



Article

Exploiting Remote Sensing Imagery for Vehicle Detection and Classification Using an Artificial Intelligence Technique

Masoud Alajmi ¹, Hayam Alamro ^{2,*}, Fuad Al-Mutiri ³, Mohammed Aljebreen ⁴, Kamal M. Othman ⁵ and Ahmed Sayed ⁶

- ¹ Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
- ² Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
- ³ Department of Mathematics, Faculty of Sciences and Arts, King Khalid University, Abha 63311, Saudi Arabia
- ⁴ Department of Computer Science, Community College, King Saud University, P.O. Box 28095, Riyadh 11437, Saudi Arabia
- ⁵ Department of Electrical Engineering, Umm Al-Qura University, Makkah 21955, Saudi Arabia
- ⁶ Research Center, Future University in Egypt, New Cairo 11835, Egypt
- * Correspondence: hmamro@pnu.edu.sa

Abstract: Remote sensing imagery involves capturing and examining details about the Earth's surface from a distance, often using satellites, drones, or other aerial platforms. It offers useful data with which to monitor and understand different phenomena on Earth. Vehicle detection and classification play a crucial role in various applications, including traffic monitoring, urban planning, and environmental analysis. Deep learning, specifically convolutional neural networks (CNNs), has revolutionized vehicle detection in remote sensing. This study designs an improved Chimp optimization algorithm with a DL-based vehicle detection and classification (ICOA-DLVDC) technique on RSI. The presented ICOA-DLVDC technique involves two phases: object detection and classification. For vehicle detection, the ICOA-DLVDC technique applies the EfficientDet model. Next, the detected objects can be classified by using the sparse autoencoder (SAE) model. To optimize the SAE's hyperparameters effectively, we introduce an ICOA which streamlines the parameter tuning process, accelerating convergence and enhancing the overall performance of the SAE classifier. An extensive set of experiments has been conducted to highlight the improved vehicle classification outcomes of the ICOA-DLVDC technique. The simulation values demonstrated the remarkable performance of the ICOA-DLVDC approach compared to other recent techniques, with a maximum accuracy of 99.70% and 99.50% on the VEDAI dataset and ISPRS Postdam dataset, respectively.

Keywords: artificial intelligence; object detector; computer vision; remote sensing; target detection; deep learning



Citation: Alajmi, M.; Alamro, H.; Al-Mutiri, F.; Aljebreen, M.; Othman, K.M.; Sayed, A. Exploiting Remote Sensing Imagery for Vehicle Detection and Classification Using an Artificial Intelligence Technique. *Remote Sens.* **2023**, *15*, 4600. <https://doi.org/10.3390/rs15184600>

Academic Editors: Yanni Dong, Xiaochen Yang and Qian Du

Received: 30 July 2023

Revised: 2 September 2023

Accepted: 7 September 2023

Published: 19 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Remote sensing target detection is used to mark the objects of interest in remote sensing imagery (RSIs) and to predict the location and type of these targets [1]. Based on the perspective of the Earth vision platform, the object strength in the aviation images always appears in a random direction and the target is only concentrated in the conventional detection dataset [2]. The object detection (OD) technique is used to detect samples of semantic objects of specific classes (for example, humans, birds, or airplanes) in digital videos and images. Small target detection has often become a hot and challenging field in target detection tasks. Transport planning, environmental management, military, and disaster control are crucial applications of RSIs [3]. Moreover, vehicles in RSIs, as a special class (whether transportation, civilian, or military), are of particular significance and increasingly difficult. First, vehicle targets in RSIs are fewer than twenty pixels or even ten

pixels in the target detection task; the smaller target is generally a target that has fewer than thirty pixels in an image [4]. Next, weather and environment images, including shadow, building, and atmospheric occlusions, and other factors, including similar colors amongst vehicles, dissimilar sizes of vehicle targets in similar images, different overhead views, and their environments, can all lead to the poor detection accuracy of car targets [5].

Vehicle detection in RSI aims to identify each instance of a vehicle [6]. In previous approaches, researchers often developed and extracted vehicle features manually and then classified them to attain vehicle detection [7]. The fundamental objective is to extract vehicle features and utilize traditional machine learning (ML) techniques for classification. Generally, the integration channel features, the scale-invariant feature transform (SIFT), and the histogram of the oriented gradient (HOG) are the features utilized in the detection process. [8]. The approaches utilized for classification are intersection kernel support vectors (IKSVM), AdaBoost, SVM, and so on. However, conventional target detection techniques pay greater consideration to completing the tasks of RSI vehicle detection, and it is challenging to balance speed and accuracy. In contrast to the tremendous growth of deep learning (DL) techniques, there is a big difference in the efficiency and accuracy of detection [9]. Network models based on DL approaches can map complex nonlinear relationships and extract richer features. Two categories of target detection network models are continually formed and optimized due to the development of hardware technology and enormous data: single-stage networks (i.e., SSD and YOLOv3) and two-stage networks (i.e., cascade RCNN and fast RCNN) [10].

This study designs an improved chimp optimization algorithm with a DL-based vehicle detection and classification (ICOA-DLVDC) technique on RSIs. The presented ICOA-DLVDC technique focuses on the utilization of the DL model for the detection of vehicles on the RSI with a hyperparameter tuning strategy. First, the ICOA-DLVDC method exploits the EfficientDet model for OD purposes. Next, the detected objects are classified using the sparse autoencoder (SAE) model. Finally, the hyperparameter tuning of the SAE method can be chosen by ICOA. An extensive set of experiments has been conducted to highlight the improved vehicle classification outcomes of the ICOA-DLVDC technique. In short, the key contributions of the paper are listed as follows.

- An intelligent ICOA-DLVDC technique comprising an EfficientDet object detector, SAE classification, and ICOA-based hyperparameter tuning for RSI has been presented, and to the best of our knowledge, the proposed model will not be found in the literature;
- SAE is able to learn informative and discriminative features with the reduction of the data dimensionality, which is helpful in handling large and complex remote sensing datasets;
- The integration of the EfficientNet object detector with SAE classification can significantly accomplish enhanced generalization and adaptability over various RSI datasets;
- Hyperparameter optimization of the SAE model using the ICOA algorithm using cross-validation helps to boost the predictive outcome of the ICOA-DLVDC model for unseen data.

The rest of the paper is organized as follows: Section 2 provides the related works and Section 3 offers the proposed model. Then, Section 4 gives the result analysis and Section 5 concludes the paper.

2. Related Works

Ahmed et al. [11] designed an IoT-assisted smart surveillance solution for multi-OD using segmentation. In particular, the study proposes the utilization of DL, IoT, and collaborative drones to enhance surveillance applications in smart cities. The study proposed an AI-based technique using a DL-based pyramid scene parsing network (PSPNet) for multiple-object segmentation and applied an aerial drone dataset. The authors in [12] developed a new one-phase OD technique termed MDCT based on a transformer block and multi-kernel dilated convolution (MDC) blocks. Initially, in the single-phase OD technique, a feature enhancement model, the MDC block, was introduced. Next, a transformer block

was incorporated into the neck network of the single-phase OD technique. Finally, a depth-wise convolutional layer was incorporated into the MDC block for reducing the computation cost. Qiu, Bai, and Chen [13] designed a new technique called YOLO-GNS for vehicle detection. First, the SSH (single-stage headless) model was devised to facilitate the detection of smaller objects and optimize the feature extraction.

The authors in [14] developed an OD technique based on YOLOv5 for aerial RSI, named KCFS-YOLOv5. The K-means++ algorithm was used for optimizing the initial cluster point to attain the suitable anchor box. Coordinate attention (CA) was embedded with the backbone network of YOLO_v5 to develop the Bi-directional FPN (BiFPN) architecture. Ye et al. [15] designed a convolution network using an adaptive attention fusion module (AAFm). Initially, the stitcher was used for developing one image with objects of different scales according to the features of object distribution in the dataset. Moreover, a spatial attention module was developed, and the semantic data of the feature map was attained. Xiaolin et al. [16] presented an S²ANET-SR model based on the S²ANET network. The original and reduced images were fed to the detection model; later, a super-resolution enhancement model for the reduced images was developed for enhancing the feature extraction of smaller objects, and the texture matching loss and perceptual loss were introduced as supervision.

Javadi et al. [17] investigated the ability of 3D feature maps for enhancing the accuracy of DNN for the recognition of vehicles. First, they introduced a DNN by using YOLOv3 with the base network, involving DenseNet201, DarkNet53, SqueezeNet, and MobileNetv2. Next, 3D depth maps were produced. Later, FCNN was trained on 3D feature mapping. Wu et al. [18] introduced a GCWNet (global context-weaving network) for object recognition in RSIs. Then, two novel modules were introduced for refinement and feature extraction.

