

## Research Article

# Moving Vehicle Detection and Classification Using Gaussian Mixture Model and Ensemble Deep Learning Technique

Preetha Jagannathan,<sup>1</sup> Sujatha Rajkumar ,<sup>2</sup> Jaroslav Frnda ,<sup>3</sup>  
Parameshchari Bidare Divakarachari ,<sup>4</sup> and Prabu Subramani <sup>5</sup>

<sup>1</sup>Department of Computer Science and Engineering, Muthayammal Engineering College Autonomous, Rasipuram, Tamil Nadu 637408, India

<sup>2</sup>Department of Embedded Technology, School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

<sup>3</sup>Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, 01026 Zilina, Slovakia

<sup>4</sup>Department of Telecommunication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru 570016, India

<sup>5</sup>Department of Electronics and Communication Engineering, Mahendra Institute of Technology, Tamil Nadu 637503, India

Correspondence should be addressed to Jaroslav Frnda; jaroslav.frnda@fpedas.uniza.sk

Received 23 February 2021; Revised 20 March 2021; Accepted 18 May 2021; Published 27 May 2021

Academic Editor: Laurie Cuthbert

Copyright © 2021 Preetha Jagannathan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent decades, automatic vehicle classification plays a vital role in intelligent transportation systems and visual traffic surveillance systems. Especially in countries that imposed a lockdown (mobility restrictions help reduce the spread of COVID-19), it becomes important to curtail the movement of vehicles as much as possible. For an effective visual traffic surveillance system, it is essential to detect vehicles from the images and classify the vehicles into different types (e.g., bus, car, and pickup truck). Most of the existing research studies focused only on maximizing the percentage of predictions, which have poor real-time performance and consume more computing resources. To highlight the problems of classifying imbalanced data, a new technique is proposed in this research article for vehicle type classification. Initially, the data are collected from the Beijing Institute of Technology Vehicle Dataset and the MIOvision Traffic Camera Dataset. In addition, adaptive histogram equalization and the Gaussian mixture model are implemented for enhancing the quality of collected vehicle images and to detect vehicles from the denoised images. Then, the Steerable Pyramid Transform and the Weber Local Descriptor are employed to extract the feature vectors from the detected vehicles. Finally, the extracted features are given as the input to an ensemble deep learning technique for vehicle classification. In the simulation phase, the proposed ensemble deep learning technique obtained 99.13% and 99.28% of classification accuracy on the MIOvision Traffic Camera Dataset and the Beijing Institute of Technology Vehicle Dataset. The obtained results are effective compared to the standard existing benchmark techniques on both datasets.

## 1. Introduction

In recent times, developing an intelligent traffic surveillance system has become an emerging research topic, where it delivers an innovative tool to improve driver satisfaction, efficiency, and transportation safety [1]. Automatic vehicle clas-

sification plays a crucial role in intelligent traffic surveillance systems, and it supports several applications like traffic flow analysis, electronic toll collection, and intelligent parking systems [2, 3]. Due to the COVID-19 outbreak and mobility restrictions, citizens were allowed to move out of the home to procure only essential goods in groceries or pharmacies.

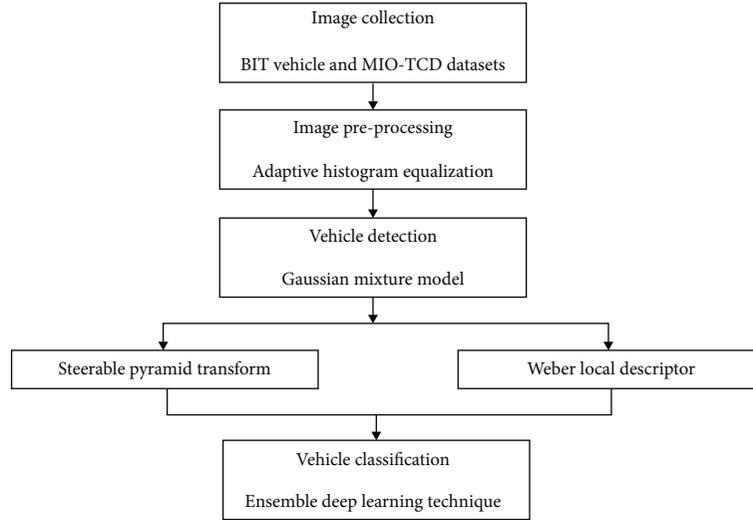


FIGURE 1: Flow diagram of the ensemble deep learning technique.

Intelligent traffic surveillance systems can track down any motorists entering to the worst-affected region from low-risk areas.

Automatic vehicle classification is a challenging task while the videos are being collected from traffic surveillance cameras [4]. Captured traffic surveillance images are lower-resolution images and are subjected to several weather conditions, illumination conditions, and occlusion [5]. In addition, vehicle types generate a lot of intra-class and inter-class similarities which affect vehicle classification performance [6]. In order to address the aforementioned problems, several machine learning methods and data manipulation techniques have been developed in order to deal with the imbalanced data classification [7–9]. Compared to other objects, vehicles have different structural characteristics, larger intra-class variations, and larger inter-class distances, and these factors make vehicle detection and classification a challenging task [10] because a single classifier in the classification stage would seem impossible to detect. The existing research on various detection mechanisms has resulted in efficient identification of incidences while others have the same limitations of standard identification versions [11, 12]. The motivation of this research study is to highlight the aforementioned issues and to deal with the imbalanced data, and a new technique is proposed in this research paper for vehicle type classification.

Initially, the surveillance videos or images are collected by the Beijing Institute of Technology (BIT) Vehicle Dataset and the MIOvision Traffic Camera Dataset (MIO-TCD). Additionally, the visual ability of the collected vehicle images is improved by implementing the Adaptive Histogram Equalization (AHE) method and then the Gaussian Mixture Model (GMM) which are utilized to detect vehicles from the denoised images. The GMM model provides higher detection accuracy, adaptation to image content, simplicity of implementation, and fast computation in vehicle detection. After recognizing the vehicles, the hybrid feature extraction is accomplished by using the Steerable Pyramid Transform (SPT) and the Weber Local Descriptor (WLD)

to extract feature vectors from the detected images. By implementing high-level global descriptors, the semantic gap between the extracted feature vectors is limited and results in better classification, reduced training time, and overfitting issues. Finally, the ensemble deep learning technique is used to classify the vehicle types such as the 11 classes in MIO-TCD and the 6 classes in the BIT Vehicle Dataset. Lastly, the proposed ensemble deep learning technique performance is analyzed in terms of the False Discovery Rate (FDR), the False Omission Rate (FOR), recall, precision, and accuracy. The simulation results confirmed that the proposed ensemble deep learning technique is significant in vehicle type classification related to the state-of-the-art techniques. In contrast, one of the drawbacks of using the ensemble deep learning technique is the vanishing gradient problem, which occurs when a large input space is mapped into a smaller one; this problem can be highlighted in future work.

