

## Driver Distraction Detection using Hybrid CNN Method

Wijdan Abd Alhussain Abd Almutalib<sup>a\*</sup>, Thekra Hydar Ali Abbas<sup>b</sup>,  
Huda Abdulaali Abdulbaqi<sup>c</sup>

<sup>a\*</sup>Computing Department, College of Science, Mustansiriyah University, Iraq.

E-mail: rahaf.odai@uomustansiriyah.edu.iq

<sup>b</sup>Computing Department, College of Science, Mustansiriyah University, Iraq.

<sup>c</sup>Computing Department, College of Science, Mustansiriyah University, Iraq.

**Article History:** Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

**Abstract:** Traffic safety is an important issue in the world due to direct relationship with unsafe driving actions. This attitude belongs to the driver irresponsible behaviors. This research presents a solution for this problem by using modern technologies to follow the driver's behavior. The present method applies two effective techniques, namely Gaussian Mixture Model GMM and YCbCr color, to extract the useful properties of the image and insert them in deep learning technique for the purpose of achieving the best way to monitor the driver's movements. The proposed system consist of 3 stages: the first stage is preprocessing of RGB image to extract the region of interest (RoI). The second stage is segmentation process. it is achieved by converting the result of preprocessing stage to YCbCr color. The final stage is to use YCbCr outputs as an input representation to convolutional neural network (CNN) model to detect the final action. The main concern in this technique is to extract face and hands of the driver's body. These two parts represent the essential in the body that can be used to monitor the driver. The proposed model applied State Farm Dataset and achieved a classification accuracy of 96.59%. The results show that this method is superior in driver action recognition.

**Keywords:** Driver Distraction, YCbCr Color, CNN, GMM.

### 1. Introduction

Driver inattention is considered one of the main causes of accidents on highways. This was published by global means such as reports of the World Health Organization, which indicated that the number of deaths reached 1.35 million people in 2016 [1]. In addition, the National Highway Traffic Safety Administration (NHTSA) indicated that approximately 25% of accidents reported by the police were related to behaviors of the driver's inattention while driving, which results in the driver being distracted [2].

As a result of in-depth research in this regard, it was found that there are three main causes of driver distraction, which are visual, cognitive, and manual [3]. This phenomenon is one of the reasons that led to a dangerous and fatal situation for the human, and therefore finding effective measures to identify and reduce this phenomenon has been a major focus of many researches and studies. Contemporary studies of the problem began to reveal the causes of tired driving and explain its causes, effects and consequences [4]. The emergence of neural networks had an effective effect on the diagnosis and matching process. Whereas the recent development in the emergence of technologies that operate on the CNNs has brought about a qualitative leap in the processing of complex and high-resolution images, it has been used in the process of image classification, processing, recognition and other problems [5].

Therefore, the latest technologies were used in this research for the purpose of processing driving pictures and the related data through images captured from inside the vehicle for the purpose of indicating the driver's position while driving. The most influential techniques in this area, including GMM, RoI, and conversion of results to YCbCr, were used to extract the influencing regions that help inferring the distraction process. This induction was divided into several sections. The second section was concerned with previous studies related to it, and the third section clarified the technical methods used to solve the research problem and methods of implementation, while the other sections clarified the results and conclusions for the overall work.

### 2. Related Works

This part of the research explains the previous studies that focused on this topic, which emerged in large measure after the emergence of cell phones, which was one of the main causes of driver distraction while driving. Tamas et al, [6] proposed a CNN-based system for detecting driver distraction. They used VGG-16 and modified it for this task to train AUC dataset. Many types of activation functions (Leaky ReLU, DReLU, SELU) were used,

in addition to, squeeze-and-excitation module have been evaluated to achieve higher performance. The system obtained a classification accuracy of 95.82%.

While Huang et al, [7] proposed an automatic system to determine the driver's behavior through a surveillance camera installed on the dashboard. This method is characterized by the advantages that characterize the models through the application of HCRF technology. This method identifies the way the driver uses the mobile phone as a surveillance tool. As the accuracy of the classification obtained by the researchers does not exceed 94.2% . Also, CNN is used for the process of classification. Tran et al. [8] used pretrained models to extract distracted drivers features and train the application of SVM as a classification tool. Researchers extracted the results They achieved the recognition accuracy of 96.7%.

Baheti et al, [9] in 2018 solved the problem of distracted detection by using modified VGG-16 model in order to regularize techniques to prevent overfitting. After using AUC dataset, the classification accuracy result of this approach 95.54%. Also, they studied the effect of dropout, L2 regularization and batch normalization on the functioning of the system. Ramazan et al. [10] used a hierarchical classification approach to address dispersed leadership styles according to spatial and temporal requirements. The researchers presented a new concept in addressing the problem of distraction while driving, as they did not deal with the image as a fixed input, but rather as a variable. The results were accurate, as it achieved an accuracy rate of 89.62%.

