



# OPEN Optimal deep learning based vehicle detection and classification using chaotic equilibrium optimization algorithm in remote sensing imagery

Youseef Alotaibi<sup>1</sup>, Krishnaraj Nagappan<sup>2</sup>, Tamilvizhi Thanarajan<sup>3</sup>✉ & Surendran Rajendran<sup>4</sup>

Remote sensing images (RSI), such as aerial or satellite images, produce a large-scale view of the Earth's surface, which gets them used to track and monitor vehicles from several settings, like border control, disaster response, and urban traffic surveillance. Vehicle detection and classification using RSIs is a vital application of computer vision and image processing. It contains locating and identifying vehicles from the image. It is done using many approaches that have object detection approaches, namely YOLO, Faster R-CNN, or SSD, which utilize deep learning (DL) to locate and identify the image. Additionally, the classification of vehicles from RSIs contains classification of them based on their variety, such as trucks, motorcycles, cars or buses, utilizing machine learning (ML) techniques. This article designed and developed an automated vehicle type detection and classification using a chaotic equilibrium optimization algorithm with deep learning (VDTC-CEOADL) on high-resolution RSIs. The VDTC-CEOADL technique presented examines high-quality RSIs for the accurate detection and classification of vehicles. The VDTC-CEOADL technique employs a YOLO-HR object detector with a residual network as the backbone model to accomplish this. In addition, CEOA based hyperparameter optimizer is designed for the parameter tuning of the ResNet model. For the vehicle classification process, the VDTC-CEOADL technique exploits the attention-based long-short-term memory (ALSTM) mod-el. Performance validation of the VDTC-CEOADL technique is validated on a high-resolution RSI dataset, and the results portrayed the supremacy of the VDTC-CEOADL technique in terms of different measures.

**Keywords** Deep learning, Vehicle detection, Vehicle classification, Chaotic equilibrium optimization algorithm, Remote sensing images

The ongoing process of urban development is intensifying the problems related to controlling and dealing with traffic in road networks. Urban population growth at high speed has created more demanding issues in managing traffic flow and reducing congestion while optimizing urban resource distribution. Modern intelligent transportation systems together with urban traffic surveillance operations strongly depend on effective systems which precisely detect and classify vehicles. The remote sensing (RS) techniques which use aerial and satellites to obtain imagery offer wide-scale urban environment observations for real-time large-scale vehicle monitoring operations. The precise detection and identification of vehicles in high-resolution RS images (RSIs) prove difficult because of complex backgrounds alongside various object sizes and poor resolution of small objects and substantial differences between similar vehicles. Due to the consistent growth of urban populations, many related issues, such as environmental pollution, traffic jams, and excessive resource consumption, present severe challenges in urban planning and management<sup>1</sup>. Smart city construction adopted intellectual technologies such

<sup>1</sup>Department of Software Engineering, College of Computing, Umm Al-Qura University, Makkah 21955, Saudi Arabia. <sup>2</sup>Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India. <sup>3</sup>Department of Computer Science and Engineering, Panimalar Engineering College, Chennai 600123, India. <sup>4</sup>Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai 602105, India. ✉email: tamilvizhi.phd.it@gmail.com

as the Internet of Things (IoT) and deep learning (DL) to realize intelligent and scientific urban development, which is a significant target. The remote sensing (RS) technique can capture urban land cover data from great heights, thereby generating remote sensing images (RSI) that have the benefits of rich data and comprehensive spatial coverage<sup>2</sup>, which is significant for urban big data and was widely utilized in urban analysis, for example, in mapping and monitoring of slums in cities, urban heat island effect analysis, urban ecosystem analysis, and urban functional region layout<sup>3</sup>. Identifying RS targets is to detect the interest in RSI and predict the location and type of this target. In the conventional detection datasets, the target was focused. However, the aviation data were not, and the strength of the object in the aviation images generally occurs in an arbitrary orientation that relies on the viewpoint of the Earth vision platform<sup>4</sup>. The study of RSI has vital applications in environmental management, military, transport planning, and disaster control. Moreover, vehicles in RSI as a special category, whether transportation, civilian, or military, have a significant meaning, while they are more difficult<sup>5</sup>.

Using computer technology to perform tasks such as segmentation, classification, and detection of RSI has been a hot topic in the image research domain<sup>6</sup>. Vehicle detection in RSI played a crucial role in safety-assisted driving, defence, traffic planning, urban vehicle supervision, etc<sup>7</sup>. For such tasks, fast and precise detection techniques were vital. Among other detection targets in RSI (such as aeroplanes, buildings, and ships), the vehicle detection task is beginning to be challenging since there are complicated background interference data, fewer target pixels, and uneven distribution of vehicle targets. Identifying vehicle targets in RSI has been advanced for many years and has achieved effective research results. RSI vehicle detection targets to identify all vehicle instances in RSI<sup>8</sup>. In early techniques, researchers generally manually extracted and designed vehicle features and classified them to obtain vehicle detection. The central ideology is to use classical machine learning (ML) approaches and classic vehicle features for classification<sup>9</sup>. However, conventional target detection approaches pay more attention to the completion of RSI vehicle detection tasks, and it needs to be improved to balance speed and accuracy. There must be more precision and effectiveness of detection compared to the rapid growth of DL technology<sup>10</sup>. Network methods related to DL techniques can derive richer features and map complicated nonlinear relations.

YOLO and SSD along with Faster R-CNN represent DL object detection models that have remarkably enhanced vehicle detection abilities within RSIs. Long short-term memory (LSTM) networks improve sequential data learning in classification tasks because of their employment in such applications. These methods face two main challenges because they lack optimized hyperparameter settings that affect detection accuracy and they do not have adequate attention mechanisms to focus effectively on key features when classifying vehicles in urban environments.

The proposed research introduces a new Vehicle Detection and Classification using Chaotic Equilibrium Optimization Algorithm with Deep Learning (VDTC-CEOADL) technique to address current limitations. The proposed method integrates three advanced approaches in a synergistic manner as its main novelty.

- (i) A YOLO-HR object detection system was designed for high-resolution RSIs by integrating ResNet to optimize backbone features and multi-scale detection techniques.
- (ii) The Chaotic Equilibrium Optimization Algorithm (CEOA) serves to optimize the YOLO-HR model hyperparameters effectively.
- (iii) Attention-based Long Short-Term Memory (ALSTM) model with an attention mechanism enhances vehicle classification by dynamically selecting important vehicle instance features for better performance.

