pretrained networks

CNN Model	Accuracy %	Learning time (mm:ss)	Model size (MB)	Execution time (Speed)
DistractNet	99.32%	92:54	7.99	0.0299 s
ResNet-50	98.16%	75:15	91.4	0.0672 s
GoogLeNet	97.96%	39:11	94.2	0.0463 s
InceptionV3	98.02%	79:02	92	0.0563 s
AlexNet	98.44%	33:58	629	0.0878 s

In this part, we have evaluated the impact of the number of training images datasets on the accuracy of classification. The result shows that the best overall accuracy is obtained when the number of images is 22,424. In deep learning, precisely in CNNs, the number of images significantly affected the classification accuracy of the CNN models. Therefore, producing higher accurate results requires a large number of image datasets [29]. Figure 7 shows the accuracy of CNN networks with a different number of training images datasets.

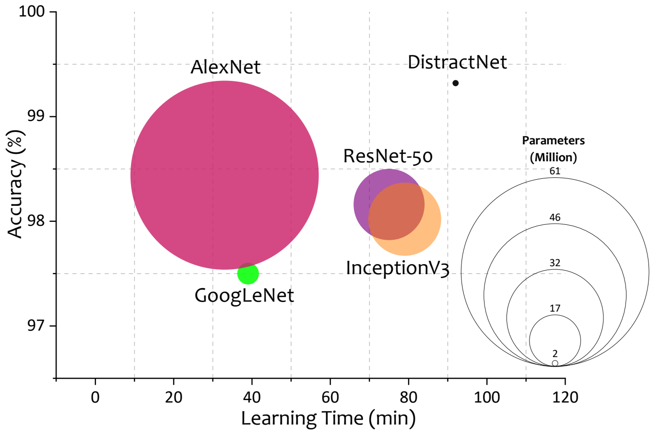


Figure 6. Accuracy, learning time, and number of parameters of CNNs models

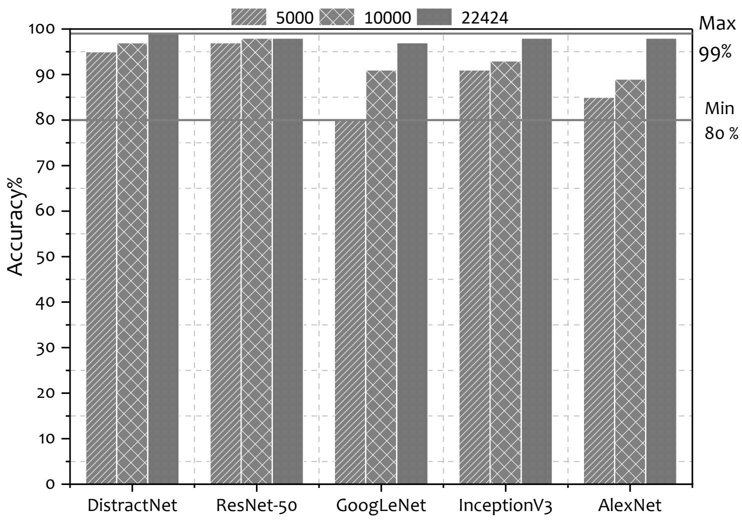


Figure 7. Accuracy comparison of CNN models with varying image number of dataset

3.2. Comparison with related works

All previously methods used for distracted driver detection in smart transportation systems [30]–[32] can be divided into two main classes. The first used method is based on data collected from sensors such as accelerometer and gyroscope using OBD-II or external sensors. The authors in [9], [10] aimed to detect distraction based on driving signals. This class of methods takes advantage of the number of sensors available within vehicles and can provide significant results. The accuracy associated with detection distractions reached more than 98%.

On the other hand, the second class is based on images processing captured by a camera module installed inside the vehicle to monitor driver behavior and classified using a machine learning algorithm, including CNN, Random Forests algorithm, and SVM. The authors in [11], [12] developed a machine learning classifier based on features extraction from image input to predict the driver behavior. Table 5 as shows the comparison of the performances of different methods.