Several automated vehicle detection and classification models have been presented in the literature. Despite the benefits of the earlier studies, it is still required to boost the vehicle classification performance. Because of the continual deepening of the model, the number of parameters of DL models also increases quickly, which results in model overfitting. At the same time, different hyperparameters have a significant impact on the efficiency of the CNN model. Particularly, hyperparameters such as epoch count, batch size, and learning rate selection are essential to attaining an effective outcome. Since the trial-and-error method for hyperparameter tuning is a tedious and erroneous process, metaheuristic algorithms can be applied. Therefore, in this work, we employ the ICOA algorithm for the parameter selection of the SAE model.

3. The Proposed Model

In this work, the ICOA-DLVDC technique is established for automated vehicle detection and classification on RSI. In the proposed ICOA-DLVDC technique, a DL-based object detector and classifier are applied. Figure 1 shows the working flow of the ICOA-DLVDC algorithm. The presented ICOA-DLVDC technique involves two phases: EfficientDet-based object detector and ICOA with SAE-based classification. Initially, the input images are passed into the EfficientDet model for the detection of vehicles. Next, the detected objects are classified by the use of SAE model. Finally, the ICOA is applied for the hyperparameter tuning of the SAE model.

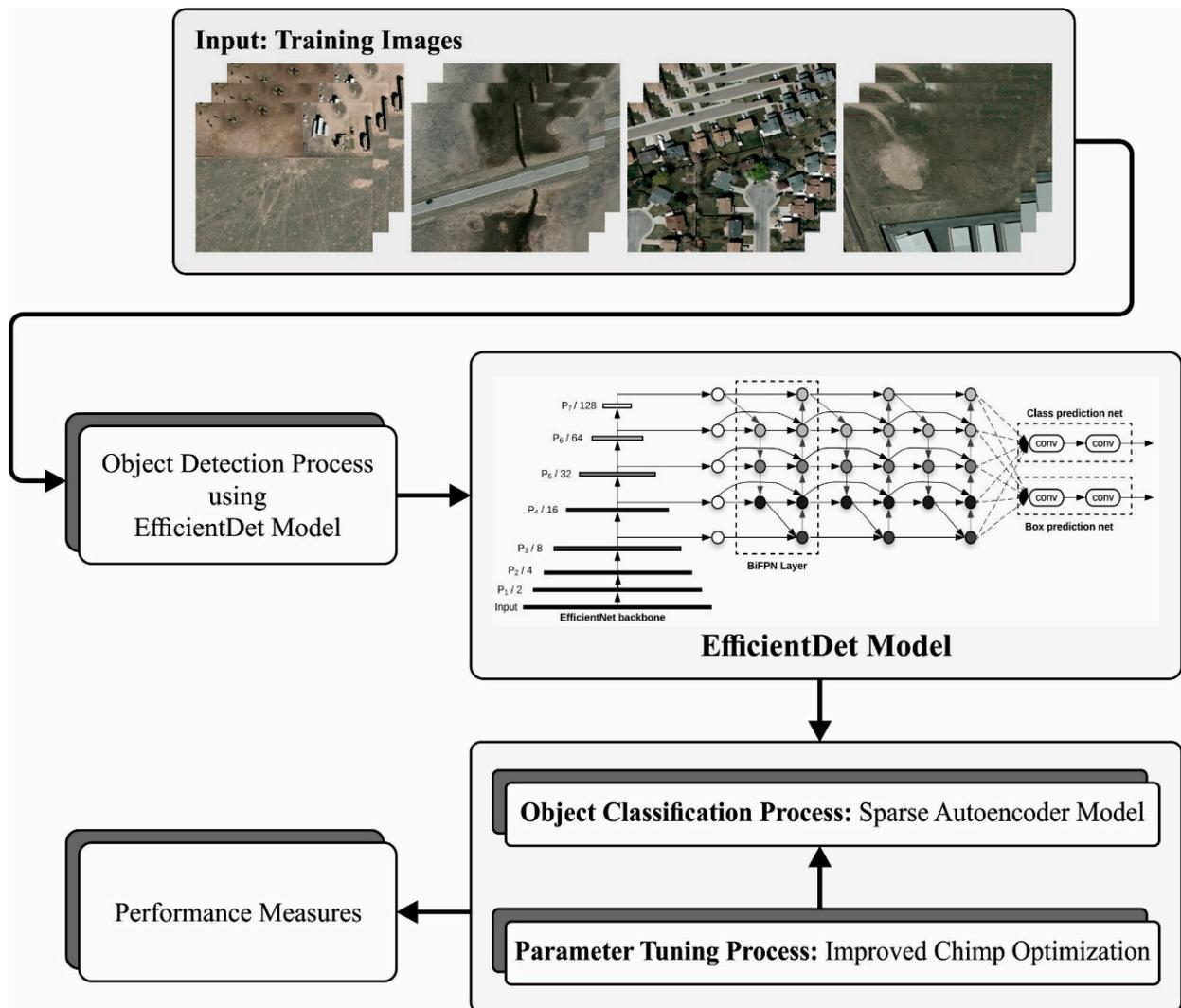


Figure 1. Working flow of the ICOA-DLVDC approach.

3.1. Stage I: Object Detector

The EfficientDet model is used to detect the objects (i.e., vehicles) in the RSI. For combining features with a top-down direction, a conventional approach, named Feature pyramid network (FPN), was used [19]. The PANet (path aggregation network) allows for the forward and reverse flows of feature fusion from low to high resolution. Lastly, the Efficient-Det architecture stacks this BiFPN block. Scaling issues were addressed for resizing the weighted BiFPN, backbone, input quality of the image, and class/box. The EfficientDet model was validated on 100,000 photographs. The network automatically scales from EfficientNetB0 to EfficientNetB6; therefore, the quantity of BiFPN stacks might affect the depth and width of the networks. In most instances, EfficientDet outperforms other OD techniques.

$$\underset{m}{\text{maximize}} \text{ACC}(m) \cdot \left[\frac{\text{FLOPS}(m)}{T} \right]^w, \quad (1)$$

where T refers to the target of FLOPS $\text{ACC}(m)$ and is defined as the accuracy of the algorithm m ; $\text{FLOPS}(m)$ denotes the FLOPS (floating point operations per second) of the algorithm m ; and $w = -0.07$ denotes the hyperparameter that controls the exchange amongst FLOPS and accuracy. The EfficientNet seems to be a solid foundation.

As a feature network, the BiFPN function accepts levels 3–7 components (P_3, P_4, P_5, P_6, P_7) from the EfficientNet (backbone network).

$$W_{BiFPN} = 64 \cdot (1.35^\varphi), D_{BiFPN} = 3 + \varphi \tag{2}$$

The width of BiFPN was exponentially expanded because the levels of BiFPN should be transformed into small integers, but it gradually enhances the depth. The depth was continuously increased; however, the width was retained at the accurate levels of BiFPN and formulated as follows:

$$D_{box} = D_{class} = 3 + \left\lceil \frac{\varphi}{3} \right\rceil. \tag{3}$$

Considering that BiFPN exploits feature levels 3–7, the input resolution should be dividable by $2^7 = 128$, which implies that it linearly improves the resolution by using the following equation:

$$R_{input} = 512 + (\varphi)(128). \tag{4}$$

Generally, a compound scaling method for OD was introduced, which exploits the φ compound coefficient to enhance each feature of the input image resolution and the backbone, featured, and class/box networks.

The Efficient-Det structure is based on the backbone network EfficientNet42. The class/box net layers and feature network BiFPN are repeated to constitute resource constraints of different magnitudes.

3.2. Stage II: Classification Model

Once the objects are detected, the SAE model is utilized for classification purposes. AE has the potential to duplicate (without learning the hidden representation) the input dataset in the output layer because of the hidden representation, $L1(x)$, and to maximize the mutual information of the input dataset, x [20]. Therefore, the application of sparsity was used to constrain AE in order to learn the hidden representation for the input dataset. Figure 2 demonstrates the infrastructure of SAE.