## Related works

Song et al.<sup>11</sup> devised a vehicle detection and counting mechanism; The road surfaces of the highways in the images are divided and extracted into proximal and remote areas using a new segmentation approach; the technique is vital to improve vehicle detection. Subsequently, the above two areas were placed in the YOLOv3 network to identify the position and kind of vehicle. Finally, with the ORB approach, vehicle trajectories were reached, which is utilized to determine the vehicle's driving direction and gain different vehicles. Rafique et al.<sup>12</sup> designed a new vehicle classification and detection technique for smart traffic monitoring that uses CNN to segment aerial images. Such segmented imageries were inspected to identify vehicles by including new custom pyramid grouping. Lastly, such vehicles were tracked using kernelized filter-based techniques and Kalman filter (KF) to deal with and achieve massive traffic flows with less human interference. Bibi et al.<sup>13</sup> introduced a new system for automatically recognise road anomalies by autonomous vehicles and to provide road data to upcoming vehicles based on VANET and Edge AI. Road images captured by cameras and the disposition of trained methods to detect vehicle anomalies could reduce the risk of hazards and the accident rate in worse road conditions. The VGG11 and ResNet-18 were implemented to classify and automatically detect roads with anomalies such as cracks, potholes, plain roads and bumps without anomalies utilizing data from various online sources.

In<sup>14</sup>, the author presented a structure to minimize the complexity of CNN-related AVS approaches. In contrast, the BS-related module was presented as a pre-processing stage for optimizing the convolution operation by the CNNs. A CNN-related detector with suitable convolution was implemented for all image candidates to manage the overlapping issue and boost the detecting performance. Four existing CNN-based detection structures have been benchmarked as base methods of the detection cores for evaluating the presented structure. Qiu et al.<sup>15</sup> introduced a new technique called YOLO-GNS for detecting special vehicles from a drone point of view. First, to enhance feature extraction and ease identification of obscured or small objects, the single stage headless (SSH) context structure was presented. In the mean time, the computational cost of the technique was minimized because of GhostNet by exchanging the complicated convolutional with linear conversion through modest operation.

In<sup>16</sup>, to attain front vehicle detection in assisted driving or fast front vehicle detection without manned vehicles, an efficient and vision-based technique was presented that depends on the spatial and temporal features of the

front vehicle. Initially, a method depends on the spatial position constraint, and Oriented FAST and Rotated BRIEF (ORB) were devised for extracting the motion vector of the front vehicle. Afterwards, by examining the changes among the motion vectors of the vehicle and the background, the feature points of the vehicle were derived. Lastly, to realize front-vehicle recognition, a feature-point clustering approach related to integrating temporal and spatial features was implemented. Zhou et al.<sup>17</sup> projected a target detection technique called IR-PANet for the RSI of an automobile. SPP was utilized to strengthen learning content in the CSPDarknet53 backbone network. Afterwards, IR-PANet is utilized as the neck network. A depth-wise separable convolution after upper sampling was used to ignore the lack of smaller targeted feature data in the convolution of shallow network and rise semantic data in the higher-level network.

The classification of high-resolution remote sensing images encounters three main difficulties which include scale variation together with long-range dependencies and similar classes. Researchers proposed a novel Residual Channel-attention (RCA) network which integrates residual structures and SE mechanisms with channel attention to solve these issues<sup>18</sup>. A combination of background subtraction and a customized YOLOv8 model for performing semi-automated unregulated object detection. The system performs automatic moving object labeling through low-rank decomposition and clustering techniques. The proposed method delivers better detection results which produce superior mAP scores on UA-DETRAC and CDnet 2014 datasets when compared to recent detection approaches<sup>19</sup>.

An immediate system that tracks and detects vehicles in aerial video content through morphological transformation analysis combined with motion data evaluation. The system merges KLT tracking with K-means clustering to divide moving vehicles before it efficiently tracks their movements<sup>20</sup>. Semi-automatic object detection system is proposed to unite background subtraction with a modified YOLOv4 model for surveillance video applications. The method performs automatic training data labeling through motion-based low-rank decomposition and clustering while eliminating the need for manual labeling<sup>21</sup>.

## Materials and methods

In this study, we have introduced a new VDTC-CEOADL technique for the identification and classification of vehicles on high-resolution RSIs. The VDTC-CEOADL technique presented examined high-quality RSIs to precisely detect and classify different types of vehicle. The VDTC-CEOADL technique comprises several subprocesses: YOLO-HR based object detector, CEOA based hyperparameter tuning, and ALSTM based vehicle classification. Figure 1 illustrates the workflow of the VDTC-CEOADL method.

### Dataset Preparation

The proposed VDTC-CEOADL framework needed extensive dataset preparation for its implementation and evaluation using both VEDAI and ISPRS Potsdam datasets. The YOLO-HR model required  $512 \times 512$  pixel resolution inputs so all remote sensing images received initial resizing for both compliance and computational optimization. A preservation of the aspect ratio method was used for the resizing process to prevent any distortion. The pixel values received normalization to the [0, 1] range for improving the numerical stability throughout the learning process. Channel-wise normalization was implemented by deducting the mean of each channel followed by dividing it with its standard deviation to boost training convergence performance.

Different data augmentation techniques were used for the detection model to improve generalization and prevent sample size overfitting. The model incorporated random image cropping for spatial variance and employed horizontal and vertical flipping at 0.5 probability together with brightness adjustments between  $\pm 20\%$  for lighting condition simulation. The object orientation diversity was expanded through small scaling operations (0.9 to 1.1) and rotational adjustments ( $\pm 10$  degrees). The model gained additional robustness against noise through the intentional addition of Gaussian noise in particular samples. The augmentation methods operated dynamically throughout training to extend data diversity without requiring additional storage capacity.

The VEDAI dataset presented annotations through XML format containing bounding boxes with labels whereas the ISPRS Potsdam dataset provided pixel-level semantic segmentation masks. All annotations received conversion to YOLO format for pipeline compatibility with YOLO-HR detection framework by standardizing each bounding box with normalized class ID and center coordinates along with width and height specifications. The semantic masks in the ISPRS dataset underwent processing to generate bounding boxes that defined vehicle objects. The cleaning process focused on eliminating duplicated annotations while removing wrong boxes and incorrect labels from the dataset. All datasets received uniform class labels for maintaining consistent training conditions.