Table 5. Accuracy comparison between proposed model and related works

Reference	Class	Method	Accuracy%
[9]	Sensor	Accelerometer, gyroscope using Random Forests based algorithm	98.76%
[10]	Sensor	Acceleration, throttle, gravity, speed, and revolutions per minute (RPM)	-
[11]	Vision	facial expressions, eye behavior, head movement, and arm position using computer vision techniques	90%
[12]	Vision	Support vector machine (SVM)	80%
[15]	Vision	Eye state and head pose using (SVM)	91%
Proposed DistractNet	Vision	convolutional neural network (CNN)	99.32%

4. CONCLUSION

In this work, we have proposed a new CNN named DistracNet for distracted driver detection and recognition nine of types of the most activities conducted to distraction such as texting, talking, operating the radio, and reaching behind. The StateFarm's distracted driver detection dataset was used for training and testing the proposed CNN model. Compared to other related works, the proposed model DistractNet was the most accurate model with an average accuracy of 99.32%. Also, the experiment results demonstrate that DistractNet has significant performances in terms of training time and model size.

In future work, we will focus on implementing the proposed CNN model in an embedded system able to be used in real-time to monitor driver states. Also, as an extension of this work, we will communicate the proposed system with other electronic control units (ECUs) in the vehicle network to take further necessary action to avoid accidents by warning drivers in a distraction state using a sound alarm, and text message.

REFERENCES

[1] K. Hurts, L. S. Angell, and M. A. Perez, "The Distracted Driver: Mechanisms, Models, and Measurement," *Rev. Hum. Factors Ergon.*, vol. 7, no. 1, pp. 3–57, Aug. 2011, doi: 10.1177/1557234X11410387.

[2] "Distracted Driving | Motor Vehicle Safety | CDC Injury Center." [Online]. Available: https://www.cdc.gov/transportationsafety/distracted_driving/index.html.

[3] Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: Less than the sum of its parts," *Accid. Anal. Prev.*, vol. 42, no. 3, pp. 881–890, May 2010, doi: 10.1016/j.aap.2009.05.001.

[4] M. Q. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," *Sensors (Switzerland)*, vol. 19, no. 11, p. 2574, Jun. 2019, doi: 10.3390/s19112574.

[5] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," Nov. 2015, [Online]. Available: https://arxiv.org/abs/1511.08458v2.

[6] K. K. Kumar, M. D. Kumar, C. Samsonu, and K. V. Krishna, "Role of convolutional neural networks for any real time image classification, recognition and analysis," *Mater. Today Proc.*, Mar. 2021, doi: 10.1016/j.matpr.2021.02.186.

[7] A. Khandakar *et al.*, "Portable system for monitoring and controlling driver behavior and the use of a mobile phone while driving," *Sensors (Switzerland)*, vol. 19, no. 7, p. 1563, Mar. 2019, doi: 10.3390/s19071563.

[8] M. Amarasinghe, S. Kottegoda, A. L. Arachchi, S. Muramudalige, H. M. N. D. Bandara, and A. Azeez, "Cloud-based driver monitoring and vehicle diagnostic with OBD2 telematics," in *15th International Conference on Advances in ICT for Emerging Regions, ICTer 2015 - Conference Proceedings*, Aug. 2016, pp. 243–249, doi: 10.1109/ICTER.2015.7377695.

[9] K. B. Ahmed, B. Goel, P. Bharti, S. Chellappan, and M. Bouhorma, "Leveraging Smartphone Sensors to Detect Distracted Driving Activities," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3303–3312, Sep. 2019, doi: 10.1109/TITS.2018.2873972.

[10] M. Shahverdy, M. Fathy, R. Berangi, and M. Sabokrou, "Driver behavior detection and classification using deep convolutional neural networks," *Expert Syst. Appl.*, vol. 149, p. 113240, Jul. 2020, doi: 10.1016/j.eswa.2020.113240.

[11] C. Craye and F. Karray, "Driver distraction detection and recognition using RGB-D sensor," Feb. 2015, [Online]. Available: https://arxiv.org/abs/1502.00250v1.

[12] M. Kutila, M. Jokela, G. Markkula, and M. R. Rué, "Driver distraction detection with a camera vision system," in *Proceedings - International Conference on Image Processing, ICIP*, 2006, vol. 6, pp. 201–204, doi: 10.1109/ICIP.2007.4379556.

[13] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 1048–1054, 1999, doi: 10.1109/72.788645.

[14] A. Tzotsos and D. Argialas, "Support vector machine classification for object-based image analysis," in *Lecture Notes in Geoinformation and Cartography*, vol. 0, no. 9783540770572, Springer Berlin Heidelberg, 2008, pp. 663–677.

[15] R. O. Mbouna, S. G. Kong, and M. G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013, doi: 10.1109/TITS.2013.2262098.

[16] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, "Real-time Distracted Driver Posture Classification," Jun. 2017, [Online]. Available: https://arxiv.org/abs/1706.09498v3.