The hidden unit was constrained to have a small pre-determined activation value, z . The calculated sparsity parameter, \check{z} , for j hidden units, was attained via Equation (5):

$$\check{z}_j = \frac{1}{N} \sum_{n=1}^N o_{jx_n}. \tag{5}$$

In Equation (5), N indicates the number of training samples, O_j denotes the activation (or output) of the hidden module, and x_n shows the training sample with index n . The sparsity is used to limit the j hidden units so that $\check{z} = z.s$. The KL (Kullback–Leibler) divergence is used for measuring the distribution deviation z from \check{z} and thus enhances the algorithm.

$$KL(z||\check{z}_j) = z \log \frac{z}{\check{z}_j} + (1 - z) \log \frac{(1 - z)}{(1 - \check{z}_j)} \tag{6}$$

Note that $(z||\check{z}) = 0$ for $\check{z} = z$. KL divergence is added to the MSE for the minimization of cost. Thus, the cost function $C(x, y; \theta)$ is formulated in (7):

$$C(x, y, \theta) = \arg \min \sum_{=n1}^N \left\{ \sum_{i=1}^u (x_i - y_i)^2 + \gamma \left(\sum_{j=1}^h KL(z||\check{z}_j) \right) \right\}. \tag{7}$$

The SAE with the convolution operation can be represented as a sparse CAE (SCAE).

The ICOA is used to finetune the hyperparameter value of the SAE technique. COA is derived from the predatory behaviors of the chimp population [21]. Attacker, driver, barrier, and chaser are four different groups based on their behaviors during hunting. Chasing and attacking prey are the two different hunting methods of chimps, which corresponds

to the exploration and development phases. Each chimpanzee participating in predation randomly changes its location to move closer to the prey as follows:

$$D = \left| c \cdot x_{prey}(t) - m \cdot x_{chimp}(t) \right| \quad (8)$$

$$x_{chimp}(t+1) = x_{prey}(t) - a \cdot d, \quad (9)$$

where x_{chimp} shows the chimp's location vector, D denotes the distance between the prey and the chimps, x_{prey} indicates the prey's location vector t signifies the existing amount of iteration, and a , m , and c represent coefficient vectors.

$$a = 2 \cdot f \cdot r_1 - f \quad (10)$$

$$c = 2 \cdot r_2 \quad (11)$$

$$m = \text{Chaotic_value} \quad (12)$$

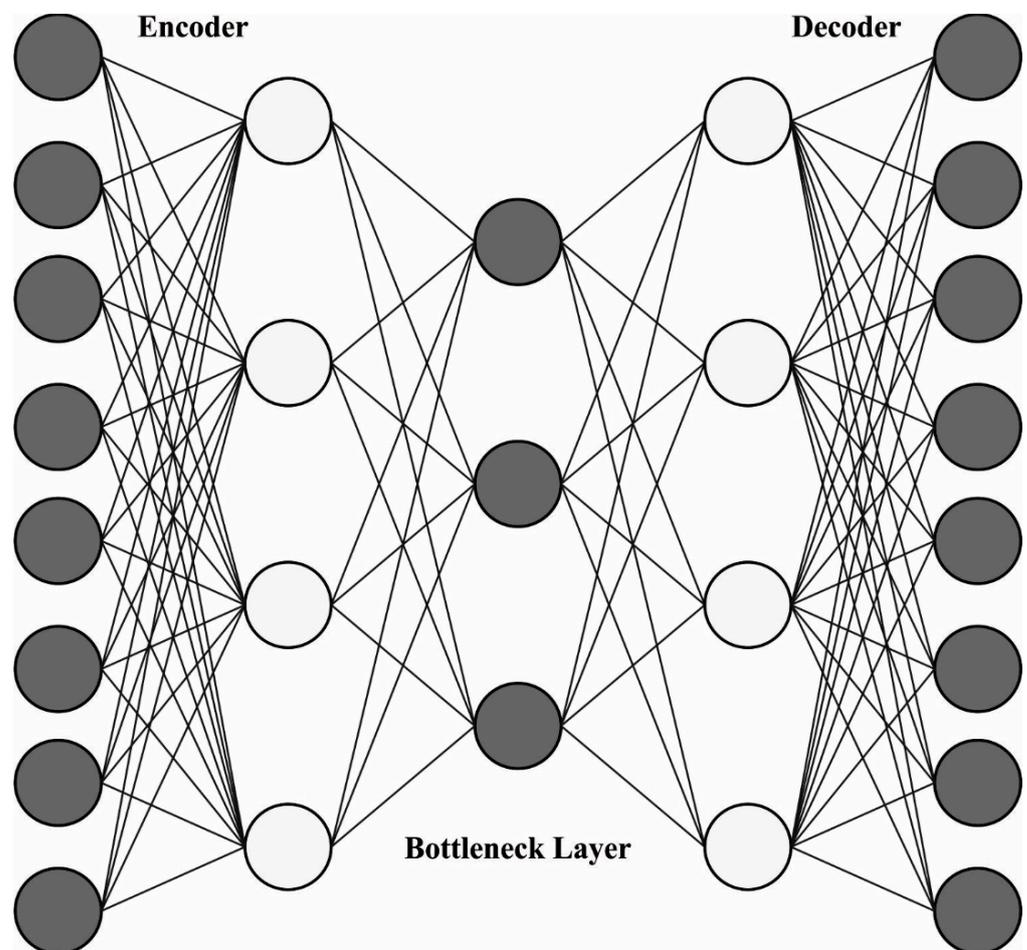


Figure 2. SAE structure.

During the iteration, the value of f reduces from 2.5 to 0, r_1 and r_2 denotes the random vector within $[0, 1]$, and m refers to the chaotic vector computed based on the chaotic map.

The present optimum solution (the first attacker), barrier, chaser, and driver are informed about the target position, and other members are forced to update the locations based on the optimum location of chimps.

$$\begin{cases} d_{Attacker} = |c_1 x_{Attacker} - m_1 x| \\ d_{Barrier} = |c_2 x_{Barrier} - m_2 x| \\ d_{Chaser} = |c_3 x_{Chaser} - m_3 x| \\ d_{Driver} = |c_4 x_{Driver} - m_4 x| \end{cases} \quad (13)$$

$$\begin{cases} V_1 = x_{Attacker} - a_1(d_{Attacker}) \\ V_2 = x_{Barrier} - a_2(d_{Barrier}) \\ V_3 = x_{Chaser} - a_3(d_{Chaser}) \\ V_4 = x_{Driver} - a_4(d_{Driver}) \end{cases} \quad (14)$$

$$x(t + 1) = \frac{V_1 + V_2 + V_3 + V_4}{4}, \quad (15)$$

where $d_{Attacker}$, $d_{Barrier}$, d_{Chaser} , and d_{Driver} denote the distance between 4 kinds of chimps and their target in the existing group; $x_{Attacker}$, $x_{Barrier}$, x_{Chaser} , and x_{Driver} indicate the location vector relative to the prey; V_1 , V_2 , V_3 , and V_4 characterize their location update vector; $x(t + 1)$ shows the location of $t + 1$ generation chimps; and $a_1 \sim a_4$, $m_1 \sim m_4$, and $c_1 \sim c_4$ denote the coefficient vector. The chimps release hunting responsibility after food satisfaction and scramble to obtain food. These chaotic behaviors assist in preventing the model from becoming trapped in local optima.

$$x_{chimp}(t + 1) = \begin{cases} x_{prey}(t) - a \cdot d, & \text{if } \mu < 0.5 \\ \text{Chaotic_value}, & \text{if } \mu \geq 0.5 \end{cases} \quad (16)$$

In Equation (16), μ represents a randomly generated value within [0,1] and *Chaotic_value* shows the chaotic mapping.