A patch-wise segmentation approach was implemented because the ISPRS Potsdam dataset includes large orthophoto tiles of  $6000 \times 6000$  pixels resolution. The YOLO-HR model required input sizes of  $512 \times 512$  pixels so each big tile received this partitioning into separate non-overlapping areas. Vehicle information loss prevention was achieved through a sliding window approach which utilized 20% overlapping between neighboring windows. The training and testing process included vehicle instance-bearing patches whose annotations properly matched the segmented areas. The patch-wise method created for segmentation served both to lower processing demands and to enhance the detection performance of vehicles especially when detecting compact objects.

### Object detector using the YOLO-HR model

The YOLO-HR model is utilized in this work. YO-LO-HR is an object detection network for high-resolution RSIs. The network is divided into Head, Backbone, and Neck<sup>22</sup>. The fundamental architecture of Backbone is a ResNet and convolution module at its core. The basis of deep ResNet was to present a deep residual learning model in the network. Compared to the characteristic that the stacked layer of CNN, including VGG16, is better suited for the preferred underlying mapping, ResNet makes the layer fit for residual mapping.

The ResNet encompassed the sequence of residual blocks:



**Fig. 1.** Workflow of the VDTC-CEOADL approach. The figure was created using Drawio tool. As the image is fully original and created by the authors, there is no copyright or licensing issue, and no permission document is required. The figure is eligible for publication under the CC BY 4.0 open access license.

$$x_{l+1} = x_l + F(x_l, W_l) \quad (1)$$

In Eq. (1),  $x_l$  and  $x_{l+1}$  denote input and output, and  $F(x_l, W_l)$  refers to a residual part and represents the gap among the observed and predicted values. Usually, the network layer is expressed as  $=H(x)$ , and the residual block belonging to the residual network is represented as  $H(x) = F(x) + x$ , i.e., in the unit mapping,  $y = H(x)$  are observed values, but the predicted value of  $H(x)$ , hence  $F(x)$  corresponds to the residual named ResNet. The number of feature maps of  $x_l$  and  $x_{l+1}$  is not identical in the convolution network, and the  $1 \times 1$  convolution is required for dimension reduction or elevation. Subsequently, the image of data enhancement was placed in the retrieved convolutional and network modules channel that then integrated with the Conv model using the kernel size = 6. They have interconnected with PANet in Neck after that feature enhancement model called SPPF<sup>23</sup>.

Bidirectional feature fusion was assumed to improve the network detection ability. The conv2d layer independently scales the fused feature layer for generating the multilayer output. The NMS technique integrated the output of each single-layer detector to produce the last detection frame. Conv includes a Silu activation function, 2D convolution layers, and BN layer, C3 includes two 2D convolution layers along with the bottleneck layer, and Up sample denotes the up sampling layer. The ECA is effectively implemented by 1D convolution of size  $k$  to capture local cross-channel interaction data after channel-level global average pooling without dimensionality reduction, considering the connection of every channel with neighbors  $k$ . The CA attention module encodes every channel along the vertical and horizontal axes, correspondingly using the channel-level global average pooling of size  $(H, 1)$  or  $(1, W)$  pooling kernels. The abovementioned two transformations gather feature alongside both spatial directions for producing two direction-aware feature maps later modified and concatenated by the sigmoid and convolution functions for the attention output.

### CEOA based hyperparameter tuning

In this study, the CEOA is used to optimally modify the hyperparameters of the YOLO-HR model. The original EOA is a metaheuristic approach related to physics to manage consistent optimization difficulties<sup>24</sup>. The EOA is a newly developed metaheuristic approach that takes advantage of an equilibrium pool and candidate to upgrade the particle. The EOA depends on the analytic solution process for the dynamic mass balance over the control volume. The EOA prevents being trapped in the local optima and has greater exploration and exploitation abilities. The benefit of EOA has been made possible by the concept of generation rate. The equation of mass balance is analytically resolved, producing the following result:

$$C = C_{eq} + (C_0 + C_{eq}) \cdot F + G/\lambda V (1 - F) \quad (2)$$

In Eq. (2),  $V$  can be regarded as 1 as a volume unit. The EOA produces an equilibrium pool having 4 candidates and other averaged ones:

$$C_{eqs, pool} = \{C_{eqs}(0), C_{eqs}(1), C_{eqs}(2), C_{eqs}(3), C_{ave}\} \quad (3)$$

$$C_{ave} = (C_{eqs}(0) + C_{eqs}(1) + C_{eqs}(2) + C_{eqs}(3)) / 4 \quad (4)$$

The fitness value of the candidate in the equilibrium pool must satisfy the subsequent rules for the issue indicated by  $f$ .

$$f(C_{eqs}(0)) \leq f(C_{eqs}(1)) \leq f(C_{eqs}(2)) \leq f(C_{eqs}(3)) \quad (5)$$

EOA exploits the initializing and iterating methods as another bio-inspired algorithm while resolving problems, they come from real engineering work or specified as benchmark. The EOA continues the exploration and exploitation process while iterating. To enhance the performance, there exist many operations to construct the EOA. Assume that the presented problem is limited by the symmetric or asymmetric range with  $[lb, ub]$  and that the candidate of the swarm is uniform across the domain. To accomplish this, the pseudorandom random value  $r_1$  is presented:

$$C_i = lb + (ub - lb) \cdot r_1 \quad (6)$$

The location vector for the  $i$ -th candidate is  $C_i$ . Then, the position of the candidate for the subsequent iteration would be important for the three stages, as shown in Eq. (2).  $C_{eq}$  represents a random selection of candidates from the swimming pool construction with Eq. (3) for the initial stages. There exists an exponential variable  $F$  for the second part of the Eq. (2) that can be defined:

$$F = a_1 \text{sign}(r_2 - 0.5)(e^{-\lambda t} - 1) \quad (7)$$

In Eq. (7)  $a_1$  denotes a constant variable that controls the exploration abilities along with parameters that divide the exploiting and exploring procedures. The higher the value of  $a_1$ , the greater the possibility for the candidate to perform exploration, and the lesser the possibility for the candidate to perform exploitation. For convenience and experience,  $a_1 = 2 \cdot r_2$  is another random integer within  $[0, 1]$ , and  $t$  denotes the variable expressed to be related to the iteration time.

$$t = (1 - \text{iter} / (\text{max Iter}))^{(a_2 \text{ iter} / \text{max Iter})} \quad (8)$$

Equation (8)  $\text{iter}$  signifies the existing number of iterations, and  $\text{maxIter}$  denotes the maximum iterations restricted initially.