## 2. Related Work

Liu et al. [13] developed the Generative Adversarial Nets (GANs) to classify vehicles from traffic surveillance videos. The developed approach consists of three steps in vehicle classification. Initially GAN was trained on a collected traffic dataset to generate adversarial samples for the rare classes. In the second step, an ensemble-based Convolutional Neural Network (CNN) was trained on the imbalanced dataset, and then sample selection was carried out to eliminate the lower quality adversarial samples. Finally, the selected adversarial samples were utilized to refine the ensemble model on the augmented dataset. Extensive experiments showed that the developed GAN approach achieved effective performance in vehicle classification on MIO-TCD by means of the Cohen kappa score, mean recall, precision, and mean precision. However, degradation issues will occur in the developed GAN approach, when the deeper networks are about to converge. Fu et al. [14] developed a new vehicle classification technique on the basis of a hierarchical multi-SVM (multi-Support Vector Machine) classifier. Initially, the foreground



FIGURE 2: Sample images of BIT Vehicle Dataset: (a) SUV, (b) sedan, (c) minivan, (d) microbus, (e) bus, and (f) truck.

objects were extracted from the surveillance videos, and then the hierarchical multi-SVM technique was developed for vehicle classification. Additionally, a voting-based correction approach was used to track the classified vehicles for the performance evaluation. In this literature study, a practical system was developed based on the hierarchical multi-SVM technique for robust vehicle classification in a heavy traffic scene. Hence, the developed technique is ineffective in practical crowded traffic scenes, due to the different views, shadows, and heavy occlusion. Further, Şentaş et al. [15] used

the tiny YOLO with the SVM classifier for vehicle detection and classification. In the experimental segment, the performance of the developed model was validated on the BIT Vehicle Dataset in light of precision and recall. The result of the experiment confirms that the developed model significantly classifies the vehicle types in real-time-streaming traffic videos. However, SVM was a binary classifier, which supports only binary classification that was a major limitation in this study. Wang et al. [16] developed a vehicle type classification system based on the faster R-CNN technique.

TABLE 1: Data statistics of the MIO-TCD classification dataset.

| Vehicle type         | Training | Testing |
|----------------------|----------|---------|
| Articulated truck    | 10,346   | 2,587   |
| Pedestrian           | 6,262    | 1,565   |
| Car                  | 260,518  | 65,131  |
| Pickup truck         | 50,906   | 12,727  |
| Bicycle              | 2,284    | 571     |
| Single-unit truck    | 5,120    | 1,280   |
| Nonmotorized vehicle | 1,751    | 438     |
| Bus                  | 10,316   | 2,579   |
| Motorcycle           | 1,982    | 495     |
| Work van             | 9,679    | 2,422   |
| Background           | 160,000  | 40,000  |
| Total                | 519, 164 | 129,795 |

The performance of the developed technique was evaluated on a real-time dataset which contains real scene images captured at the crossroads. As a future enhancement, a novel technique is needed to improve the ability to detect a vehicle which is occluded due to different illumination conditions, angles, and scales of the images. Zhuo et al. [17] developed a CNN model for vehicle classification which includes two important steps such as fine tuning and pretraining. In the pretraining step, GoogLeNet was applied on the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) dataset in order to get the initial model with connection weight. In the fine tuning step, the obtained initial model was fine-tuned on the vehicle dataset to achieve final classification. In this literature study, the collected highway surveillance videos include six vehicle categories like van, minibus, truck, bus, car, and motorcycle. In the experimental phase, the performance analysis was carried out on the vehicle dataset by means of accuracy. However, the developed CNN model is computationally expensive and has a major problem of “overfitting.” Murugan and Vijaykumar [18] developed a new framework for vehicle type classification that includes six main phases such as data preprocessing, detection of the vehicles, vehicle tracking, structural matching, extraction of the features, and vehicle classification. After collecting the traffic surveillance videos, data preprocessing was accomplished by using noise removal and color conversion. Further, the Otsu thresholding algorithm and background subtraction were used to detect the vehicles. Then, vehicle tracking was accomplished using the Kalman filter in order to track the moving vehicles. Additionally, the log Gabor filter and the Harrish corner detector were used to extract the feature vectors, and then the obtained features were fed to the Artificial Neural Fuzzy Inference System (ANFIS) for classification of the vehicles. Extensive experiments showed that the developed framework achieved significant performance in vehicle classification in light of error rate and accuracy. The developed framework increases the dimensionality issue that accounts for the model complexity. Dong et al. [19] implemented a new semisupervised CNN architecture for vehicle type classification. In the developed architecture, a sparse Laplacian filter was applied to extract

the rich and discriminative information of the vehicles. In the output layer, a softmax classifier was trained by multitask learning for vehicle type classification. In this literature study, the features learned by the semisupervised CNN architecture were discriminative to work significantly in the complex scenes. Extensive experiments were evaluated on the BIT Vehicle Dataset and a public dataset in order to analyze the efficiency of the developed architecture in light of classification accuracy. The semisupervised CNN architecture includes several layers, so the training process consumes more time. Hedeya et al. [20] introduced a new densely connected single-split super learner and applied variants for vehicle type classification on the BIT Vehicle Dataset and MIO-TCD. The developed model was simple, and it does not require any logic reasoning and hand-crafted features to achieve better vehicle type classification performance. In the complex datasets, the developed model introduces the vanishing gradient problem that is a major concern in this literature study. Soon et al. [21] implemented a new semisupervised model for vehicle type classification on the basis of Principal Component Analysis Convolutional Network (PCN). In the developed model, convolutional filters were utilized to extract the hierarchical and discriminative features. The simulation result showed that the developed model obtained better performance in real-time applications, due to its robustness against noise contaminations, illumination conditions, rotation, and translation. The developed PCN model contains a greater number of training parameters that lead to an overfitting problem.

Awang et al. [22] developed the Sparse-Filtered CNN with Layer Skipping (SF-CNNLS) approach for vehicle type classification. In this literature study, three channels of the SF-CNNLS approach were applied to extract discriminant and rich vehicle features. Additionally, the global and local features of the vehicles were extracted from the three channels of an image based on their color, brightness, and shape. In the Experimental Results and Discussion, the performance of the developed SF-CNNLS approach was validated on a benchmark dataset. Finally, the softmax regression classifier was used to classify the vehicle types like truck, minivan, bus, passenger, taxi, car, and SUV. The developed softmax regression classifier includes higher-level layers; however, by embedding lower-resolution vehicle images, there may be a loss of vehicle type information. Nasaruddin et al. [23] developed an attention-based approach and a deep CNN technique for lightweight moving vehicle classification. In this literature, the developed model performance was validated on a real-time dataset by means of specificity, precision, and  $f$ -score. However, the developed model performance was limited in such circumstances as baseline, camera jitter classes, and bad weather. The methods undertaken, datasets, advantage of using the developed methods in vehicle type classification, and disadvantage of the methods are clearly given for each literature paper. In order to address the above stated issues, a new ensemble deep learning technique is proposed in this research paper to improve vehicle type classification.

This paper is organized as follows. Methodology introduces two vehicle datasets and their parameters, as well as



(a)



(b)



(c)



(d)



(e)



(f)

FIGURE 3: Continued.