In ICOA, reverse learning is used to attain the reverse solution of an individual and, later, retained the individual with the higher fitness value to enhance the individual quality of COA and the population diversity. The refraction of light was combined with reverse learning. The refraction angle takes place while attaining the reverse location of the existing individuals, thereby optimizing the generalization capability of the algorithm and extending the search range of the individual. The upper and lower boundaries of the search region are represented as u and l ; correspondingly, $\chi \in [u, l]$ and O represent the midpoint of the $[u, l]$ interval.

$$\begin{cases} \sin\theta_1 = ((u + l)/2 - x) / |PO| \\ \sin\theta_2 = (x - (u + l)/2) / |OQ| \end{cases} \quad (17)$$

$$\eta = \frac{\sin\theta_1}{\sin\theta_2}, \quad (18)$$

where η signifies the refractive index. Consider $k = |PO|/|OQ|$; thenm the refraction reverse learning solution was defined:

$$x' = \frac{u + l}{2} + \frac{u + l}{2k\eta} - \frac{x}{k\eta}. \quad (19)$$

The common form of the inverse solution was attained by expanding Equation (19) to n -dimensional space.

$$x'_i = \frac{u_i + l_i}{2} + \frac{u_i + l_i}{2k\eta} - \frac{x_i}{k\eta} \quad (20)$$

In Equation (20), u_i and l_i denote the i^{th} dimension of the upper and lower boundaries, respectively. Thus, the study introduced hyper-parametric ω . According to dissimilar

iteration processes, it adaptively adjusts to improve the randomness of the solution enhance the capability of the model with respect to escaping the local optimum.

$$\begin{cases} x'_i = \frac{u_i+l_i}{2} + \frac{u_i+l_i}{2\omega} - \frac{x_i}{\omega} \\ \omega = \frac{\sigma}{2} - \left(\frac{e^{t/T}-1}{e-1}\right)^\sigma \end{cases} \quad (21)$$

In Equation (21), T embodies the iteration count and t shows the existing iteration count. σ controls the attenuation rate of ω ; the larger the σ , the slower ω decays. By using the greedy approach, individuals with lower fitness value are rejected while individuals with high fitness value are retained after attaining the reverse location of chimps, as follows:

$$x_{update} = \max_fitness(x_i, x'_i). \quad (22)$$

The ICOA method derives an FF to achieve high efficiency of classification. It describes a positive integer to portray the better outcomes of the solution. The decline of the classification error rate is considered FF.

$$\begin{aligned} fitness(x_i) &= ClassifierErrorRate(x_i) \\ &= \frac{No. of misclassified samples}{Total No. of samples} * 100 \end{aligned} \quad (23)$$

4. Results and Discussion

The proposed model is simulated using the Python 3.6.5 tool on PC i5-8600k, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD, and 1TB HDD. The parameter settings are given as follows: learning rate: 0.01; dropout: 0.5; batch size: 5; epoch count: 50; activation: ReLU.

The experimental evaluation of the ICOA-DLVDC technique is performed on two datasets: the VEDAI [22] and ISPRS Postdam [23] datasets. The former dataset includes 3687 images; and the latter dataset has 2244 images. Tables 1 and 2 defined a detailed description of the two datasets. Figure 3 depicts the sample images.

Table 1. Details of VEDAI dataset.

Class	No. of Instances
Car	1340
Truck	300
Van	100
Pickup Car	950
Boat	170
Camping Car	390
Other	200
Plane	47
Tractor	190
Total Instances	3687

Table 2. Details on ISPRS Postdam dataset.

Class	No. of Instances
Car	1990
Truck	33
Van	181
Pickup Car	40
Total Instances	2244

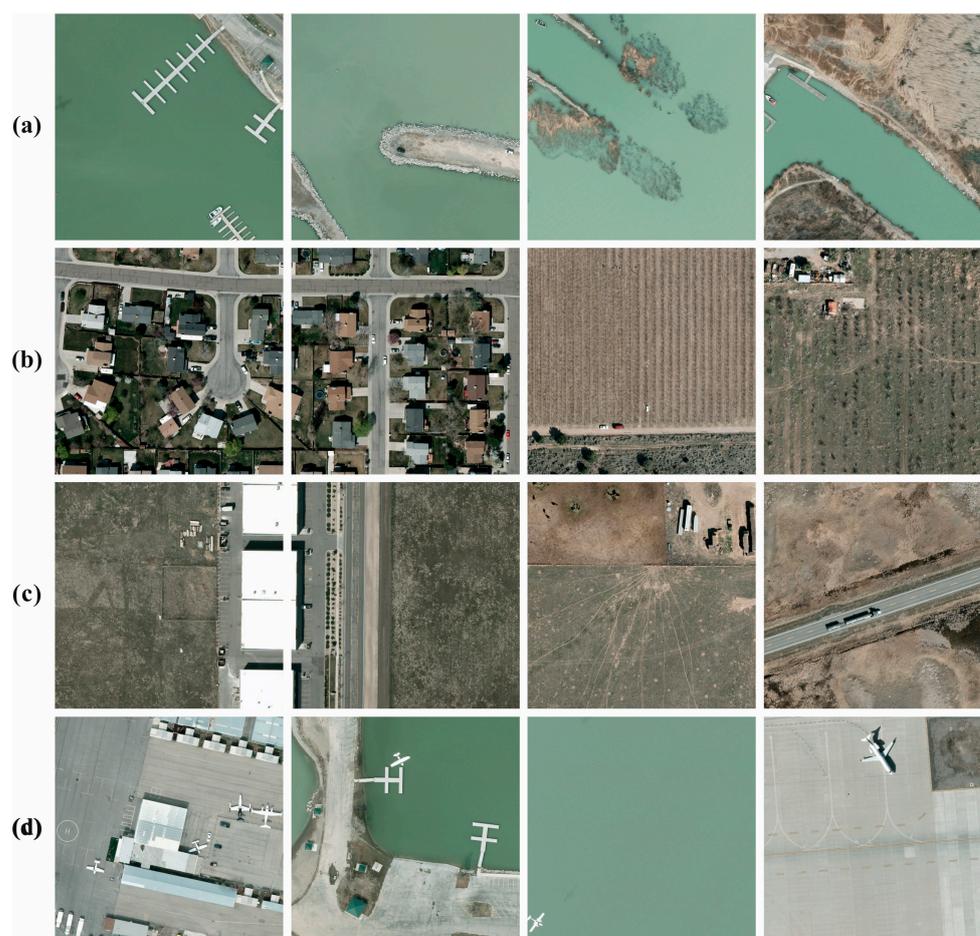


Figure 3. Sample images: (a) boat, (b) car, (c) pickup car, (d) airplane.

Figure 4 illustrates the classifier outcomes of the ICOA-DLVDC method under the VEDAI dataset. Figure 4a,b describes the confusion matrix presented by the ICOA-DLVDC technique at 70:30 of the TR set/TS set. The figure denoted that the ICOA-DLVDC method has detected and classified all nine class labels accurately. Similarly, Figure 4c demonstrates the PR examination of the ICOA-DLVDC system. The figure showed that the ICOA-DLVDC method has accomplished maximal PR outcomes under nine classes. Finally, Figure 4d demonstrates the ROC examination of the ICOA-DLVDC method. The figure demonstrates that the ICOA-DLVDC method has resulted in proficient outcomes with the highest ROC values under nine class labels.

In Table 3, the vehicle classification outcomes of the ICOA-DLVDC method on the VEDAI dataset are reported. The table values state that the ICOA-DLVDC technique properly recognized all the vehicle types. With 70% of the TR set, the ICOA-DLVDC technique gains average $accu_y$, $prec_n$, $reca_1$, F_{score} , and MCC of 99.43%, 96.66%, 94.45%, 95.43%, and 95.15% respectively. Moreover, with 30% of the TS set, the ICOA-DLVDC method gains average $accu_y$, $prec_n$, $reca_1$, F_{score} , and MCC of 99.50%, 97.27%, 94.45%, 95.94%, and 95.72%, respectively.

Figure 5 shows the training accuracy TR_accu_y and VL_accu_y of the ICOA-DLVDC method on the VEDAI dataset. The TL_accu_y is determined by the evaluation of the ICOA-DLVDC technique on the TR dataset; whereas the VL_accu_y is computed by evaluating the performance on a separate testing dataset. The outcomes demonstrate that TR_accu_y and VL_accu_y increase with an upsurge in epochs. Thus, the performance of the ICOA-DLVDC method is improved on the TR and TS datasets, with a rise in several epochs.