The parameter  $G$  refers to the generation rate that increases the exploitability of the candidate. The generation rate was proportional to an exponential parameter that was shown below:

$$G_o = GCP(C_{eq} - \lambda C) \quad (9)$$

$$GCP = \{(0.5, \&r_2 \geq GP@0, \&r_2 \geq GP)\} \quad (10)$$

Where  $GP$  represents generation probability set as 0.5 to accomplish the best outcomes in balancing the possibility in-between exploitation and exploration. Upgrading rules of EO are given below:

$$C = C_{eq} + (C - C_{eq}) \cdot F + G/\lambda V (1 - F), \quad (11)$$

In Eq. (11),  $F$  is determined in Eq. (7) and  $V$  is regarded as a unit. CEOA is derived by the integration of EOA with chaotic maps. A chaotic system has the features of ergodicity and randomness<sup>25</sup>. A variety of populations are produced by applying those features, thus speeding up the convergence rate and enhancing the performance of the model. Rather than the arbitrarily generated population, it was applied to the chaotic map, which produces the initial population, making it easier for them to escape from local optima and improving the performance of the whale algorithm. Presently, various chaotic maps exist in the optimization field, primarily involving Gauss, logistic, and tent maps. The study exploits the logistic map for generating the initial population:

$$x_{i+1} = \mu x_i (1 - x_i), i = 1, 2, \dots, N - 1 \quad (12)$$

In Eq. (12),  $\mu$  is 4,  $x_{-1}$  represents the random integer within [0,1], and N represents the individual population number. The position of the initial population produced by the logistic map was normal distribution of population location associated with the arbitrarily produced that increases the population diversity and extends search space of hawks in space. For the specific range, it improves the limitations of EOA that it can be easier to be trapped in local optimum. Fitness selection was a decisive factor in the CEOA approach. The encoding of the solution was exploited to judge the aptitude of the candidate solution. Here, the accuracy value was the main condition leveraged to model a fitness function.

$$Fitness = \max (P) \tag{13}$$

$$P = (true\ positive) / (true\ positive + false\ positive) \tag{14}$$

Sub-su3.3. Vehicle Classification using the ALSTM Model At the final stage, the class labels of the detected vehicles can be identified using the ALSTM model. Hochreiter and Schmidhuber project LSTM as the key to the problem of gradient disappearance<sup>26</sup>. LSTM was planned to adaptably control learned features' storage length and illustrates optimum outcomes in the Seq2Seq problem. To solve this problem that RNNs could not learn data in long data orders, the LSTM nodes include a novel state value c, termed cell state, and its function saves the input data far away in the present moment, that is, long-term memory. The structure named gates is fixed in the LSTM node for removing or adding data to the c cell state. Gates can be fully connected (FC) layers with outcomes between zero and one. Usually, the LSTM node is the following 3 kinds of gates that upgrade and control cell states: output, forget, and input. The forget gate manages those data from cell state to forget, whose output was provided by subsequent formula:

$$f((t)) = \varphi(w_{-f} [h((t-1)), x((t))] + b_{-f}) \tag{15}$$

Whereas  $b_{-f}$  defines the bias vector of the forgetting gate,  $w_{-f}$  refers to the weighted matrix of the forget gate,  $h((t-1))$  implies the output of the hidden state of the preceding time step,  $[,]$  denotes 2 cascading vectors,  $x((t))$  represents the input of the present time step, and  $\varphi$  refers to the activation function. An input gate control that novel data has been encoded as cell state. The input gate output is written formula as:

$$\hat{i}((t)) = \varphi(w_{-i} [h((t-1)), x((t))] + b_{-i}) \tag{16}$$

in which  $w_{-i}$  represents the weighted matrix and  $b_{-i}$  indicates the bias vector. The resultant of the input gate  $i$  defines if the candidate value C created by the novel input is along with the cell state or not and can be formulated as:

$$\hat{C}((t)) = \tanh(w_{-c} [h((t-1)), x((t))] + b_{-c}) \tag{17}$$

whereas  $b_{-c}$  and  $w_{-c}$  denote the equivalent bias vector and weight matrix. The cell state of the current moment step  $c((t))$  was calculated according to the cell state of the preceding time step and candidate values of the present time steps.

$$\hat{c}((t)) = f((t)) \odot \hat{c}((t-1)) + \hat{i}((t)) \odot \hat{C}((t)) \tag{18}$$

In which,  $\odot$  implies the multiplying equivalent elements of 2 vectors. The output gate controls that the data encoder of the cell state was assigned to the network as the output of the hidden state. The outcome of the output gate was provided as follows:

$$O((t)) = \varphi(w_{-o} [h((t-1)), x((t))] + b_{-o}) \tag{19}$$

Whereas  $w_{-o}$  denotes the weighted matrix,  $b_{-o}$  implies the bias vector, and  $O$  signifies the resultant of the output gate. The LSTM layers resultant vector was provided as:

$$h^{(t)} = O^{(t)} \odot \tanh(h^{(t)}) \tag{20}$$

In which  $h^{(t)}$  is broadcast to the next network layer and the LSTM layer of the next time step.

With establishing state value and three types of gates, the LSTM network efficiently study the long-term trend by effectively battling the gradient exploitation. It could be predicted that the automated scam predictive method dependent on LSTM is the optimum solution.

In the ALSTM model, the attention module stimulates the way the human brain processes data and focuses only on the crucial part of the data while disregarding the unrelated parts. Figure 2 shows the structure of ALSTM. Subsequently, the attention module was added to the LSTM layer to assign probability weight to the output vector of the final LSTM layer for vehicle type classification. Therefore, the NN focuses more on the key content of the vector to produce an extremely effective classification. This study applied Luong's multiplicative attention to create the custom attention layer, and the computing process was listed below.