(g)



(h)



(i)



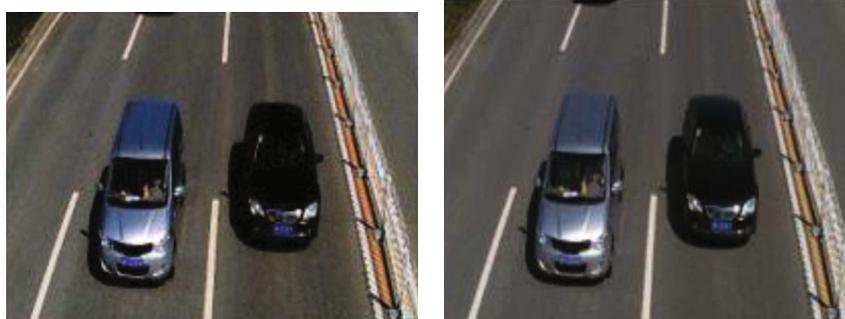
(j)

FIGURE 3: Continued.



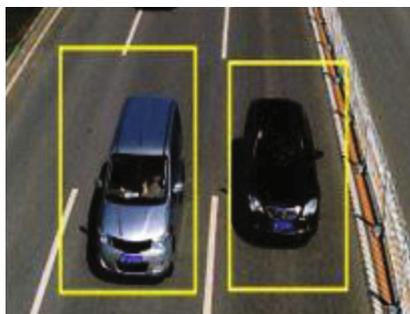
(k)

FIGURE 3: Sample images of the MIO-TCD classification dataset: (a) pickup truck, (b) articulated truck, (c) car, (d) work van, (e) pedestrian, (f) single-unit truck, (g) bus, (h) bicycle, (i) motorcycle, (j) nonmotorized vehicle, and (k) background.



(a)

(b)



(c)

FIGURE 4: (a) Original image, (b) preprocessed image, and (c) vehicle-detected image.

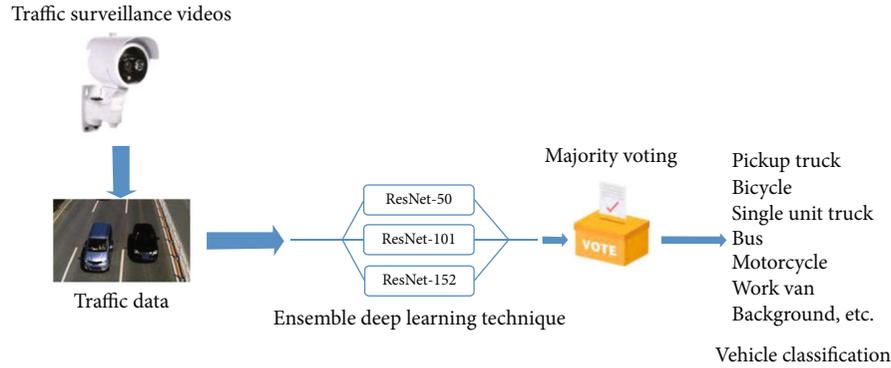


FIGURE 5: Architecture of the ensemble deep learning technique.

**Input:** Size of feature space, training set, size of feature subspace, feature set, number of feature subspace, one test sample, and number of classes.  
**Output:** Classification of vehicle types.  
**Process:**  
 For  $i = 1$ : classes  
     Label the samples of  $i^{\text{th}}$  class.  
     Train the feature subsets using ResNet-152, ResNet-101, and ResNet-50.  
**End for**  
 Calculate the value of counter  
**Output.**

PSEUDOCODE 1: Pseudocode of ensemble deep learning technique.

preprocessing data techniques and selected machine learning algorithms. The Experimental Results and Discussion describes the performance of the ensemble deep learning technique in terms of classification accuracy, provides comparative analysis between the proposed and existing technique, and discusses benefits and weaknesses of selected models. Finally, the last section presents our conclusions.

### 3. Methodology

In a recent scenario, vehicle type classification is the emerging research area in intelligent traffic systems, due to its wide range of applications that includes intelligent parking systems and traffic flow statistics [24]. Many approaches have been developed using vehicle type classification, which are commonly based on cameras, magnetic induction, and optic fibres [25]. The image-based approaches received great attention in the computer vision community with the extensive use of traffic surveillance cameras. The flow diagram of the ensemble deep learning technique is given in Figure 1.

**3.1. Image Collection.** In this research study, the proposed ensemble deep learning technique performance is tested on the BIT Vehicle Dataset and MIO-TCD. The BIT Vehicle Dataset is comprised of 9850 vehicle images with pixel sizes of  $1920 \times 1080$  and  $1600 \times 1200$ , which have been captured using two different cameras at different places and time. The BIT Vehicle Dataset consists of six vehicle types, namely, sedan, microbus, SUV, minivan, bus, and truck, and there are

5922, 883, 1392, 476, 558, and 822 images for each corresponding vehicle type [26]. The captured images are varied in terms of view points, surface color of the vehicles, scales, position of the vehicles, and illumination conditions. Due to the sizes of the vehicles and capturing delay, the top and bottom parts of the vehicles are not included in the images. The location of every vehicle is preannotated in the BIT Vehicle Dataset, because some images include one or two vehicles in the same image. The sample images of the BIT Vehicle Dataset are given in Figure 2. The BIT Vehicle Dataset link is as follows: <https://www.programmersonought.com/article/7654351045/>.

In addition, the MIO-TCD classification dataset is comprised of 648,959 vehicle images, and it includes eleven vehicle types: bicycle, articulated truck, motorcycle, nonmotorized vehicle, bus, car, pedestrian, work van, pickup truck, single-unit truck, and background [27]. The data statistics of the MIO-TCD classification dataset is stated in Table 1. Every annotated image in the BIT Vehicle Dataset and MIO-TCD is stored in a structured format. Sample images of the MIO-TCD classification dataset are given in Figure 3. The MIO-TCD classification dataset link is as follows: <https://github.com/hakimamarouche/MIO-TCD-classification>.

**3.2. Image Preprocessing and Vehicle Detection.** After collecting the vehicle images, the AHE technique is used to enhance the visual ability of the images by altering the global image contrast. Additionally, the AHE technique calculates several

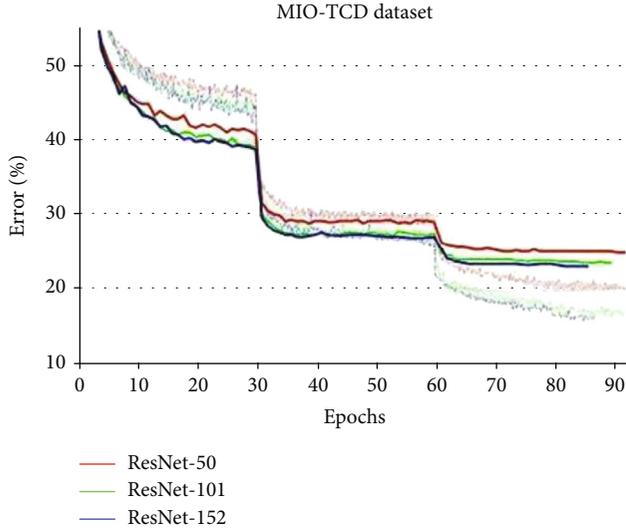


FIGURE 6: Training error of ResNet-152, ResNet-101, and ResNet-50 on MIO-TCO.