Valeriano et al. [11] develop a novel technique to solved automatic recognition for driving distraction. The method used different deep learning-based methods to analyze driver behavior by applying data from cameras 2D. The evaluation was performed on the State Farm dataset. Results show that all evaluation methods observe more than 90% of accuracy. A contest Kaggle presented image dataset to classify pictures of the driver's behavior while driving [12] to cover the lack of data on problems and causes of distraction or distraction for the driver. It classified ten different behaviors for the driver while driving, as in Figure 4.

The new generation of technologies that address computer vision in driver distraction image have been used the deep learning [13] that work on high-resolution computers, which gave important advantages and unprecedented results in this field.

### **3. Districted Driving Recognition Method by Using CNN**

#### **3.1. Methodology Design**

Convolutional neural network design CNN is a multi-layer supervised learning method that can be considered a new generation of artificial neural network technique due to its unique operation. Its modified the unique operation based on the weight of parameters and specific layers operation through the training process which improved the model accuracy. The structure of traditional convolutional algorithm is the input layer, convolution layer, flatten layer, pooling layer, full connection layer and finally the output layer. These layers are classified the input data, extract the features and construct the outcomes [13].

The present method develop a novel structural design using enhanced input representation to apply them in CNN technique in order to enhance the capability of detecting and specifying driver image condition. This method develops a system for distracted driving recognition based on collected data by in-vehicle camera. The first step of this method is by loading the input RGB image and split it into 3 bands, then, preprocessing of each band by using GMM method to extract the ROI. After that, convert the results of GMM technique to YCbCr color. This step is essential to extract the face and hand of driver's body from the converted image. Finally, classify output YCbCr segmented image by using CNN model which is expand the traditional CNN network from "one convolution and one pooling" to "three convolutions and three pooling". The network structure of distracted driving recognition model based on CNN is illustrated in the Figure 1.

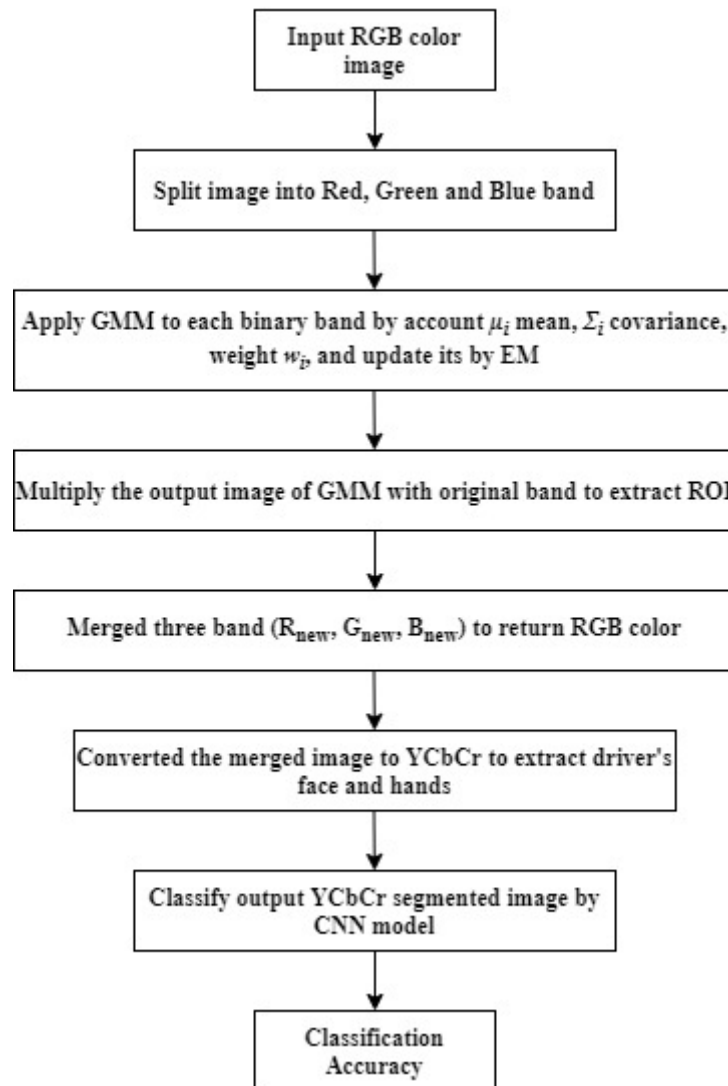


Figure 1. The proposed model for driver distraction

### 3.2. Proposal Method

The proposed method include 3 steps: preprocessing as illustrated in subsection (3.2.1) that load image and apply Gaussian filter for smooth image to extract the region of interest (ROI). while subsection (3.2.2) involved the process steps that skin segmentation of the driver's. While subsection (3.2.3) explains the structure of the CNN classifier.