$$a_{-t} = (\exp(\text{score}(h_{-t}h)) / (\sum_{s'=1}^S \exp(\text{score}(h_{-t'}h_{-s'})))) \tag{21}$$

$$\text{score}(h_{-t}, h) = h_{-t}^T W_{-a} h \tag{22}$$

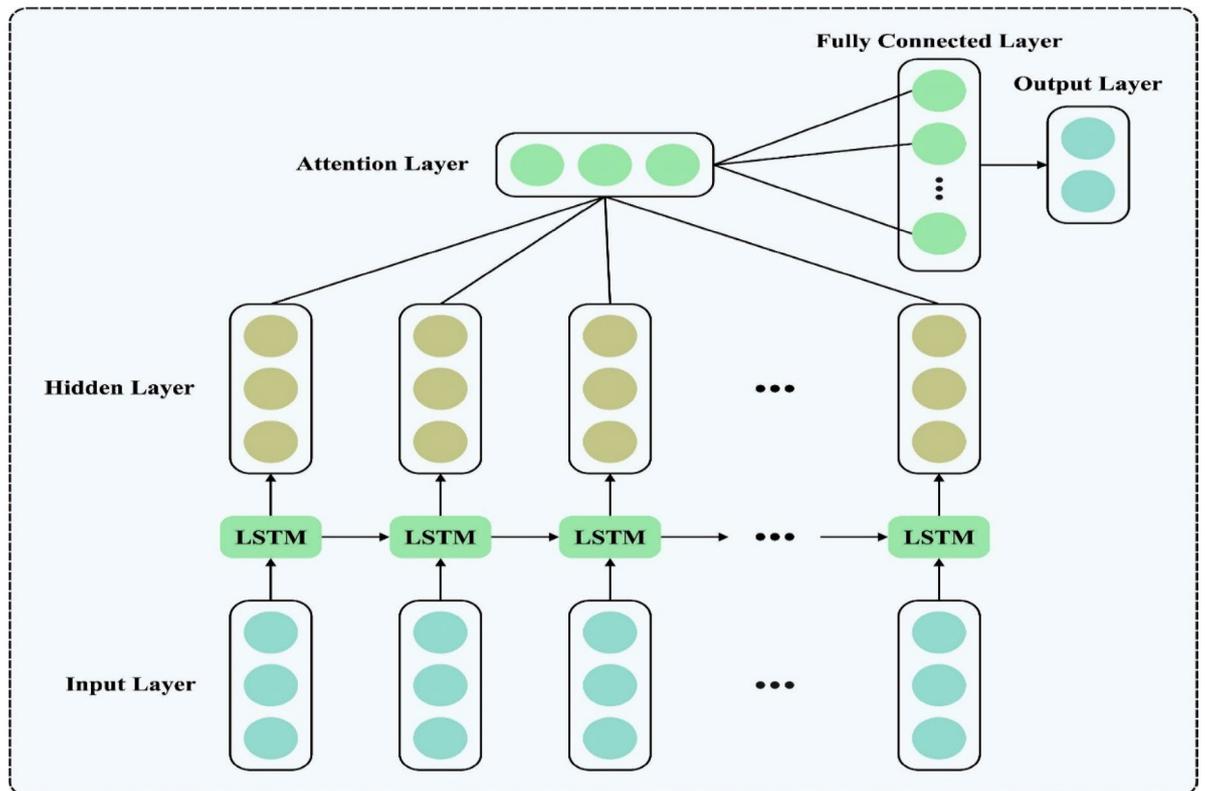


Fig. 2. Structure of ALSTM.

$$c_t = \sum_{s=1}^S a_{t,s} h_s \quad (23)$$

$$h_{att} = \tanh(W_c [c_t; h]) \quad (24)$$

here  $h_{att}$  signifies attention vector,  $h_t$  and  $h$  represent hidden present and source states, correspondingly;  $a_t$  characterizes attention weight and  $c_t$  denotes context vector.

## Results and discussion

The vehicle detection and classification results of the VDTC-CEOADL method can be studied on two datasets: VEDAI data set<sup>27</sup> and ISPRS Potsdam data set<sup>28</sup>. The first VEDAI dataset includes 3687 samples with 9 classes. Next, the ISPRS Potsdam dataset comprises 2244 samples with 4 classes. Table 1 describes the details of the VEDAI data set and the ISPRS Potsdam data set.

Figure 3 shows the classifier results of the VDTC-CEOADL method in the VEDAI data set. Figure 3a-b depict the confusion matrix offered by the VDTC-CEOADL method. The results imply that the AOAAE-SA approach has identified the nine distinct classes of vehicle detection. Similarly, Figure 3c reveals the PR analysis of the VDTC-CEOADL technique. The figures reported that the VDTC-CEOADL method has gained maximum PR performance in all classes. Finally, Figure 3d proves the ROC analysis of the VDTC-CEOADL technique. The figure shows that the VDTC-CEOADL technique has proficient outcomes with higher ROC values under distinct class labels.

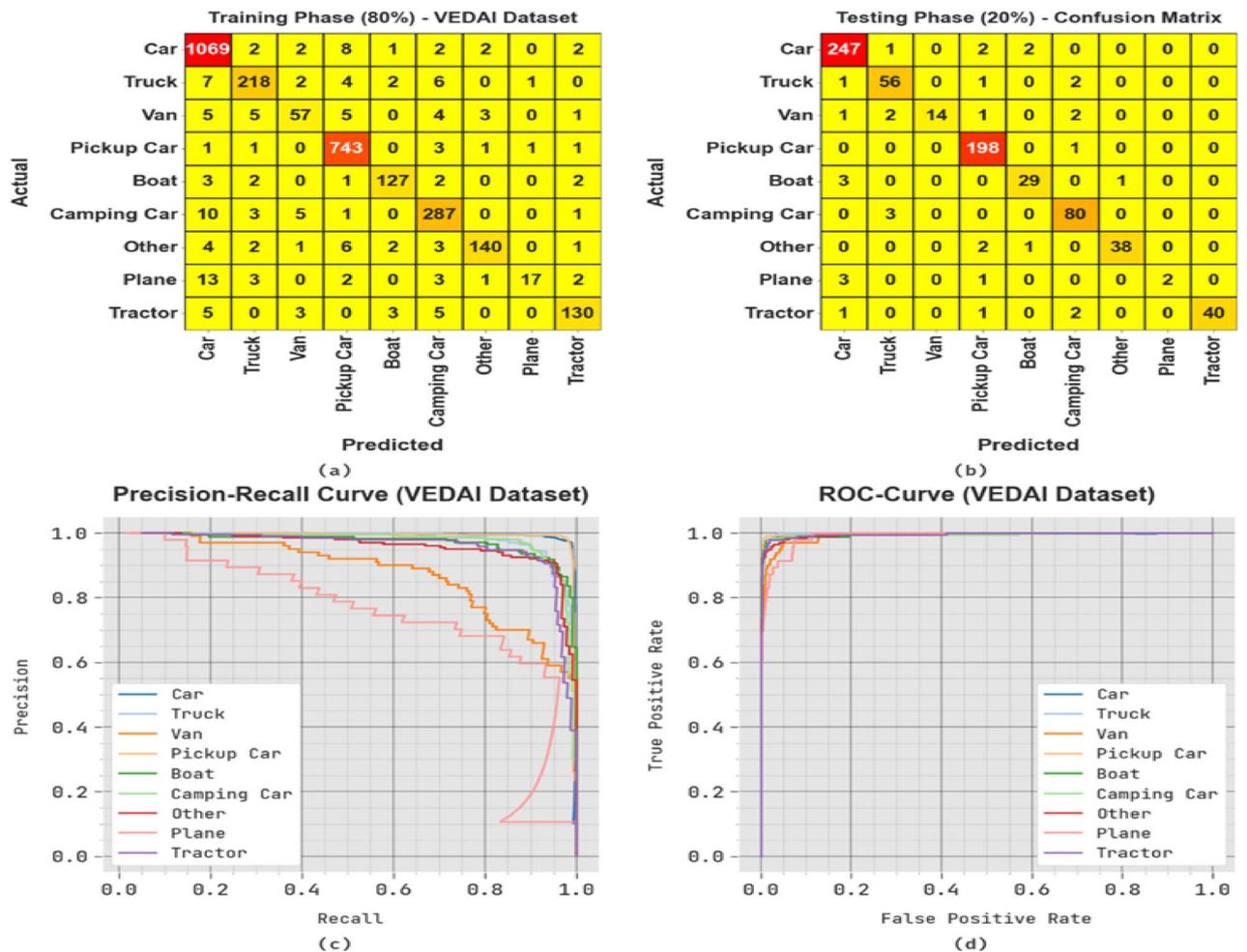
In Table 2, a brief vehicle detection result from the VDTC-CEOADL technique is given on the applied VEDAI dataset. The results represented the effectual ability of the VDTC-CEOADL technique in identifying different types of vehicles. For example, with 80% of TRP, the VDTC-CEOADL method obtains an average  $acc_y$  of 98.79%, a  $sens_y$  of 84.89%, a  $spe_y$  of 99.23%, an  $F\_score$  of 87.44%, and an  $AUC\_score$  of 92.06%. At the same time, with 20% TSP, the VDTC-CEOADL method achieves an average  $acc_y$  of 98.98%,  $sens_y$  of 84.67%, a  $spec_y$  of 99.35%,  $F\_score$  of 88.08%, and  $AUC\_score$  of 92.01%.