TABLE 2: Performance analysis of the proposed ensemble deep learning technique on the BIT Vehicle Dataset in terms of precision, recall, and accuracy.

| Feature extraction | Classifier | Precision (%) | Recall (%) | Accuracy (%) |
|--------------------|------------|---------------|------------|--------------|
| SPT                | MSVM       | 64.90         | 78         | 75           |
|                    | KNN        | 69            | 60.83      | 72           |
|                    | DNN        | 70.45         | 72         | 62.43        |
|                    | LSTM       | 73.97         | 80.80      | 79.60        |
|                    | Ensemble   | 78.91         | 86.82      | 90           |
| WLD                | MSVM       | 70            | 75         | 80           |
|                    | KNN        | 70.02         | 79.97      | 81           |
|                    | DNN        | 78.20         | 86.55      | 81.02        |
|                    | LSTM       | 79            | 84.60      | 83.20        |
| Hybrid (SPT + WLD) | Ensemble   | 81.02         | 84.39      | 86.22        |
|                    | MSVM       | 82.94         | 92.19      | 93           |
|                    | KNN        | 87            | 92         | 94.94        |
|                    | DNN        | 92.28         | 97.20      | 96.66        |
|                    | LSTM       | 93.90         | 96.97      | 98.98        |
|                    | Ensemble   | 98.24         | 99.72      | 99.28        |

histogram values for redistributing the lightness values of the vehicle images that enhances the local contrast and definitions of edges in every region of a vehicle image. Firstly, the collected images are considered as  $x$ , and the number of gray level occurrences  $i$  in the collected images is indicated as  $n_i$  [28]. Hence, the probability of a gray level occurrence is computed using

$$p_x(i) = p(x = i) = \frac{n_i}{n}, \quad 0 \leq i < L, \quad (1)$$

where  $L$  is indicated as the number of image gray levels,

TABLE 3: Performance analysis of the proposed ensemble deep learning technique on the BIT Vehicle Dataset in terms of FDR and FOR.

| Feature extraction | Classifier | FDR (%) | FOR (%) |
|--------------------|------------|---------|---------|
| SPT                | MSVM       | 34.7    | 20.91   |
|                    | KNN        | 20      | 34.72   |
|                    | DNN        | 28      | 22      |
|                    | LSTM       | 18.90   | 17      |
|                    | Ensemble   | 11.02   | 12      |
| WLD                | MSVM       | 29      | 18.42   |
|                    | KNN        | 29.98   | 14      |
|                    | DNN        | 24      | 13.18   |
|                    | LSTM       | 18.07   | 7.2     |
|                    | Ensemble   | 12.03   | 6.5     |
| Hybrid (SPT + WLD) | MSVM       | 13      | 11      |
|                    | KNN        | 9       | 5.01    |
|                    | DNN        | 7.67    | 3.10    |
|                    | LSTM       | 6.65    | 2.87    |
|                    | Ensemble   | 3.92    | 1.90    |

which ranges between 0 and 255;  $n$  is denoted as total image pixels; and  $p_x(i)$  is stated as the histogram value of the image pixel which is normalized between  $[0, 1]$ . Further, the cumulative distribution function (CDF) is computed for  $p_x$  using

$$\text{cdf}_x(i) = \sum_{j=0}^i p_x(x=j). \quad (2)$$

Then, a transformation form  $y = (x)$  is developed to generate a new image  $y$  with the flat histogram values. The transformed vehicle images have a linear CDF which is mathematically stated in

$$\text{cdf}_y(i) = iK, \quad (3)$$

$$(y') = \text{cdf}_y(T(k)) = \text{cdf}_x(k), \quad (4)$$

where  $K$  and  $T$  are represented as constant variables that range between  $[0, 1]$ , and the  $k$  variable is in the range of  $[0, L]$ . In the AHE technique, a simple transformation is applied to map the pixel values back into their original image, which is mathematically determined in

$$y' = y \times (\max(x) - \min(x)) + \min(x). \quad (5)$$

After image preprocessing, GMM is applied to detect vehicles from the preprocessed images,  $y'$ . In the field of vehicle type classification, GMM is used for detecting and recognizing moving objects [29]. GMM is a statistical model, which describes spatial distribution and the properties of the data in the parameter space. GMM includes a parametric probability density function, which is comprised of numerous Gaussian component functions for detecting vehicles

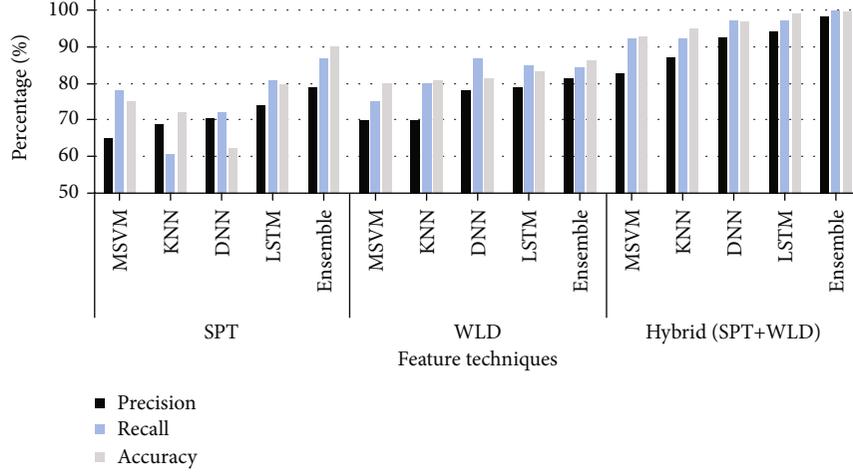


FIGURE 7: Graphical comparison of the ensemble deep learning technique on the BIT Vehicle Dataset in terms of precision, recall, and accuracy.

from the images [30], that is mathematically defined in equation (6). Sample preprocessed and vehicle-detected images are graphically represented in Figure 4:

$$\text{Prob}(y') = \sum_{i=1}^Q \pi_i R_i(y' | \mu_i, C_i), \quad (6)$$

where  $R_i(y' | \mu_i, C_i)$  is denoted as bivariate normal distribution with mean vector  $\mu_i$ ,  $\pi_i$  is denoted as the  $i$ th prior probability of Gaussian distribution, where the data sample produces, and  $C_i$  is indicated as a covariance matrix.