#### 3.2.1. Image Pre-processing Method by GMM

Gaussian mixture model GMM is a parametric probability density function which can be considered as a Gaussian summation of weighted densities [14].

The GMM parameters are specified by using Expectation-Maximization EM method or maximum a posteriori MAP estimation based on training data [15]. On the other hand, the original images of driver will be stored in the format of in the statefarm dataset choice the Specifications with 640x360x3 which is suitable with this system methodology, all details will be describe in section (4.1). The driver body region in the image will be selected by extract the ROI [16]. The RGB space color involves three channels in each pixel (i.e. red, green and blue). The rang of color value in these pixels are (0 to 255) [17].

In this step started by loading the input RGB image and split into 3bands (R,G,B), then preprocessing the image by apply Gaussian filter for smooth image to extract the region of interest (ROI) by used **Equation (1)** [15], and clustering by GMM each band of the RGB image in two group by using **Equation (2)** [14] to extract the skin and non-skin region which will be used later.

$$f(y) = \sum_{j=1}^k \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp \left\{ -\frac{1}{2} (y - \mu_j)^T \Sigma_j^{-1} (y - \mu_j) \right\} \tag{1}$$

where  $\mu_j$  is mean,  $\Sigma_j$  is covariance and  $y$  is image (matrix).

A Gaussian mixture model is a weighted sum of  $M$  Gaussian components which is given by [14]:

$$p(y | skin) = \sum_{i=1}^M w_i f(y | \mu_i, \Sigma_i) \tag{2}$$

where,  $y$  is a  $d$  dimensional feature vector,  $w_i, i = 1, \dots, M$ , are the mixture weights with constraints  $\sum_i w_i = 1$  and  $f(y|\mu_i, \Sigma_i)$  where  $i = 1, \dots, M$ , are the Gaussian components.

where, weight  $w_i$ , mean  $\mu_i$  and covariance  $\Sigma_i$  are obtained by Expectation Maximization (EM) algorithm. then determine the threshold of each band by take mean of center point of each group and convert each band according threshold.

In order to detect the ROI, the binary band will be multiplied by the original band as shown in Figure 2.

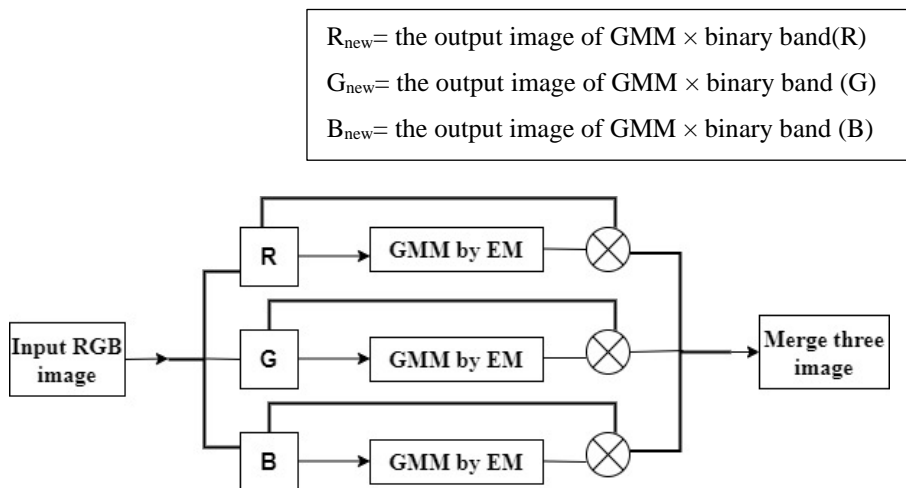


Figure 2. Preprocessing the input image by GMM method

### 3.2.2. Skin Segmentation by YCbCr Color

In this step, The resulting image is from GMM was converted to YCbCr color according to **Equation (3)** to extract face and hands of the driver's body. we simply segment the region of skin based on the following decision rules for the pixel value in YCbCr color space:

$$\left\{ \begin{array}{l} 0 \leq Y \leq 255 \\ 133 \leq Cb \leq 173 \\ 120 \leq Cr \leq 180 \end{array} \right.$$

The basic equations to convert between RGB and YCbCr are [18]:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.256 & 0.504 & 0.098 \\ -0.148 & -0.292 & 0.441 \\ 0.441 & -0.369 & -0.071 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{3}$$

After extract face and hands of the driver's The skin segmented region is obtained as an input ready for the classification process through the CNN model To improve classification accuracy.