Figure 4 shows the precision of the VDTC-CEOADL method during the training and validation process on the VEDAI dataset. The figure specifies that the VDTC-CEOADL method reaches increasing accuracy values over increasing epochs. Additionally, the increase in validation accuracy over training accuracy shows that the VDTC-CEOADL method learns efficiently on the VEDAI data set.

The loss analysis of the VDTC-CEOADL approach at the time of training and validation is shown in the VEDAI data set in Figure 5. The results highlighted by the VDTC-CEOADL method reach closer values of training loss and validation. The VDTC-CEOADL method learns efficiently on the VEDAI dataset.

Particulars	VEDAI dataset	ISPRS potsdam dataset
Purpose	Designed specifically for small object detection (vehicles) in aerial imagery	Urban semantic segmentation and object classification, including vehicles
Data size	3687 cropped vehicle samples from aerial images	2244 annotated vehicle instances extracted from 38 large tiles
Spatial resolution	12.5 cm/pixel	5 cm/pixel
Image type	RGB + IR (infrared)	True orthophoto with 4-band imagery (RGB + NIR)
Scene type	Rural, urban, and industrial zones	Dense urban scenes with diverse infrastructure
Object scale	Small object instances (vehicles ~ 20–50 pixels)	Small to medium vehicle sizes with occlusion and overlap
Annotations	Bounding boxes per object (class-labeled)	Pixel-level semantic masks with object IDs and classes
Vehicle classes	9: Car, Truck, Van, Pickup Car, Boat, Camping Car, Other, Plane, Tractor	4: Car, Truck, Van, Pickup Car
Relevance	Ideal for training models for fine-grained vehicle type classification under varying environments	Represents complex urban environments with dense traffic and shadows—ideal for real-world deployment evaluation
Preprocessing applied	Cropping to fixed-size patches, normalization, data augmentation (rotation, flipping, brightness), class balancing	Patch-based tiling from large orthophotos, normalization, data augmentation, class balancing

**Table 1.** Dataset description.

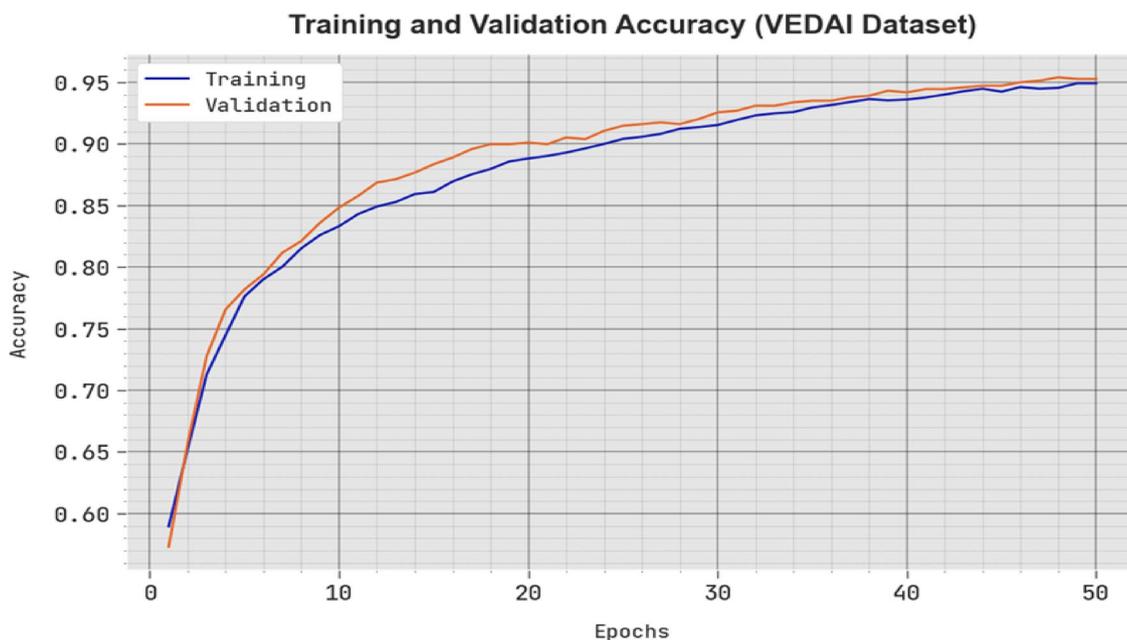


**Fig. 3.** Classifier result in VEDAI data set (a, b) Confusion matrices, (c) PR curve and (d) ROC curve.

Figure 6 shows the classifier results of the VDTC-CEOADL approach in the ISPRS Potsdam dataset. Figure 6a-b describe the confusion matrix rendered by the VDTC-CEOADL technique. The results indicate that the AOAAE-SA approach has detected the four distinct classes of vehicle detection. In the same way, Figure 6c demonstrates the PR analysis of the VDTC-CEOADL model. The figures stated that the VDTC-CEOADL technique has gained a higher PR performance in all classes. Eventually, Figure 6d exemplifies the ROC analysis of the VDTC-CEOADL technique. The figure exposed that the VDTC-CEOADL approach has resulted in proficient results with higher ROC values under distinct class labels.