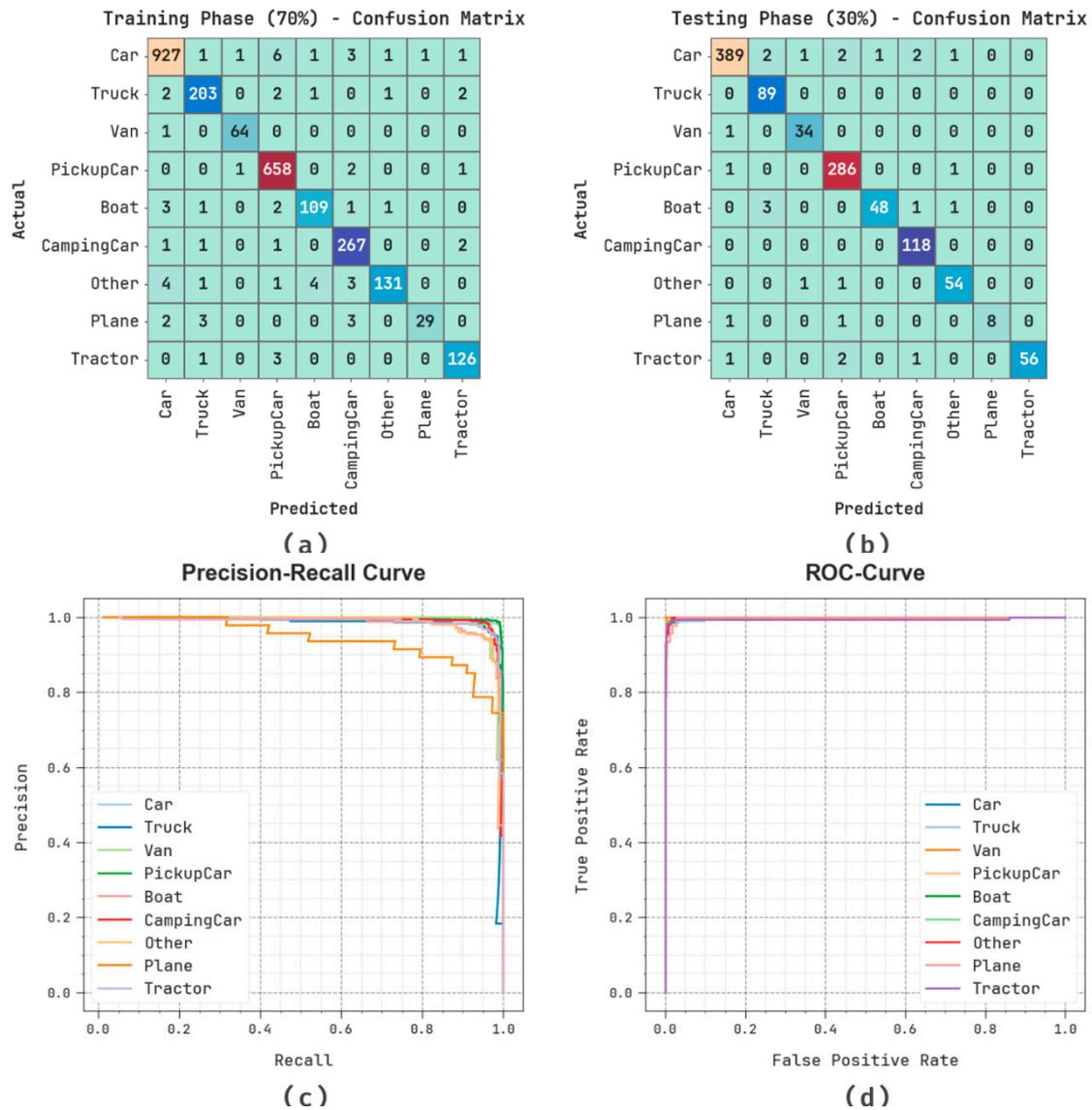


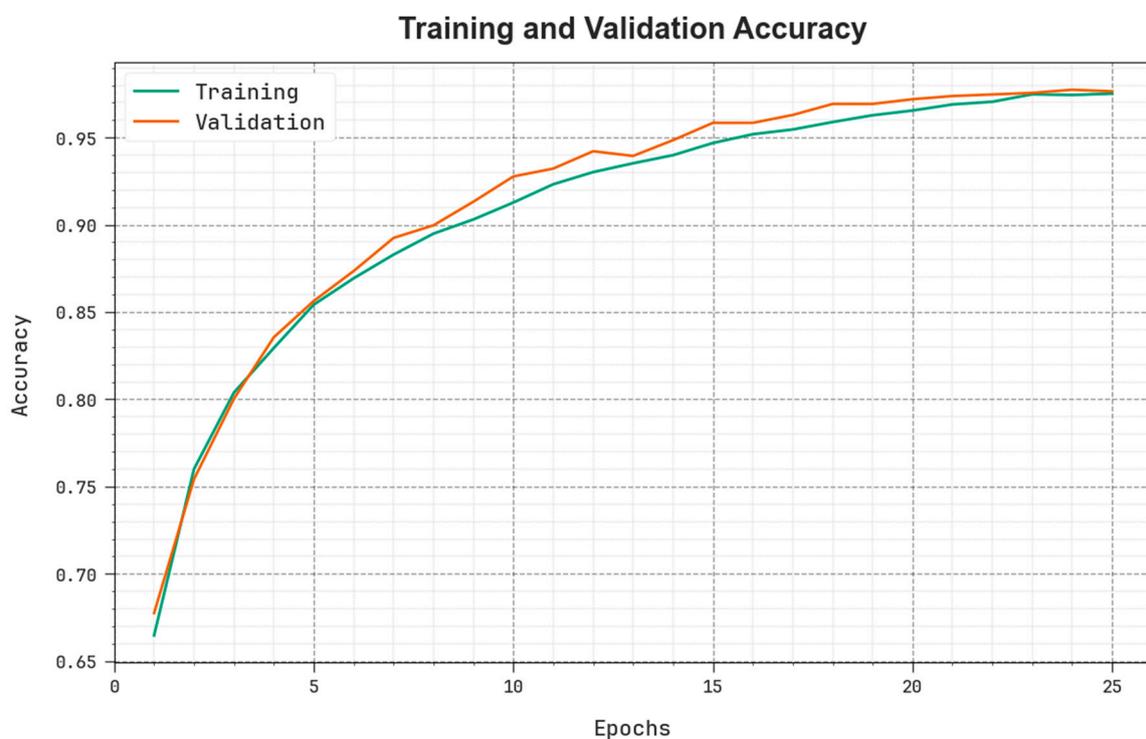
Figure 4. Performance on VEDAI datasets: (a,b) confusion matrices; (c) PR_curve; (d) ROC_curve.

Table 3. Vehicle classifier outcome of ICOA-DLVDC technique on VEDAI dataset.

Labels	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC
Training Phase (70%)					
Car	98.91	98.62	98.41	98.51	97.66
Truck	99.38	96.21	96.21	96.21	95.87
Van	99.88	96.97	98.46	97.71	97.65
Pickup Car	99.26	97.77	99.40	98.58	98.09
Boat	99.46	94.78	93.16	93.97	93.69
Camping Car	99.34	95.70	98.16	96.91	96.56
Other	99.38	97.76	90.97	94.24	93.99
Plane	99.65	96.67	78.38	86.57	86.88
Tractor	99.61	95.45	96.92	96.18	95.98
Average	99.43	96.66	94.45	95.43	95.15

Table 3. Cont.

Labels	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC
Testing Phase (30%)					
Car	98.83	98.98	97.74	98.36	97.45
Truck	99.55	94.68	100.00	97.27	97.06
Van	99.73	94.44	97.14	95.77	95.64
Pickup Car	99.28	97.95	99.31	98.62	98.14
Boat	99.46	97.96	90.57	94.12	93.91
Camping Car	99.64	96.72	100.00	98.33	98.15
Other	99.55	94.74	96.43	95.58	95.34
Plane	99.82	100.00	80.00	88.89	89.36
Tractor	99.64	100.00	93.33	96.55	96.43
Average	99.50	97.27	94.95	95.94	95.72

Figure 5. $Accu_y$ curve of the ICOA-DLVDC technique on the VEDAI dataset.

In Figure 6, the TR_{loss} and VR_{loss} outcomes of the ICOA-DLVDC method on the VEDAI dataset are shown. The TR_{loss} defines the error among the predictive performance and original values on the TR data. The VR_{loss} represents the measure of the performance of the ICOA-DLVDC technique on individual validation data. The results indicate that the TR_{loss} and VR_{loss} tend to decrease with rising epochs. They portray the enhanced performance of the ICOA-DLVDC method and its capability to generate accurate classification. The reduced value of TR_{loss} and VR_{loss} demonstrates the enhanced performance of the ICOA-DLVDC technique in capturing patterns and relationships.