### 3.2.3. The structural of the CNN Classifier

#### 3.2.3.1. Design of Convolution Layer and Pooling Layer

In the process of deep learning, convolution layer and pooling layer are the unique neuron layers in CNN network, which are used for feature extraction of image data [19]. In proposed method, the input representation of ConvNets is RGB of size 240x240. Table 1 illustrates the CNN model that was used throughout our approach. According to the sequence of "convolution layer pooling layer convolution layer pooling layer convolution layer pooling layer", the features of convolution and pooling are extracted three times. In this experiment, ReLU function is adopted, like  $f(x) = \max(0, x)$ , and the size of convolution core is  $3 \times 3$ . In the design of pooling layer, the max pooling method is used to calculate the input features. Max-pooling is performed with a window size of  $2 \times 2$  [20] pixels.

#### 3.2.3.2. Full connection layer and output layer design

The proposed method adopts a stack of convolutional layers, as they are two fully connected methods. They arrange and organize the data in a one-dimensional scanning method with a specific vector, and build the layer with 1024 neurons, then 256 neurons [21] as illustrates in Table 1. It also uses the ReLU Activation function.

That the model contains 10 categories leads to the classification of the information system into ten classes.

So the final layer was used with a fully connected layer with 10 channels to perform the 10-way classifications, and the weights and displacement values were updated with the optimizer Adam[22]. Adam successfully improved performance while ensuring that the "categorization\_ crossentropy" function gradually shrank and converged perfectly [23]. We set the number of epochs in CNN's training to 10 and the minibatch size to 32 as illustrate in [21] with learning rate 10-3 as illustrate in [24].

The pseudo code of an Algorithm (1) is given below for training the CNN

Algorithm (1): Training the CNN
Input: dataset result of GMM and YCbCr, epochs = 10, options. Minibatch = 32, learning rate = 0.001
Output: model of CNN
Step1: Initialize CNN configuration by prepare (call) network cnn train.net =CNN Initialization according Table (1)
Step2: Prepare Data set by split the dataset to input as image and label (10 class) To prepare data input to network.
Step3: Start training with CNN train.net, input image, Label, epochs, minibatch, learning rate
Step4: End

**Table 1.** The CNN model

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 238, 238, 128)	3584
max_pooling2d_1	(MaxPooling2 (None, 119, 119, 128)	0
conv2d_2 (Conv2D)	(None, 117, 117, 64)	73792
max_pooling2d_2	(MaxPooling2 (None, 58, 58, 64)	0
conv2d_3 (Conv2D)	(None, 56, 56, 32)	18464
max_pooling2d_3	(MaxPooling2 (None, 28, 28, 32)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_1 (Dense)	(None, 1024)	25691136
dense_2 (Dense)	(None, 256)	262400
dense_3(Dense)	(None, 10)	2570
Total params: 26,051,946		
Trainable params: 26,051,946		
Non-trainable params: 0		

Found 17943 images belonging to 10 classes.

Found 4481 images belonging to 10 classes.

The proposed method steps can be describe in Figure (3)