Class	Accu <sub>y</sub>	Sens <sub>y</sub>	Spec <sub>y</sub>	F <sub>score</sub>	AUC <sub>score</sub>
Training phase (80%)					
Car	97.73	98.25	97.42	96.96	97.84
Truck	98.64	90.83	99.34	91.60	95.08
Van	98.78	71.25	99.55	76.00	85.40
Pickup car	98.81	98.93	98.77	97.70	98.85
Boat	99.39	92.70	99.72	93.38	96.21
Camping car	98.37	93.49	98.94	92.28	96.21
Other	99.12	88.05	99.75	91.50	93.90
Plane	99.12	41.46	99.93	56.67	70.70
Tractor	99.12	89.04	99.64	90.91	94.34
Average	98.79	84.89	99.23	87.44	92.06
Testing phase (20%)					
Car	98.10	98.02	98.15	97.24	98.08
Truck	98.64	93.33	99.12	91.80	96.22
Van	99.19	70.00	100.00	82.35	85.00
Pickup car	98.78	99.50	98.52	97.78	99.01
Boat	99.05	87.88	99.57	89.23	93.73
Camping car	98.64	96.39	98.93	94.12	97.66
Other	99.46	92.68	99.86	95.00	96.27
Plane	99.46	33.33	100.00	50.00	66.67
Tractor	99.46	90.91	100.00	95.24	95.45
Average	98.98	84.67	99.35	88.08	92.01

**Table 2.** Vehicle detection results of the VDTC-CEOADL algorithm in VEDAI dataset.



**Fig. 4.** Accuracy curve of the VDTC-CEOADL algorithm in the VEDAI dataset.

In Table 3, a brief vehicle detection result from the VDTC-CEOADL method is given on the applied ISPRS Potsdam data set. The figure pointed out the effective ability of the VDTC-CEOADL method in identifying different types of vehicles. For example, with 80% of TRP, the VDTC-CEOADL method attains an average accu<sub>y</sub> of 98.47%, a sens<sub>y</sub> of 78.45%, a spec<sub>y</sub> of 95.49%, an F<sub>score</sub> of 83.53% and an AUC<sub>score</sub> of 86.97%. Simultaneously, with 20% of TSP, the VDTC-CEOADL approach acquires an average accu<sub>y</sub> of 99.11%, a sens<sub>y</sub> of 73.09%, a spe<sub>y</sub> of 97.32%, a F<sub>score</sub> of 79.58%, and an AUC<sub>score</sub> of 85.20%.

Figure 7 inspects the accuracy of the VDTC-CEOADL approach during the training and validation process on the ISPRS Potsdam data set. The figure notified that the VDTC-CEOADL approach reaches increasing

### Training and Validation Loss (VEDAI Dataset)

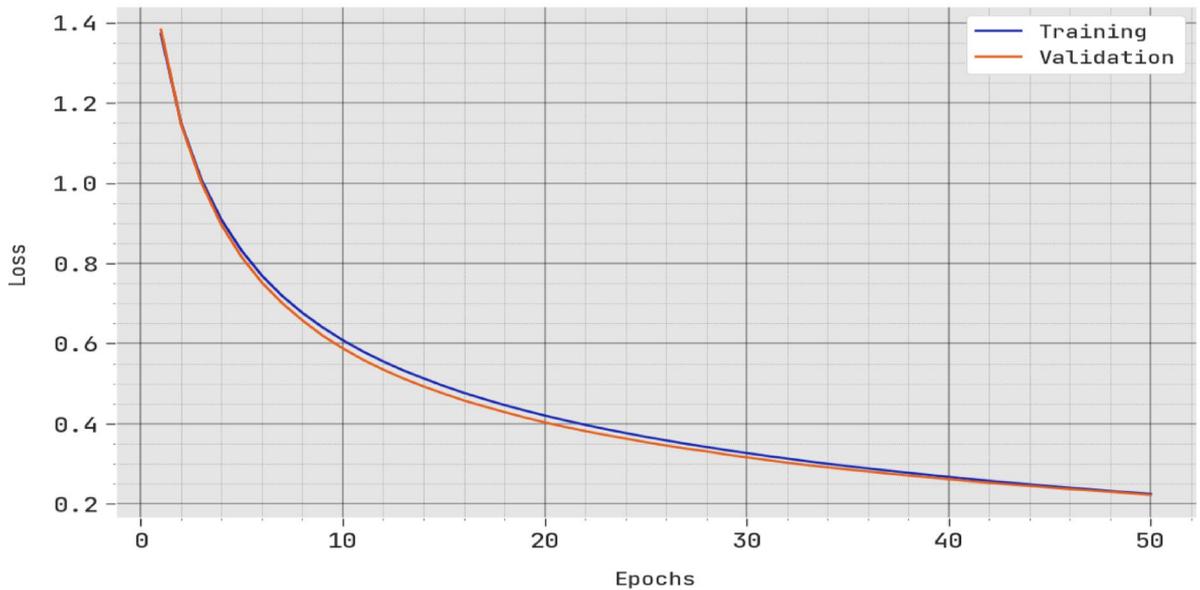


Fig. 5. Loss curve of the VDTC-CEOADL algorithm in VEDAI dataset.