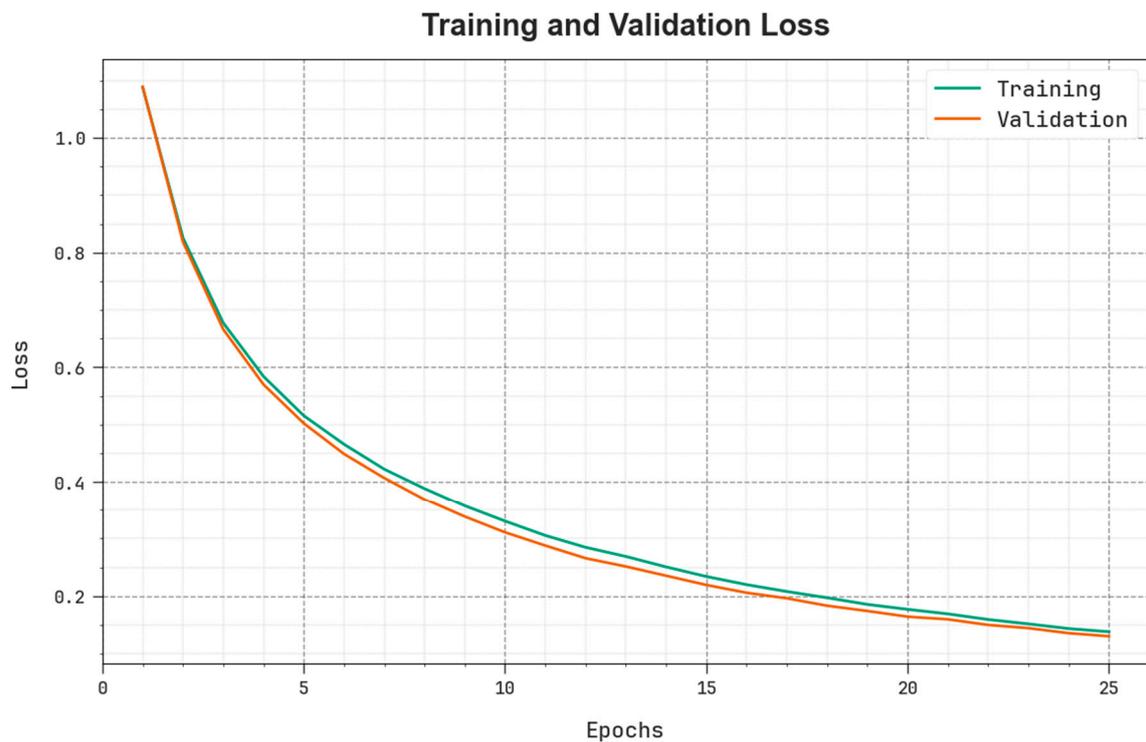


Figure 6. Loss curve of the ICOA-DLVDC technique on the VEDAI dataset.

The comparison study of the ICOA-DLVDC technique with other DL models on the VEDAI dataset is highlighted in Table 4 and Figure 7 [24]. The outcomes show that the ICOA-DLVDC technique accomplishes improved performance with an $accu_y$ of 99.50%. On the other hand, the CSOTL-VDCRS, LeNet, AlexNet, and VGG-16 models achieve reduced performance with $accu_y$ of 98.07%, 79.78%, 88.98%, and 94.46%, respectively.

Table 4. $Accu_y$ outcome of the ICOA-DLVDC technique with recent methods on the VEDAI dataset.

VEDAI Dataset	
Methods	Accuracy (%)
ICOA-DLVDC	99.50
CSOTL-VDCRS	98.07
LeNet Model	79.74
AlexNet Model	88.98
VGG-16 Model	94.46

Figure 8 illustrates the classifier results of the ICOA-DLVDC technique on the ISPRS Postdam dataset. Figure 8a,b demonstrates the confusion matrix presented by the ICOA-DLVDC system at 70:30 of the TR set/TS set. The figure demonstrates that the ICOA-DLVDC method has detected and classified all four class labels accurately. Similarly, Figure 8c demonstrates the PR examination of the ICOA-DLVDC model. The figure shows that the ICOA-DLVDC technique has accomplished high PR outcomes under four classes. Lastly, Figure 8d elucidates the ROC examination of the ICOA-DLVDC model. The figure shows that the ICOA-DLVDC method has resulted in proficient outcomes, with the highest ROC values under four class labels.

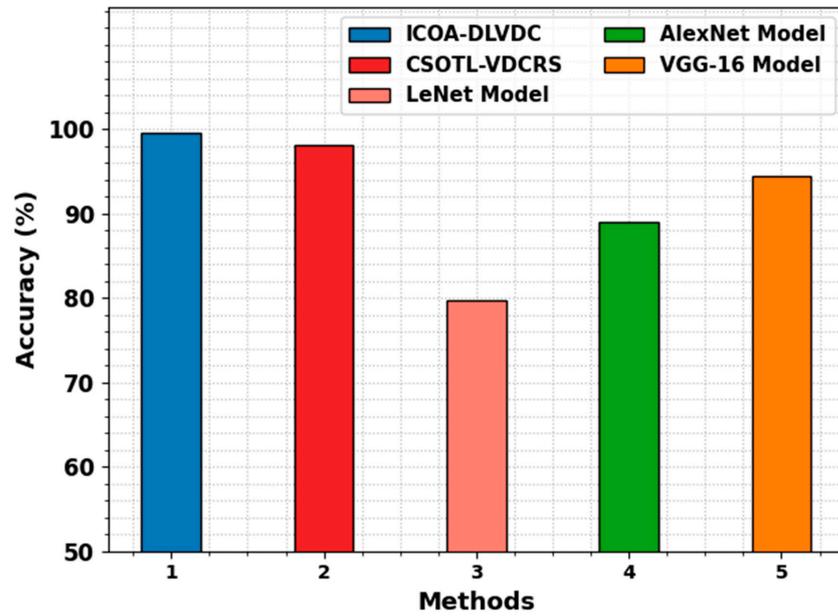


Figure 7. Accuracy outcome of the ICOA-DLVDC technique on the VEDAI dataset.

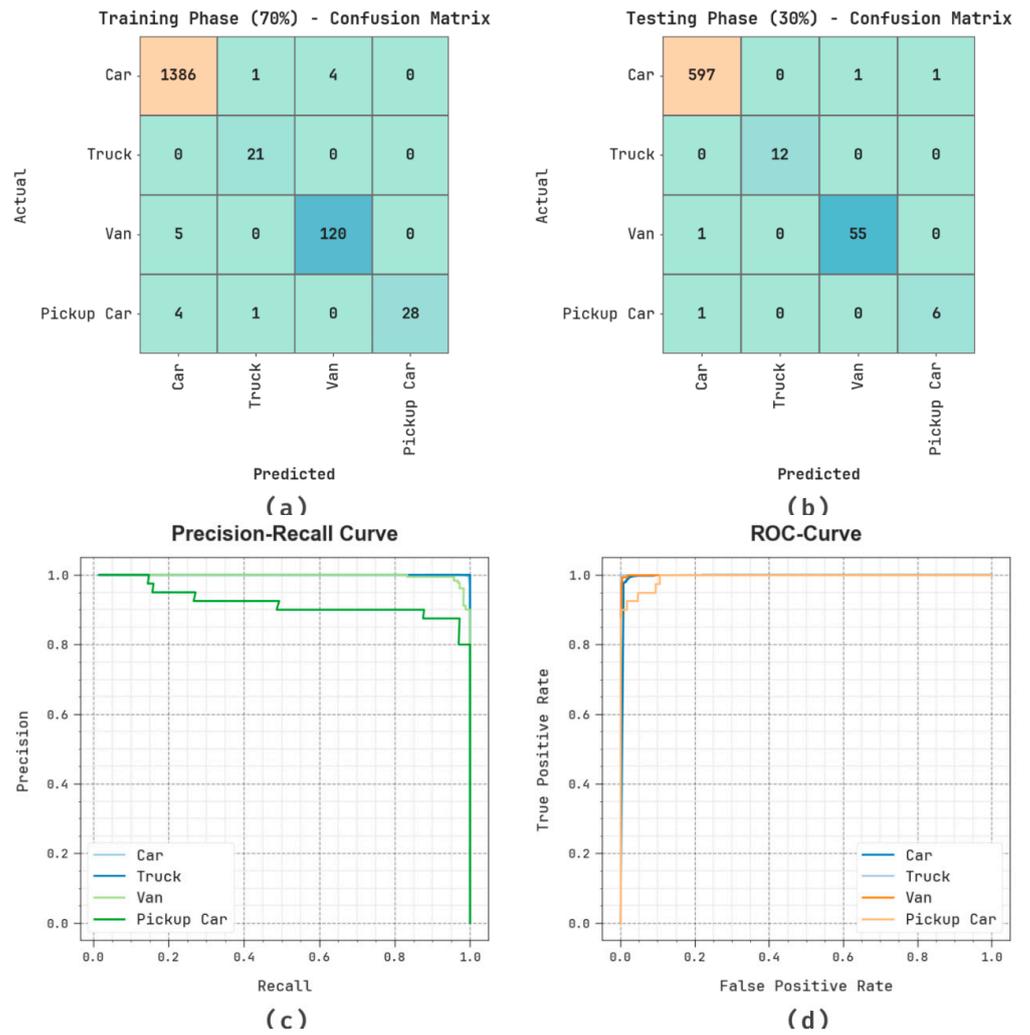


Figure 8. Performance on the ISPRS Postdam dataset: (a,b) confusion matrices; (c) PR_curve; (d) ROC_curve.