**3.3. Feature Extraction and Vehicle Classification.** After vehicle detection, SPT and WLD are combined to extract feature vectors from the detected images, which decreases the overfitting risks, speeds up the training process, and enhances the data visualization ability. SPT is a linear multiorientation and multiscale image decomposition method, and it is developed to overcome the concerns of orthogonal separable wavelet decomposition [31]. At first, the SPT decomposition method categorizes the detected images into several orientations, and then scales the images based on the derivate operators in different directions with variable sizes, even though the bandwidth orientation of the subbands are equal to  $2\pi/o$ , where  $o$  is stated as the number of orientations. The resultant subbands of the SPT method are rotation invariant and translation invariant [32].

In the SPT method, the detected images are decomposed into high- and low-frequency components using H0 and L0 filters. Additionally, the lower-frequency components are decomposed into two oriented band-pass components, and the low-frequency components are decomposed by using the oriented band-pass filters B0 and B1 and the low-pass filter L1. The more the number of orientations (increasing the derivative degree), the greater the number of pyramid levels produced and the finer is the orientation and scale tuning, which means a more robust representation of the images. In the SPT method, orientation of the filters should satisfy the following conditions:

- (i) The linear combination of the filters generates a filter of any orientation
- (ii) The filters are copied and rotated to develop another filter. So, all the filters are copies rotated from their counterparts

Next, every subband is convolved with the texture descriptor WLD to extract the active features from the images. WLD is a robust local texture feature descriptor, which is inspired from Weber's law. WLD is comprised of two components such as image orientation and differential excitation to extract texture features from the vehicle-detected image. Hence, the differential excitation component is used for reflecting the changes of current pixels [33–35], which is computed by utilizing

$$\xi(x_a) = \arctan(G_{\text{ratio}}(x_a)) = \arctan\left[\sum_{i=0}^{m-1} \left(\frac{x_i - x_a}{x_a}\right)\right], \quad (7)$$

where  $\xi(x_a)$  is represented as the differential excitation of the current pixel  $x_a$ ,  $G_{\text{ratio}}(x_a)$  is stated as the ratio of the difference in current pixel intensity,  $x_i$  ( $i = 0, 1, 2, 3 \dots m-1$ ) is represented as  $i$ th neighboring pixel of  $x_a$ , and  $m$  is stated as the number of neighbors. Further, the gradient orientation component of the current pixel  $x_a$  is calculated using

$$\theta(x_a) = \arctan\left(\frac{v_s^{11}}{v_s^{10}}\right), \quad (8)$$

where  $v_s^{11}$  and  $v_s^{10}$  are the outputs of two filters, filter<sub>11</sub> and filter<sub>10</sub>, which are used to compute the differences between current and neighborhood image pixels, and  $\theta$  is in the range of  $[-\pi/2, (\pi/2)]$ . Next, the extracted active feature vectors  $F$  are fed to the ensemble deep learning technique for vehicle classification.

Additionally, an ensemble deep learning technique is proposed for vehicle type classification on traffic surveillance videos. The extracted features  $F$  are fed to the ensemble deep

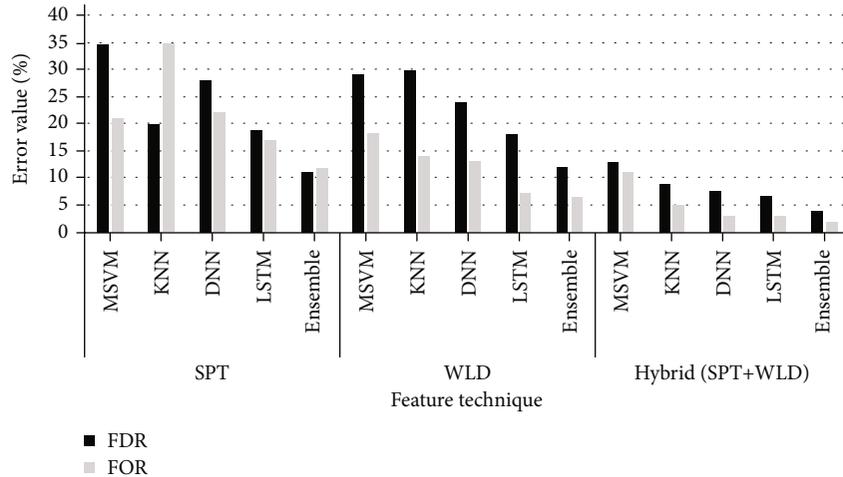


FIGURE 8: Graphical comparison of the proposed ensemble deep learning technique on the BIT Vehicle Dataset in terms of FDR and FOR.

learning technique in the input layer that reduces the classification bias and the training time. In order to highlight the concerns occurring because of extreme imbalanced data distributions, hybrid feature extraction (SPT and WLD) is devised in this research. Additionally, the size of the minority vehicle classes is reduced to a small number compared to the majority classes in the practical applications to avoid overfitting problems. One of the major advantages of using ResNet-152, ResNet-101, and ResNet-50 models is while it increases network depth, it also effectively eliminates negative outcomes. The proposed ensemble deep learning technique consists of a set CNN models which are trained on the balanced dataset with good initialization (pretrained on ImageNet). At last, the output of the ensemble techniques are combined by maximum voting policy based on the predictions of an individual technique.

As represented in Figure 5, the ensemble deep learning technique includes ResNet-152, ResNet-101, and ResNet-50. The proposed ensemble deep learning technique consists of three key phases: CNN techniques with good initial parameters, fine tuning of network parameters, and averaging models.

The residual networks (ResNets) are easy to optimize with limited training error, and it also gains higher classification accuracy from large datasets like the BIT Vehicle Dataset and MIO-TCD. The training error of ResNet-152, ResNet-101, and ResNet-50 on MIO-TCD is indicated in Figure 6. By increasing the number of epochs, the error percentage gradually decreases in the ResNet-152, ResNet-101, and ResNet-50 techniques. Pseudocode of the ensemble deep learning technique is given below.

#### 4. Experimental Results and Discussion

In this research, the proposed ensemble deep learning technique performance is simulated using MATLAB 2019a software with the following system requirements: operating system—Windows 10 (64 bit); processor—Intel core i9; hard disk—3 TB; and RAM—16 GB. In this research, the ensemble deep learning technique performance is validated by compar-

TABLE 4: Performance analysis of the proposed ensemble deep learning technique on MIO-TCD by means of precision, recall, and accuracy.