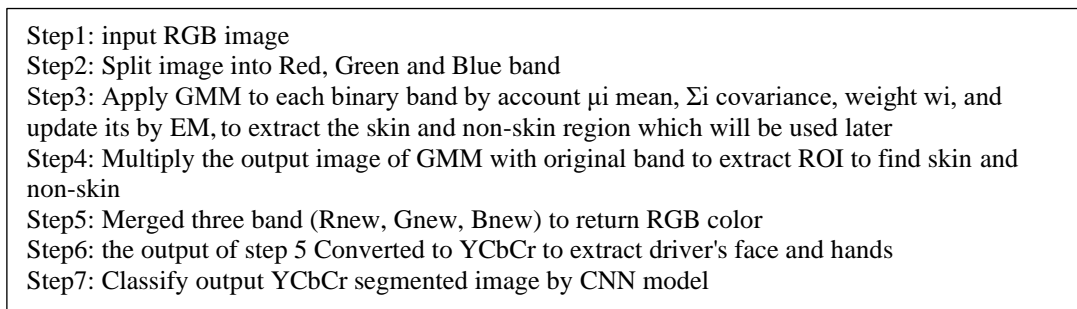


Figure 3: the steps of methodology

#### 4. Experimental Analysis

##### 4.1. Description of the Dataset

The dataset can be described as the basic units of probability that can be used to determine the causes and consequences of the driver distraction investigation. Therefore, the current data was obtained from a company kaggle state farm in the United States of America. The company collected and organized the data in the form of systematic pictures of different scenes of drivers. The company took into consideration the color, age, and tire type, as well as the driver’s physical characteristics such as skin color. The collection was completed in 2016 to be ready for use by researchers to find ways to organize studies and research in this field. The data divided the driver’s condition into ten sections, which are safe driving, talking on the phone, writing messages correctly, in addition to other activities such as distracting driving [19]. Each image in the dataset is a 640 x 480 JPE G file with a storage capacity of about 45 KB. The total number of samples is 22424 files as it appears in each image as shown in (Table 2) and the sample data is illustrated in Figure 4.

**Table 2.** List of distracted driving activities

Activity	Class	Sample count
Safe driving	C0	2489
Texting using right hand	C1	2267
Talking on the phone using right hand	C2	2317
Texting using left hand	C3	2346
Talking on the phone using left hand	C4	2326
Operating the radio	C5	2312
Drinking	C6	2325
Reaching behind	C7	2002
Hair and makeup	C8	1911
Talking to passenger	C9	2129
Total	10	22424



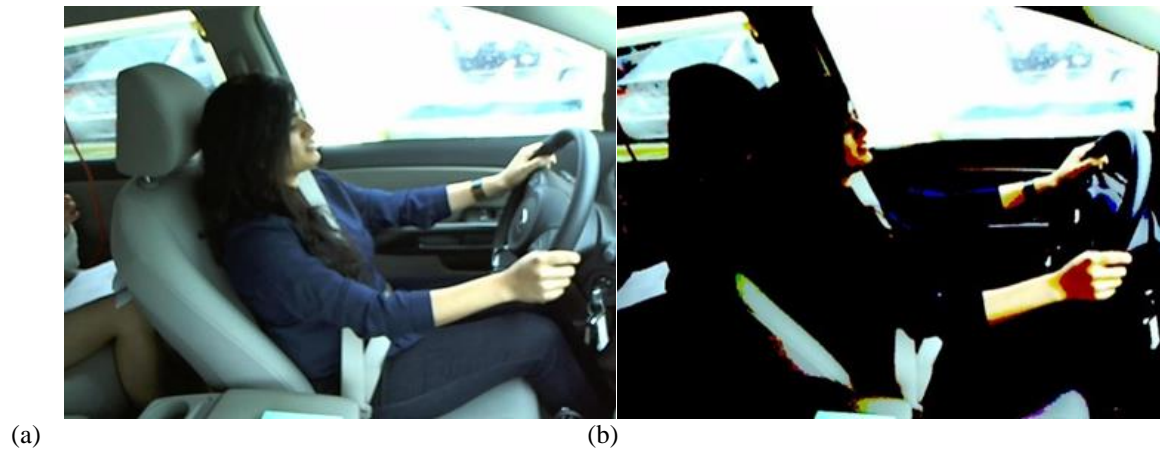
Figure 4. Sample of kaggle state frame images [25]

#### 4.2. Results of Proposed Method

In this experiment The code was mainly implemented in the Python language. the Keras deep learning framework (version :2,2,4) based on TensorFlow (version: 1.12.0) is used to develop and implement the algorithm . Furthermore, the total data sets are randomly selected 80% for model training and 20% for model testing.

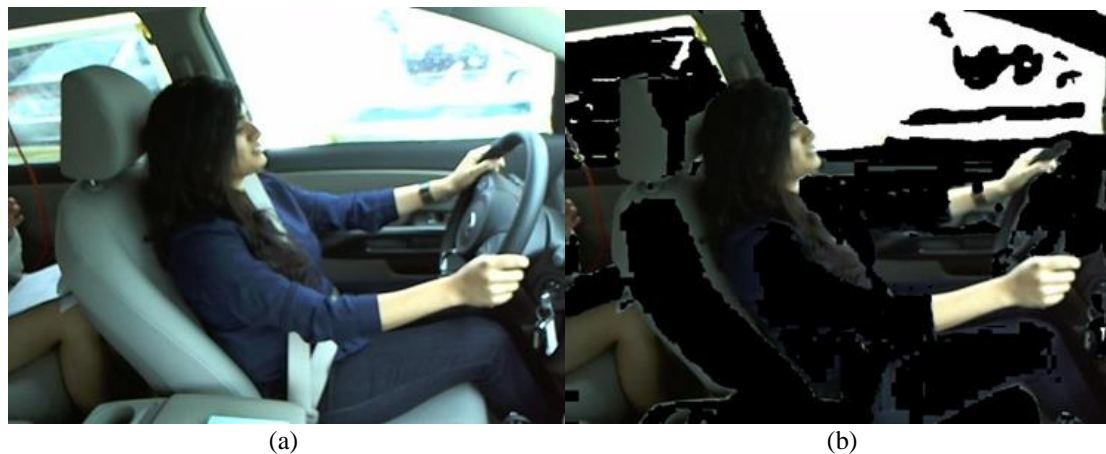
After extracting the ROI from the preprocessing on the input RGB image, the result shown in Figure 5 was obtained, where the important areas appear more luminous than the rest of the image. When testing this model with CNN it had a good classification accuracy of 96.78%.