In Table 5, the vehicle classification outcomes of the ICOA-DLVDC technique on the ISPRS Postdam dataset are reported. The table values stated that the ICOA-DLVDC technique properly recognized all the vehicle types. With 70% of the TR set, the ICOA-DLVDC method gains average $accu_y$, $prec_n$, $reca_1$, F_{score} , and MCC of 99.52%, 96.86%, 95.12%, 95.79%, and 94.77%, respectively. Furthermore, with 30% of the TS set, the ICOA-DLVDC method gains average $accu_y$, $prec_n$, $reca_1$, F_{score} , and MCC of 99.70%, 95.90%, 95.90%, 95.90%, and 95.15%, respectively.

Table 5. Vehicle classifier outcome of the ICOA-DLVDC technique on the ISPRS Postdam dataset.

Labels	$Accu_y$	$Prec_n$	$Reca_1$	F_{score}	MCC
Training Phase (70%)					
Car	99.11	99.35	99.64	99.50	95.55
Truck	99.87	91.30	100.00	95.45	95.49
Van	99.43	96.77	96.00	96.39	96.08
Pickup Car	99.68	100.00	84.85	91.80	91.96
Average	99.52	96.86	95.12	95.79	94.77
Testing Phase (30%)					
Car	99.41	99.67	99.67	99.67	97.00
Truck	100.00	100.00	100.00	100.00	100.00
Van	99.70	98.21	98.21	98.21	98.05
Pickup Car	99.70	85.71	85.71	85.71	85.56
Average	99.70	95.90	95.90	95.90	95.15

Figure 9 shows the training accuracy TR_accu_y and VL_accu_y of the ICOA-DLVDC technique on the ISPRS Postdam dataset. The TL_accu_y is determined by the evaluation of the ICOA-DLVDC technique on the TR dataset; whereas the VL_accu_y is computed by evaluating the performance on a separate testing dataset. The outcomes demonstrate that TR_accu_y and VL_accu_y increase with an upsurge in epochs. As a result, the performance of the ICOA-DLVDC technique is improved on the TR and TS dataset, with a rise in the number of epochs.

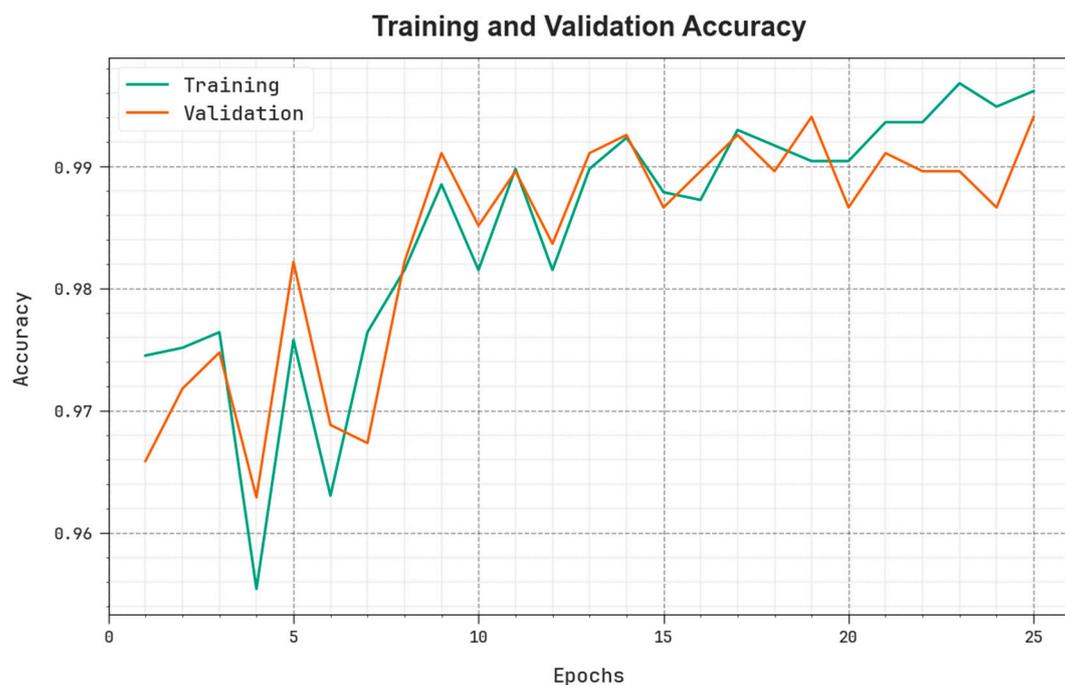


Figure 9. $Accu_y$ curve of the ICOA-DLVDC technique on the ISPRS Postdam dataset.

In Figure 10, the TR_loss and VR_loss outcomes of the ICOA-DLVDC technique on ISPRS Postdam dataset are shown. The TR_loss defines the error among the predictive performance and original values on the TR data. The VR_loss represents the measure of the performance of the ICOA-DLVDC technique on individual validation data. The results indicate that the TR_loss and VR_loss tend to decrease with rising epochs. The portray the enhanced performance of the ICOA-DLVDC technique and its capability to generate accurate classification. The reduced value of TR_loss and VR_loss demonstrates the enhanced performance of the ICOA-DLVDC technique in capturing patterns and relationships.



Figure 10. Loss curve of the ICOA-DLVDC technique on the ISPRS Postdam dataset.

The comparison analysis of the ICOA-DLVDC method with other DL techniques [24] on the ISPRS Postdam dataset is highlighted in Table 6 and Figure 11. The outcome specified that the ICOA-DLVDC technique accomplishes improved performance, with an accuracy of 99.70%. On the other hand, the CSOTL-VDCRS, LeNet, AlexNet, and VGG-16 models achieve reduced performance, with accuraciwa of 98.67%, 94.54%, 95.86%, and 89.54%, respectively.

Table 6. $Accu_y$ outcome of ICOA-DLVDC technique with recent methods on ISPRS Postdam dataset.

Methods	Accuracy (%)
ICOA-DLVDC	99.70
CSOTL-VDCRS	98.67
LeNet Model	94.54
AlexNet Model	95.86
VGG-16 Model	89.54

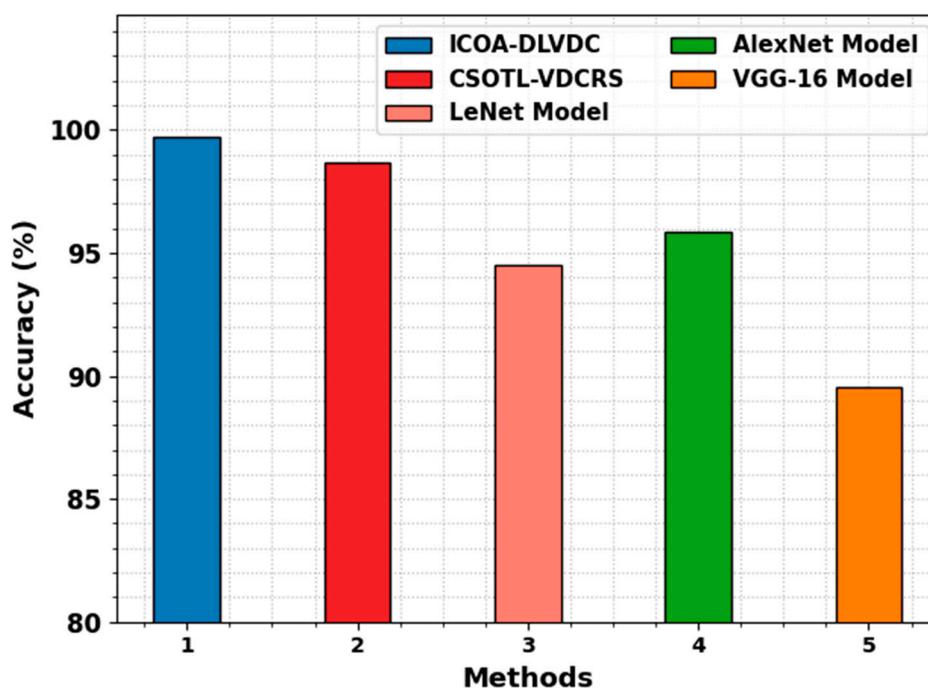


Figure 11. Accuracy outcome of the ICOA-DLVDC technique on the ISPRS Postdam dataset.