| Feature extraction | Classifier | Precision (%) | Recall (%) | Accuracy (%) |
|--------------------|------------|---------------|------------|--------------|
| SPT                | MSVM       | 67.84         | 70         | 82           |
|                    | KNN        | 70.40         | 70.27      | 65           |
|                    | DNN        | 78.22         | 89         | 80.10        |
|                    | LSTM       | 85            | 88.29      | 88           |
|                    | Ensemble   | 89            | 89.54      | 90.72        |
| WLD                | MSVM       | 70            | 80         | 90           |
|                    | KNN        | 78            | 80.90      | 92.04        |
|                    | DNN        | 83            | 90.72      | 93           |
|                    | LSTM       | 87.09         | 92         | 94.20        |
|                    | Ensemble   | 89            | 93.39      | 97           |
| Hybrid (SPT + WLD) | MSVM       | 72.34         | 93.02      | 96           |
|                    | KNN        | 82.02         | 94         | 97.77        |
|                    | DNN        | 92.31         | 96.20      | 96           |
|                    | LSTM       | 98.70         | 98.82      | 98.62        |
|                    | Ensemble   | 99.12         | 99.69      | 99.13        |

ing with a few benchmark techniques such as the GAN-based deep ensemble technique [13], the tiny YOLO with SVM [15], the semisupervised CNN model [19], PCN [21], and the three channels of SF-CNNLS (TC-SF-CNNLS) approach [22]. The primary goal of this research study is to classify the vehicle types from the BIT Vehicle Dataset and MIO-TCD. The proposed ensemble deep learning technique performance is validated using 10-crossfold validation. Let FP be indicated as false positive, FN be denoted as false negative, TP be stated as true positive, and TN be represented as true negative. Five performance measures are used to analyze the performance of the proposed ensemble deep learning technique such as accuracy, precision, recall, FDR, and FOR [34]. The mathematical expressions of accuracy, precision,

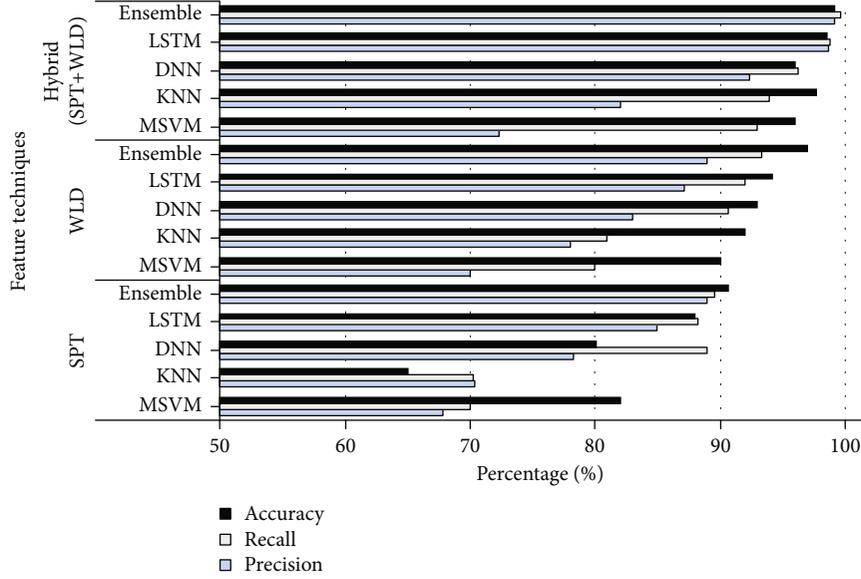


FIGURE 9: Graphical comparison of the proposed ensemble deep learning technique on MIO-TCD in terms of precision, recall, and accuracy.

recall, FDR, and FOR are represented in

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \times 100, \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100, \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \quad (11)$$

$$\text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TP}} \times 100, \quad (12)$$

$$\text{FOR} = \frac{\text{FN}}{\text{TN} + \text{FN}} \times 100. \quad (13)$$

**4.1. Quantitative Analysis on BIT Vehicle Dataset.** Here, the proposed ensemble deep learning technique performance is investigated using the BIT Vehicle Dataset, which consists of six vehicle types such as sedan, microbus, SUV, minivan, bus, and truck. In this scenario, performance analysis is carried out by different classifiers such as the Long Short-Term Memory (LSTM) network, Multisupport Vector Machine (MSVM),  $K$ -Nearest Neighbor (KNN), Deep Neural Network (DNN), and the ensemble deep learning technique with individual and hybrid feature extraction. By inspecting Table 2, the combination ensemble deep learning technique with the hybrid feature extraction achieved significant performance in vehicle type classification compared to other combinations by means of precision, recall, and accuracy. In Table 2, the proposed ensemble deep learning technique achieved a maximum recall of 99.72%, a precision of 98.24%, and an accuracy of 99.28% on the BIT Vehicle Dataset. The graphical comparison of the proposed ensemble deep learning technique on the BIT Vehicle Dataset in terms of precision, recall, and accuracy is denoted in Figure 7.

Similarly, in Table 3, the proposed ensemble deep learning technique performance is validated in terms of FDR

TABLE 5: Performance analysis of the proposed ensemble deep learning technique on MIO-TCD in terms of FDR and FOR.

| Feature extraction | Classifier | FDR (%) | FOR (%) |
|--------------------|------------|---------|---------|
| SPT                | MSVM       | 22      | 18.03   |
|                    | KNN        | 12      | 17      |
|                    | DNN        | 9.72    | 10.74   |
|                    | LSTM       | 8       | 3.20    |
|                    | Ensemble   | 6       | 1.29    |
| WLD                | MSVM       | 11.91   | 9.5     |
|                    | KNN        | 6.03    | 5.93    |
|                    | DNN        | 3       | 3.07    |
|                    | LSTM       | 2.09    | 1.02    |
|                    | Ensemble   | 1.72    | 0.86    |
| Hybrid (SPT + WLD) | MSVM       | 4       | 3.29    |
|                    | KNN        | 1.20    | 1.95    |
|                    | DNN        | 0.98    | 0.83    |
|                    | LSTM       | 0.79    | 0.35    |
|                    | Ensemble   | 0.44    | 0.32    |

and FOR on the BIT Vehicle Dataset. By inspecting Table 3, the combination of the ensemble deep learning technique with hybrid feature extraction achieved a minimum FDR of 3.92% and an FOR of 1.90% which are effective compared to other combinations in vehicle type classification. In the BIT Vehicle Dataset, 7,880 vehicle images are utilized for training, and 1,970 vehicle images are utilized for testing. The graphical comparison of the proposed ensemble deep learning technique on the BIT Vehicle Dataset in terms of FDR and FOR is represented in Figure 8. In addition to this, the running time of the proposed ensemble deep learning technique on the BIT Vehicle Dataset is 1.6 seconds per frame.

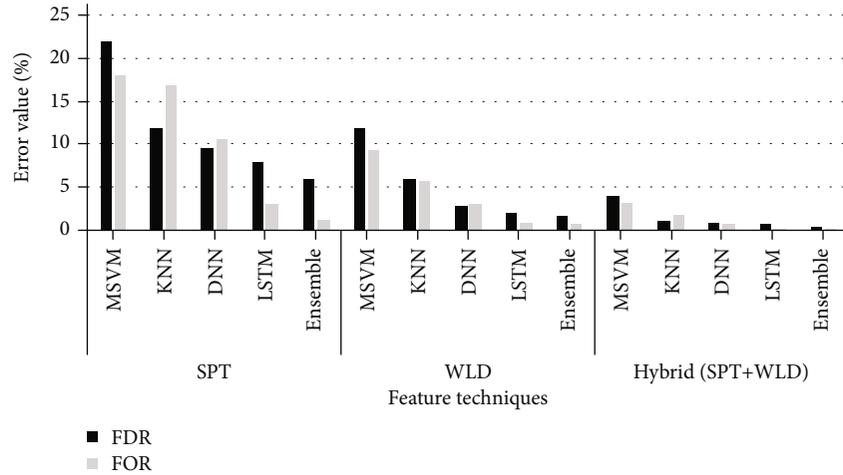


FIGURE 10: Graphical comparison of the proposed ensemble deep learning technique on MIO-TCD in terms of FDR and FOR.