[17] "State Farm Distracted Driver Detection | Kaggle." Accessed: Jun. 13, 2021. [Online]. Available: https://www.kaggle.com/c/state-farm-distracted-driver-detection.

[18] M. Shaha and M. Pawar, "Transfer Learning for Image Classification," in *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018*, Mar. 2018, pp. 656–660, doi: 10.1109/ICECA.2018.8474802.

[19] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowl. Inf. Syst.*, vol. 36, no. 3, pp. 537–556, Jun. 2013, doi: 10.1007/s10115-013-0665-3.

[20] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Syst.*, vol. 80, pp. 14–23, May 2015, doi: 10.1016/j.knosys.2015.01.010.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-Decem, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[23] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2015, vol. 07-12-June, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-Decem, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 {IEEE} Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 248–255, doi: 10.1109/cvpr.2009.5206848.

[27] B. Zhou, A. Lapedriza, A. Torralba, and A. Oliva, "Places: An Image Database for Deep Scene Understanding," *J. Vis.*, vol. 17, no. 10, p. 296, Aug. 2017, doi: 10.1167/17.10.296.

[28] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," in *Lecture Notes in Computer Science*, vol. 3408, Springer Berlin Heidelberg, 2005, pp. 345–359.

[29] I. Nasri, M. Karrouchi, H. Snoussi, A. Messaoudi, and K. Kassmi, "MaskNet: CNN for Real-Time Face Mask Detection Based on Deep Learning Techniques," in *Lecture Notes in Networks and Systems*, vol. 211 LNNS, Springer International Publishing, 2021, pp. 87–97.

[30] M. Ahmad *et al.*, "Adaptation of vehicular ad hoc network clustering protocol for smart transportation," *Comput. Mater. Contin.*, vol. 67, no. 2, pp. 1353–1368, 2021, doi: 10.32604/cmc.2021.014237.

[31] F. Ullah, A. H. Abdullah, O. Kaiwartya, J. Lloret, and M. M. Arshad, "EETP-MAC: energy efficient traffic prioritization for medium access control in wireless body area networks," *Telecommun. Syst.*, vol. 75, no. 2, pp. 181–203, 2020, doi: 10.1007/s11235-017-0349-5.

[32] I. Nasri, M. Karrouchi, H. Snoussi, K. Kassmi, and A. Messaoudi, "Detection and prediction of driver drowsiness for the prevention of road accidents using deep neural networks techniques," in *WITS 2020, Singapore*, 2022, pp. 57–64, doi: 10.1007/978-981-33-6893-4_6.

BIOGRAPHIES OF AUTHORS

Ismail Nasri received his M.Sc. degree in mechanical engineering from the Mohammed V University in Rabat, Morocco, in 2014. He is currently working toward a Ph.D. degree in Electrical Engineering and Maintenance Laboratory, High School of Technology, Mohammed First University, Oujda, Morocco. Since 2019, Ismail's research interests include Artificial Intelligence, Deep Learning, Computer Vision Techniques, and automotive technology. He can be contacted at email: i.nasri@ump.ac.ma.

Mohammed Karrouchi received his M.Sc. degree in optics and materials from Mohammed First University in Oujda, Morocco, in 2018. He is currently working toward a Ph.D. degree in Electrical Engineering and Maintenance Laboratory, High School of Technology, Oujda, Morocco. Since 2018, Mohammed's research interests include implementing an onboard system dedicated to inspecting the vehicle condition. He can be contacted at email: m.karrouchi@ump.ac.ma.

Hajar Snoussi received her M.Sc. degree in optics and materials from Mohammed First University in Oujda, Morocco. She is currently working toward a Ph.D. degree in Electrical Engineering and Maintenance Laboratory, High School of Technology, Mohammed First University, Oujda, Morocco. She can be contacted at email: h.snoussi@ump.ac.ma.

Kamal Kassmi has been a Professor at the High School of Technology, Mohammed First University, Oujda, Morocco, since October 1994, member of the Electrical Engineering and Maintenance Laboratory, High School of Technology. Kamal is the author of several publications in electrical electronics engineering. He can be contacted at email: kassmikamal@yahoo.fr.

Abdelhafid Messaoudi has been a Professor at the High School of Technology, Mohammed First University, Oujda, Morocco, since 2018, member of the Energy, Embedded Systems and Information Processing Laboratory at National School of Applied Sciences, Oujda, Morocco. Abdelhafid is the author of several publications in electrical engineering, including photovoltaic systems, image processing, renewable energy. He can be contacted at email: messaoudi1968abdelhafid@gmail.com.