5. Conclusions

In this study, we have introduced the ICOA-DLVDC technique for automated vehicle detection and classification on RSI. In the presented ICOA-DLVDC technique, DL-based object detectors and classifiers are applied. The presented ICOA-DLVDC technique involves two phases: EfficientDet-based object detector and ICOA with SAE-based classification. An extensive set of experiments has been conducted to highlight the improved vehicle classification outcomes of the ICOA-DLVDC method. The experimental outcomes demonstrated the remarkable performance of the ICOA-DLVDC technique over other recent approaches, with maximum accuracy of 99.70% and 99.50% on the VEDAI dataset and the ISPRS Postdam dataset, respectively. In the future, we will examine the performance of the ICOA-DLVDC algorithm in different environments, such as day and night times, as well as cloudy and rainy environments. In addition, the computational time of the proposed model can be examined in the future. Moreover, the vehicle detection results can be integrated into geographic information systems (GIS) for better spatial analysis and decision-making. Finally, lightweight models can be developed for edge computing and deployment on resource-constrained devices such as drones and IoT devices.

Author Contributions: Conceptualization, M.A. (Masoud Alajmi) and H.A.; Methodology, M.A. (Masoud Alajmi), H.A., F.A.-M. and K.M.O.; Software, K.M.O.; Validation, K.M.O. and A.S.; Formal analysis, F.A.-M.; Investigation, M.A. (Masoud Alajmi); Data curation, M.A. (Mohammed Aljebreen) and A.S.; Writing—original draft, M.A. (Masoud Alajmi), H.A., F.A.-M. and M.A. (Mohammed Aljebreen); Writing—review & editing, H.A., F.A.-M., M.A. (Mohammed Aljebreen), K.M.O. and A.S.; Visualization, M.A. (Mohammed Aljebreen); Funding acquisition, H.A., F.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through large group Research Project under grant number (RGP2/35/44). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R361), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Research Supporting Project number (RSP2023R459), King Saud University, Riyadh, Saudi Arabia. This study is partially funded by the Future University in Egypt (FUE).

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated during the current study.

Conflicts of Interest: The authors declare that they have no conflict of interest. The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

References

1. Wang, Y.; Peng, F.; Lu, M.; Asif Iqbal, M. Information Extraction of the Vehicle from High-Resolution Remote Sensing Image Based on Convolution Neural Network. *Recent Adv. Electr. Electron. Eng. (Former. Recent Pat. Electr. Electron. Eng.)* **2023**, *16*, 168–177.
2. Anusha, C.; Rupa, C.; Samhitha, G. Region-based detection of ships from remote sensing satellite imagery using deep learning. In Proceedings of the 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 23–25 February 2022; IEEE: New York, NY, USA, 2022; Volume 2, pp. 118–122.
3. Chen, Y.; Qin, R.; Zhang, G.; Albanwan, H. Spatial-temporal analysis of traffic patterns during the COVID-19 epidemic by vehicle detection using planet remote-sensing satellite images. *Remote Sens.* **2021**, *13*, 208. [[CrossRef](#)]
4. Wang, L.; Shoulin, Y.; Alyami, H.; Laghari, A.A.; Rashid, M.; Almotiri, J.; Alyamani, H.J.; Alturise, F. A novel deep learning—based single shot multibox detector model for object detection in optical remote sensing images. *Geosci. Data J.* **2022**, 1–15. [[CrossRef](#)]
5. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Remote Sensing: Classification, Detection, and Segmentation. *Remote Sens.* **2023**, *15*, 1821. [[CrossRef](#)]
6. Karnick, S.; Ghalib, M.R.; Shankar, A.; Khapre, S.; Tayubi, I.A. A novel method for vehicle detection in high-resolution aerial remote sensing images using YOLT approach. *Multimed. Tools Appl.* **2022**, *109*, 1–16.
7. Wang, B.; Xu, B. A feature fusion deep-projection convolution neural network for vehicle detection in aerial images. *PLoS ONE* **2021**, *16*, e0250782. [[CrossRef](#)]
8. Wang, J.; Teng, X.; Li, Z.; Yu, Q.; Bian, Y.; Wei, J. VSAI: A Multi-View Dataset for Vehicle Detection in Complex Scenarios Using Aerial Images. *Drones* **2022**, *6*, 161. [[CrossRef](#)]
9. Safarov, F.; Temurbek, K.; Jamoljon, D.; Temur, O.; Chedjou, J.C.; Abdusalomov, A.B.; Cho, Y.I. Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture. *Sensors* **2022**, *22*, 9784. [[CrossRef](#)]
10. Momin, M.A.; Junos, M.H.; Mohd Khairuddin, A.S.; Abu Talip, M.S. Lightweight CNN model: Automated vehicle detection in aerial images. *Signal Image Video Process.* **2022**, *17*, 1–9. [[CrossRef](#)]
11. Ahmed, I.; Ahmad, M.; Chehri, A.; Hassan, M.M.; Jeon, G. IoT Enabled Deep Learning Based Framework for Multiple Object Detection in Remote Sensing Images. *Remote. Sens.* **2022**, *14*, 4107. [[CrossRef](#)]
12. Chen, J.; Hong, H.; Song, B.; Guo, J.; Chen, C.; Xu, J. MDCT: Multi-Kernel Dilated Convolution and Transformer for One-Stage Object Detection of Remote Sensing Images. *Remote. Sens.* **2023**, *15*, 371. [[CrossRef](#)]
13. Qiu, Z.; Bai, H.; Chen, T. Special Vehicle Detection from UAV Perspective via YOLO-GNS Based Deep Learning Network. *Drones* **2023**, *7*, 117. [[CrossRef](#)]
14. Tian, Z.; Huang, J.; Yang, Y.; Nie, W. KCFS-YOLOv5: A High-Precision Detection Method for Object Detection in Aerial Remote Sensing Images. *Appl. Sci.* **2023**, *13*, 649. [[CrossRef](#)]
15. Ye, Y.; Ren, X.; Zhu, B.; Tang, T.; Tan, X.; Gui, Y.; Yao, Q. An Adaptive Attention Fusion Mechanism Convolutional Network for Object Detection in Remote Sensing Images. *Remote. Sens.* **2022**, *14*, 516. [[CrossRef](#)]
16. Xiaolin, F.; Fan, H.; Ming, Y.; Tongxin, Z.; Ran, B.; Zenghui, Z.; Zhiyuan, G. Small object detection in remote sensing images based on super-resolution. *Pattern Recognit. Lett.* **2022**, *153*, 107–112. [[CrossRef](#)]
17. Javadi, S.; Dahl, M.; Pettersson, M.I. Vehicle Detection in Aerial Images Based on 3D Depth Maps and Deep Neural Networks. *IEEE Access* **2021**, *9*, 8381–8391. [[CrossRef](#)]
18. Wu, Y.; Zhang, K.; Wang, J.; Wang, Y.; Wang, Q.; Li, X. GCWNet: A Global Context-Weaving Network for Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–12. [[CrossRef](#)]
19. AlDahoul, N.; Karim, H.A.; De Castro, A.; Tan, M.J.T. Localization and classification of space objects using EfficientDet detector for space situational awareness. *Sci. Rep.* **2022**, *12*, 21896. [[CrossRef](#)]
20. Akila, S.M.; Imanov, E.; Almezghwi, K. Investigating Beta-Variational Convolutional Autoencoders for the Un-supervised Classification of Chest Pneumonia. *Diagnostics* **2023**, *13*, 2199. [[CrossRef](#)]
21. Chen, Q.; He, Q.; Zhang, D. UAV Path Planning Based on an Improved Chimp Optimization Algorithm. *Axioms* **2023**, *12*, 702. [[CrossRef](#)]
22. Razakarivony, S.; Jurie, F. Vehicle detection in aerial imagery: A small target detection benchmark. *J. Vis. Commun. Image Represent.* **2016**, *34*, 187–203. [[CrossRef](#)]

23. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Benitez, S.; Breitkopf, U. The ISPRS Benchmark on Urban Object Classification and 3D Building Reconstruction. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *1*, 293–298. [[CrossRef](#)]
24. Ahmed, M.A.; Alhubiti, S.A.; de Albuquerque, V.H.C.; dos Reis, M.C.; Shashidhar, C.; Murthy, T.S.; Lydia, E.L. Fuzzy wavelet neural network driven vehicle detection on remote sensing imagery. *Comput. Electr. Eng.* **2023**, *109*, 108765. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.