TABLE 6: Comparative analysis between the proposed and the existing techniques.

| Algorithm                              | Dataset     | Precision (%) | Recall (%) | Accuracy (%) |
|--|-------------|---------------|------------|--------------|
| GAN-based deep ensemble technique [13] | MIO-TCD     | 96.41         | —          | —            |
| Tiny YOLO with SVM [15]                | BIT Vehicle | 97.90         | 99.60      | —            |
| Semisupervised CNN model [19]          | BIT Vehicle | —             | —          | 88.11        |
| PCN with softmax classifier [21]       | BIT Vehicle | —             | —          | 88.52        |
| TC-SF-CNNLS [22]                       | BIT Vehicle | 90.52         | 90.41      | 93.80        |
|  | MIO-TCD     | 99.12         | 99.69      | 99.13        |
| Ensemble deep learning technique       | BIT Vehicle | 98.24         | 99.72      | 99.28        |
|  | Combined    | 99.27         | 99.77      | 99.32        |

4.2. *Quantitative Analysis on MIO-TCD.* Here, MIO-TCD is used to validate the efficiency of the proposed ensemble deep learning technique in terms of precision, recall, accuracy, FDR, and FOR. MIO-TCD includes 648,959 images with 11 vehicle classes like single-unit truck, pickup truck, nonmotorized vehicle, car, pedestrian, articulated truck, background, motorcycle, bicycle, work van, and bus. In this scenario, 80% of the images are used for training, and 20% of the vehicle images are used for testing. By investigating Table 4, the combination of ensemble deep learning technique with hybrid feature extraction achieved a maximum precision of 99.12%, a recall value of 99.69%, and an accuracy of 99.13% on MIO-TCD. In this article, hybrid feature extraction significantly detects the statistical interactions and extracts the active feature vectors from the vehicle images. The graphical comparison of the proposed ensemble deep learning technique on MIO-TCD by means of precision, recall, and accuracy is denoted in Figure 9.

In Table 5, the proposed ensemble deep learning technique achieved a minimum FDR value of 0.44 and an FOR value of 0.32 compared to other combinations on MIO-TCD. In this study, the ensemble deep learning technique effectively maximizes the percentage of correct predictions that reduces the misclassification in dominant and minority vehicle categories. Graphical comparison of the ensemble deep learning technique on MIO-TCD by means of FDR

and FOR is stated in Figure 10. Similarly, the running time of the proposed ensemble deep learning technique on MIO-TCD is 1.44 seconds per frame.

4.3. *Comparative Analysis.* The comparative analysis between the proposed and existing techniques are given in Table 6. Liu et al. [13] introduced a deep learning technique, namely, GANs, for classifying vehicles in traffic surveillance videos. Extensive experiments showed that the developed GANs achieved 96.41% precision on MIO-TCD. Additionally, Şentaş et al. [15] utilized the Tiny YOLO with the SVM classification technique for vehicle detection and classification. The simulation outcome showed that the developed model obtained 97.9% precision and 99.6% recall on the BIT Vehicle Dataset in vehicle type classification. Dong et al. [19] developed a novel semisupervised CNN model for vehicle type classification. The semisupervised CNN model used a sparse Laplacian filter to extract rich and discriminative features of the vehicles. The features learned by the CNN model were discriminative which works effectively in complex scenes. In the experimental phase, the developed semisupervised CNN model achieved 88.11% accuracy on the BIT Vehicle Dataset.

Soon et al. [21] developed a semisupervised model, namely, PCN for vehicle type classification. The developed PCN model utilized convolutional filters to extract

hierarchical and discriminative features of the vehicles for better classification. The simulation results showed that the developed PCN model with the softmax classifier achieved 88.52% classification accuracy, and the PCN model with the SVM classifier achieved 88.39% accuracy on the BIT Vehicle Dataset. Additionally, Awang et al. [22] developed the TC-SF-CNNLS approach for vehicle type classification. In the experimental phase, the developed approach performance was validated on the BIT Vehicle Dataset in terms of recall, precision, and accuracy. The developed TC-SF-CNNLS approach achieved 93.8% accuracy by classifying the vehicle types like truck, minivan, bus, passenger, taxi, car, and SUV.

**4.4. Discussion.** As previously discussed, feature extraction and classification are the integral parts of vehicle type classification. In this research study, hybrid (SPT + WLD) descriptors are used to extract active feature vectors from the vehicle images that speed up the training process, reduce overfitting risk, and improve the data visualization ability. Hence, the effect of hybrid feature extraction in vehicle type classification is given in Tables 2, 3, 4, and 5. Additionally, a new ensemble deep learning technique is proposed in this research paper for learning the original dataset in order to classify unknown data. In most of the existing research works, an individual classifier causes bias in terms of a fixed set of parameters, where such bias is reduced by developing an ensemble classifier. In contrast, the performance of the ensemble classifier completely depends on the accuracy of the constituent classifiers, which has stronger generalization ability than the individual classifiers.

## 5. Conclusion

In this article, an ensemble deep learning technique is proposed for vehicle type classification which was primarily used for traffic surveillance systems. Nowadays, video surveillance has been utilised for additional reasons across the world during the COVID-19 pandemic. Our application uses a deep learning approach that consists of two major phases in vehicle type classification such as feature extraction and classification. In this research, hybrid (SPT+WLD) feature descriptors are applied to extract active feature vectors that reduce training time, improve classification accuracy, and diminish overfitting problems in the ensemble deep learning technique. In this study, the ensemble deep learning technique classifies 11 classes in MIO-TCD and 6 classes in the BIT Vehicle Dataset. In Experimental Results and Discussion, the ensemble deep learning technique achieved better performance in vehicle type classification compared to other classification techniques in terms of precision, recall, accuracy, FDR, and FOR. Compared to the existing benchmark techniques like the GAN-based deep ensemble technique, the Tiny YOLO with SVM, the semisupervised CNN model, the TC-SF-CNNLS, and the PCN with a softmax classifier, the proposed technique showed a maximum of 11.17% improvement in vehicle type classification by means of classification accuracy. In future work, a clustering-based segmentation algorithm is included in the proposed technique for improving vehicle type detection and classification. In

addition to this, three-dimensional modelling, vehicle tracking, and occlusion handling are given emphasis for an effective intelligent transportation system.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This publication was created thanks to the support from the Operational Program Integrated Infrastructure for the Project: Identification and possibilities of implementation of new technological measures in transport to achieve safe mobility during a pandemic caused by COVID-19 (ITMS code: 313011AUX5), cofinanced by the European Regional Development Fund.