This proves that this model can determine the driver's head and hands area through GMM segmentation method and driver behavior classification with high efficiency.



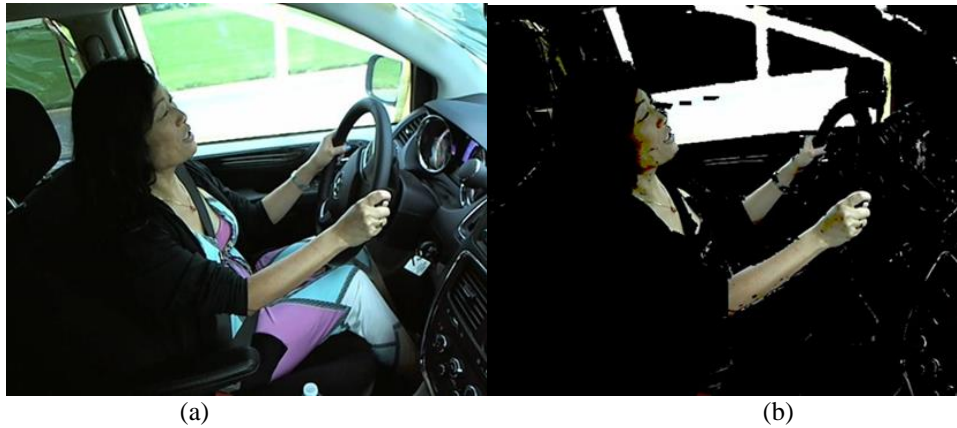
**Figure 5.** The effect of GMM segmentation pre-processing a. original image, b. extracting the ROI from the preprocessing on the input RGB image

At the same time, another color space, YCbCr, was used to extract the skin pixels. This model pre-processed the input RGB image by using **Equation (3)** and in the testing the model based on CNN not satisfactory result as the classification accuracy was 93. 27%. As in Figure 6.



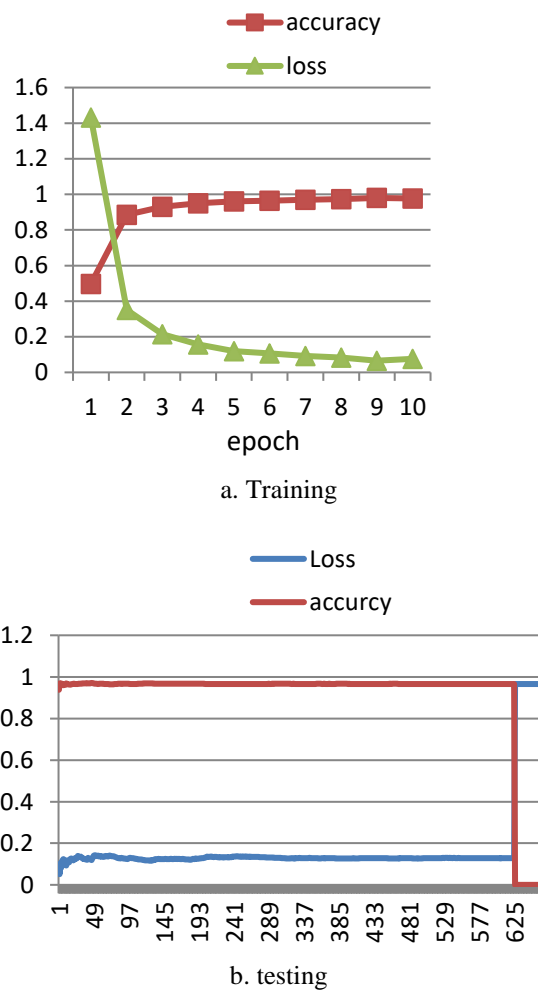
**Figure 6.** The effect YCbCr color space on input image. (a). original image, (b). after pre-processed the input RGB image by used YCbCr color to extract the skin pixels.

Also, the two models were merged together to extract the ROI by apply GMM and convert the output from preprocessing to YCbCr for extracting the driver's face and hands and then classify the output YCbCr by CNN as shown in Figure 7, with a rating of 96.59% accuracy for recognizing driver distraction.



**Figure 7.** The two models were merged together to extract the ROI by apply GMM and convert the output from preprocessing to YCbCr for extracting the driver's face and hands a. original image, b. the result of merged the two models

On the other hand, setting the number of epochs in the training of the CNN to 10 and the minibatch size to 32 with learning rate 0.001. Since with dealing a multiclass classification problem, used the loss function categorical cross entropy and the optimizer Adam. According to these uses, the accuracy of CNN network model and the convergence of loss function within the training and testing period are shows in Figure 8.



**Figure 8.** Convergence function and accuracy through training and testing period

In Figure 8, we can see that in the last epoch we have the lowest error rates in both the training set and testing set. These characteristics inform that we have a low bias and variance and that our proposed approach can realize the necessary generalization to reduce the possibility of overfitting.



### 4.3. Comparison with Other Approaches

Selected several approaches for detecting distracted driving behaviors for comparison with this proposed. The results reveal that the CNN-based approaches perform much better than other approaches, which occurs because the CNN-based approaches learn more features.

As shows in the Figure 9 Whereas, GMM based CNN to detect driver dispersion gave a satisfactory result than the other our models (Ycber based CNN) and (GMM, YCbCr based CNN).

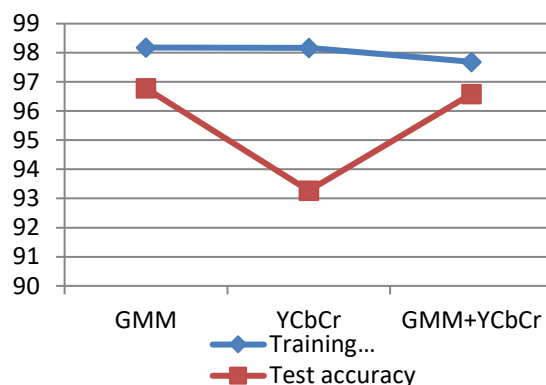


Figure 9. Compared between the three models GMM, YCbCr, and GMM+YCbCr

In Figure 10, Eraqi et al. uses Gaussian distributions (Likelihoods) are constructed for the skin and the non-skin classes after 30 epoch with an initial learning rate of  $10^{-2}$ , which achieved accuracy 94.66% [26], while after 10 epoch our method reached accuracy 96.59% that used GMM to preprocessing the image and converted to another color space for skin segmentation As shown in the (Table 3) .

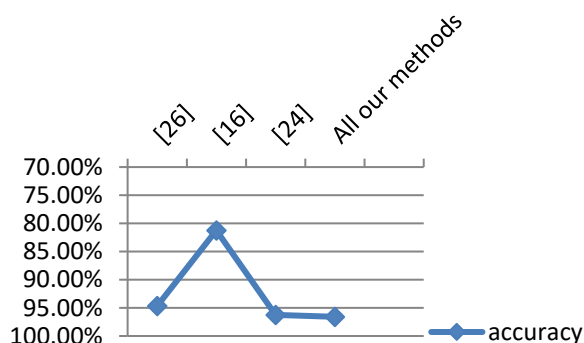


Figure 10. Comparison with other models

While Alotaibi et al. The approaches proposed that combines three of the most advanced models in deep learning after 30 epoch their method achieved accuracy 96.23% [24], while our method obtain better accuracy after 10 epoch and learning rate to 0.001.

Also Xing et al. proposed a method using GMM segmentation and transfer learning models achieved detection accuracy 81.23% [16], while our method apply GMM on raw RGB image based CNN that obtained classification accuracy 96.78% as on (Table 3).

Table 3. A proposed of three models

Method based CNN	Training accuracy	Test accuracy
GMM	98.18	96.78
YCbCr	98.17	93.27
GMM+YCbCr(all our methods)	97.69	96.59

### 5. Conclusion

Distracted driving behaviors are a primary cause of traffic accidents. Hence, it is necessary to find methods to effectively identify distracted driving behaviors.

The present paper proposed a hybrid GMM, YCbCr and CNN framework to recognize distracted driver behaviors.

To identify the driver's behavior and classify this approach, it preprocessed the input image by GMM to extract region of interest, then converted this image to YCbCr to extract the driver's face and hands.

The results showed that the proposed system achieve high performance due to extract the region of interest that help inferring the distraction process and improvement of accuracy, as the classification accuracy reached 96.59%.

### Acknowledgements

Authors thankful Department of Computer Science, Collage of Science, Mustansiriyah University, for supporting this work.