## References

- [1] S. Yu, Y. Wu, W. Li, Z. Song, and W. Zeng, "A model for fine-grained vehicle classification based on deep learning," *Neuro-computing*, vol. 257, pp. 97–103, 2017.
- [2] H. Asaidi, A. Aarab, and M. Bellouki, "Shadow elimination and vehicles classification approaches in traffic video surveillance context," *Journal of Visual Languages & Computing*, vol. 25, no. 4, pp. 333–345, 2014.
- [3] G. Yan, M. Yu, Y. Yu, and L. Fan, "Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification," *Optik*, vol. 127, no. 19, pp. 7941–7951, 2016.
- [4] W. Sun, G. Zhang, X. Zhang, X. Zhang, and N. Ge, "Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy," *Multimedia Tools and Applications*, pp. 1–14, 2020.
- [5] H. Kim, "Multiple vehicle tracking and classification system with a convolutional neural network," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2019.
- [6] Y. Nam and Y. C. Nam, "Vehicle classification based on images from visible light and thermal cameras," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, 2018.
- [7] C. R. Kumar and R. Anuradha, "Feature selection and classification methods for vehicle tracking and detection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4269–4279, 2020.
- [8] X. Wang, X. Chen, and Y. Wang, "Small vehicle classification in the wild using generative adversarial network," *Neural Computing and Applications*, vol. 33, no. 10, pp. 5369–5379, 2020.
- [9] B. Zhang, "Reliable classification of vehicle types based on cascade classifier ensembles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 322–332, 2012.
- [10] N. Shvai, A. Hasnat, A. Meicler, and A. Nakib, "Accurate classification for automatic vehicle-type recognition based on ensemble classifiers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1288–1297, 2019.

- [11] C. Iwendi, S. Khan, J. H. Anajemba, M. Mittal, M. Alenezi, and M. Alazab, "The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems," *Sensors*, vol. 20, no. 9, article 2559, 2020.
- [12] C. Iwendi, G. Srivastava, S. Khan, and P. K. R. Maddikunta, "Cyberbullying detection solutions based on deep learning architectures," *Multimedia Systems*, pp. 1–14, 2020.
- [13] W. Liu, Z. Luo, and S. Li, "Improving deep ensemble vehicle classification by using selected adversarial samples," *Knowledge-Based Systems*, vol. 160, pp. 167–175, 2018.
- [14] H. Fu, H. Ma, Y. Liu, and D. Lu, "A vehicle classification system based on hierarchical multi-SVMs in crowded traffic scenes," *Neurocomputing*, vol. 211, pp. 182–190, 2016.
- [15] A. Şentaş, İ. Tashiev, F. Küçükayvaz et al., "Performance evaluation of support vector machine and convolutional neural network algorithms in real-time vehicle type and color classification," *Evolutionary Intelligence*, vol. 13, no. 1, pp. 83–91, 2020.
- [16] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang, "Real-time vehicle type classification with deep convolutional neural networks," *Journal of Real-Time Image Processing*, vol. 16, no. 1, pp. 5–14, 2019.
- [17] L. Zhuo, L. Jiang, Z. Zhu, J. Li, J. Zhang, and H. Long, "Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks," *Machine Vision and Applications*, vol. 28, no. 7, pp. 793–802, 2017.
- [18] V. Murugan and V. R. Vijaykumar, "Automatic moving vehicle detection and classification based on artificial neural fuzzy inference system," *Wireless Personal Communications*, vol. 100, no. 3, pp. 745–766, 2018.
- [19] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.
- [20] M. A. Hedeya, A. H. Eid, and R. F. Abdel-Kader, "A super-learner ensemble of deep networks for vehicle-type classification," *IEEE Access*, vol. 8, pp. 98266–98280, 2020.
- [21] F. C. Soon, H. Y. Khaw, J. H. Chuah, and J. Kanesan, "Semisupervised PCA convolutional network for vehicle type classification," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8267–8277, 2020.
- [22] S. Awang, N. M. A. N. Azmi, and M. A. Rahman, "Vehicle type classification using an enhanced sparse-filtered convolutional neural network with layer-skipping strategy," *IEEE Access*, vol. 8, pp. 14265–14277, 2020.
- [23] N. Nasaruddin, K. Muchtar, and A. Afdhal, "A lightweight moving vehicle classification system through attention-based method and deep learning," *IEEE Access*, vol. 7, pp. 157564–157573, 2019.
- [24] J. Lian, J. Zhang, T. Gan, and S. Jiang, "Vehicle type classification using hierarchical classifiers," *Journal of Physics: Conference Series*, vol. 1069, no. 1, article 012099, 2018.
- [25] J. Nedoma, M. Kostelansky, M. Fridrich et al., "Fiber optic phase-based sensor for detection of axles and wheels of tram vehicles," *Communications - Scientific letters of the University of Zilina*, vol. 22, no. 3, pp. 119–127, 2020.
- [26] M. N. Roecker, Y. M. Costa, J. L. Almeida, and G. H. Matsushita, "Automatic vehicle type classification with convolutional neural network," in *25th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 1–5, Maribor, Slovenia, 2018.
- [27] Z. Luo, F. Branchaud-Charron, C. Lemaire et al., "MIO-TCD: a new benchmark dataset for vehicle classification and localization," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5129–5141, 2018.
- [28] Y. Zhu and C. Huang, "An adaptive histogram equalization algorithm on the image gray level mapping," *Physics Procedia*, vol. 25, pp. 601–608, 2012.
- [29] H. Bi, H. Tang, G. Yang, H. Shu, and J. L. Dillenseger, "Accurate image segmentation using Gaussian mixture model with saliency map," *Pattern Analysis and Applications*, vol. 21, no. 3, pp. 869–878, 2018.
- [30] V. Anand, D. Pushp, R. Raj, and K. Das, "Gaussian Mixture Model (GMM) based dynamic object detection and tracking," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 1365–1371, Atlanta, GA, USA, 2019.
- [31] A. Alelaiwi, W. Abdul, M. S. Dewan, M. Migdadi, and G. Muhammad, "Steerable pyramid transform and local binary pattern based robust face recognition for e-health secured login," *Computers & Electrical Engineering*, vol. 53, pp. 435–443, 2016.
- [32] G. Muhammad, M. H. Al-Hammadi, M. Hussain, and G. Bebis, "Image forgery detection using steerable pyramid transform and local binary pattern," *Machine Vision and Applications*, vol. 25, no. 4, pp. 985–995, 2014.
- [33] X. H. Han, Y. W. Chen, and G. Xu, "High-order statistics of Weber local descriptors for image representation," *IEEE transactions on cybernetics*, vol. 45, no. 6, pp. 1180–1193, 2015.
- [34] J. Frnda, M. Durica, M. Savrasovs, P. Fournier-Viger, and J. C. Lin, "QoS to QoE mapping function for Iptv quality assessment based on Kohonen map: a pilot study," *Transport and Telecommunication Journal*, vol. 21, no. 3, pp. 181–190, 2020.
- [35] D. Zhang, Q. Li, G. Yang, L. Li, and X. Sun, "Detection of image seam carving by using weber local descriptor and local binary patterns," *Journal of information security and applications*, vol. 36, pp. 135–144, 2017.