### References

1. Naomi, T. (2018). Traffic accidents are eighth leading cause of death globally, according to WHO. *Cable News Network*. <https://edition.cnn.com/2018/12/07/health/who-road-safety-report-intl/index.html>
2. SADD. "Distracted Driving". In: <https://www.sadd.org/initiatives/trafficsafety/distracted-driving>.
3. CDC. "Distracted Driving - Motor Vehicle Safety."
4. [https://www.cdc.gov/motorvehiclesafety/distracted\\_driving/index.html](https://www.cdc.gov/motorvehiclesafety/distracted_driving/index.html)
5. T. Suzuki et al., "Steering behavior model of drivers on driving simulator through visual information," 2016 *Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA 2016*, pp. 7–10, 2017.
6. H. Salman, J. Grover, and T. Shankar, "Hierarchical Reinforcement Learning for Sequencing Behaviors," vol. 2733, pp. 2709–2733, 2018.
7. Tamas, V., & Maties, V. (2019). Real-time distracted drivers detection using deep learning. *American Journal of Artificial Intelligence*, 3(1), 1-8.
8. Hoang Ngan Le, T., Zheng, Y., Zhu, C., Luu, K., & Savvides, M. (2016). Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 46-53.
9. Tran, D., Do, H.M., Lu, J., & Sheng, W. (2020). Real-time Detection of Distracted Driving using Dual Cameras. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014-2019.
10. Baheti, B., Gajre, S., & Talbar, S. (2018). Detection of distracted driver using convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 1032-1038.
11. Ramzan, M., Khan, H.U., Awan, S.M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A survey on state-of-the-art drowsiness detection techniques. *IEEE Access*, 7, 61904-61919.
12. Valeriano, L.C., Napoletano, P., & Schettini, R. (2018). Recognition of driver distractions using deep learning. In *2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, 1-6.
13. State Farm Corporate, "State farm distracted driver detection," April 2016. Competition website is Kaggle.com, <https://www.kaggle.com/c/state-farm-distracted-driver-detection>, <https://www.kaggle.com/c/state-farm-distracted-driver-detection>
14. 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.

15. Jaisakthi, S. M., & Mohanavalli, S. (2015). Skin Segmentation using Ensemble Technique. *Research Journal of Applied Sciences, Engineering and Technology*, 9(11), 963-968.
16. A.N. Hidayanto, & E.M. Koeanan. Journey on Image Clustering Based on Color Composition, 4(7), 1188–1193, 2010.
17. Xing, Y., Lv, C., Wang, H., Cao, D., Velenis, E., & Wang, F.Y. (2019). Driver activity recognition for intelligent vehicles: A deep learning approach. *IEEE transactions on Vehicular Technology*, 68(6), 5379-5390.
18. Matroushi, G. I. A. (2018). *Object detection, recognition and classification using computer vision and artificial intelligence approaches* (Doctoral dissertation, Loughborough University).
19. M.A. Yousif, “Tongue Print Recognition Based on Extreme Learning Machine and Convolutional Neural Network,” 2019.
20. Rao, X., Lin, F., Chen, Z., & Zhao, J. (2021). Distracted driving recognition method based on deep convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing*, 12(1), 193-200.
21. Koedwiady, A., Bedawi, S.M., Ou, C., & Karray, F. (2017). End-to-end deep learning for driver distraction recognition. *In International Conference Image Analysis and Recognition*, 11-18.
22. Masood, S., Rai, A., Aggarwal, A., Doja, M.N., & Ahmad, M. (2020). Detecting distraction of drivers using convolutional neural network. *Pattern Recognition Letters*, 139, 79-85.
23. Eraqi, H.M., Abouelnaga, Y., Saad, M.H., & Moustafa, M.N. (2019). Driver distraction identification with an ensemble of convolutional neural networks. *Journal of Advanced Transportation*, 2019.
24. Mase, J.M., Chapman, P., Figueredo, G.P., & Torres, M.T. (2020). A hybrid deep learning approach for driver distraction detection. *In 2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 1-6.
25. Alotaibi, M., & Alotaibi, B. (2020). Distracted driver classification using deep learning. *Signal, Image and Video Processing*, 14(3), 617-624.
26. O.G. Basubeit, D.N.T. How, Y.C. Hou, and K.S.M. Sahari, “Distracted Driver Detection with Deep Convolutional Neural Network,” *Int. J. Recent Technol. Eng.*, 8(4), 6159–6163, 2019.
27. Eraqi, H. M., Abouelnaga, Y., Saad, M.H., & Moustafa, M.N. (2019). Driver distraction identification with an ensemble of convolutional neural networks. *Journal of Advanced Transportation*, 2019.