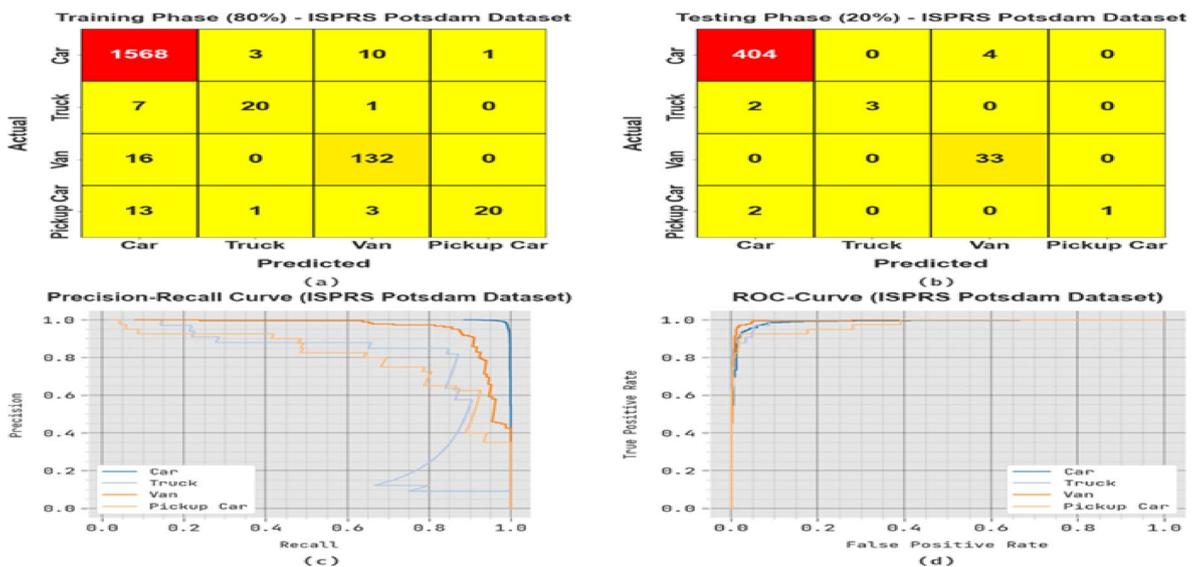


Fig. 6. Classifier result in the ISPRS Potsdam dataset (a, b) Confusion matrices, (c) PR curve, and (d) ROC curve.

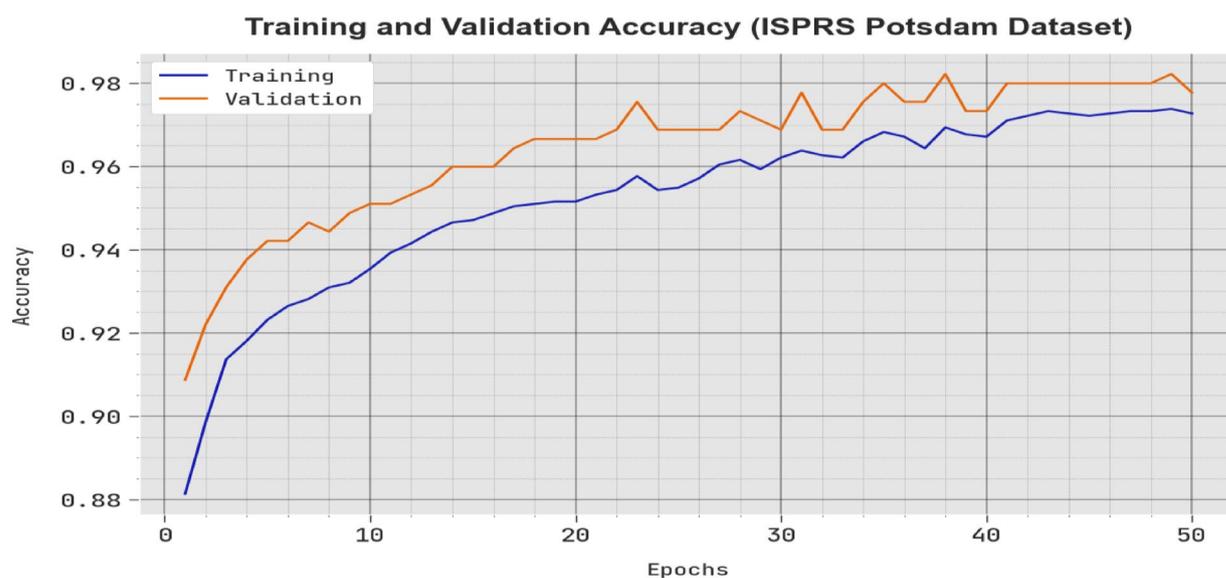
accuracy values over increasing epochs. Moreover, the increasing validation accuracy over training accuracy shows that the VDTC-CEOADL approach learns efficiently on the ISPRS Potsdam dataset. The loss analysis of the VDTC-CEOADL algorithm at the time of training and validation is demonstrated on the ISPRS Potsdam dataset in Figure 8. The results specify that the VDTC-CEOADL technique reaches closer values of training loss and validation loss. The VDTC-CEOADL technique learns efficiently on the ISPRS Potsdam dataset.

The accurate inspection of the VDTC-CEOADL technique is compared with recent DL methods in Table 4; Figure 9<sup>29,30</sup>. The results indicate that the LeNet and VGG-19 models produced poor results with lower acc<sub>y</sub> values of 96.56% and 97.06%, respectively. On the other hand, the TFC and AMS-DAT techniques have accomplished moderately closer results with acc<sub>y</sub> values of 98.82% and 98.01% respectively. However, the VDTC-CEOADL technique gains maximum performance with an acc<sub>y</sub> of 99.11%.

A detailed comparative computation time (CT) assessment of the VDTC-CEOADL method with other DL methods is carried out in Table 5; Figure 10. The results obtained indicated the worse results of the VGG-19 and LeNet techniques with significantly higher CT of 8.73 and 6.89, respectively.

Class	Accu <sub>y</sub>	Sens <sub>y</sub>	Spec <sub>y</sub>	F <sub>score</sub>	AUC <sub>score</sub>
Training phase (80%)					
Car	97.21	99.12	83.10	98.43	91.11
Truck	99.33	71.43	99.77	76.92	85.60
Van	98.33	89.19	99.15	89.80	94.17
Pickup car	99.00	54.05	99.94	68.97	77.00
Average	98.47	78.45	95.49	83.53	86.97
Testing phase (20%)					
Car	98.22	99.02	90.24	99.02	94.63
Truck	99.55	60.00	100.00	75.00	80.00
Van	99.11	100.00	99.04	94.29	99.52
Pickup car	99.55	33.33	100.00	50.00	66.67
Average	99.11	73.09	97.32	79.58	85.20

**Table 3.** Vehicle detection results of the VDTC-CEOADL algorithm in ISPRS Potsdam dataset.



**Fig. 7.** Accuracy curve of the VDTC-CEOADL algorithm in the ISPRS Potsdam dataset.

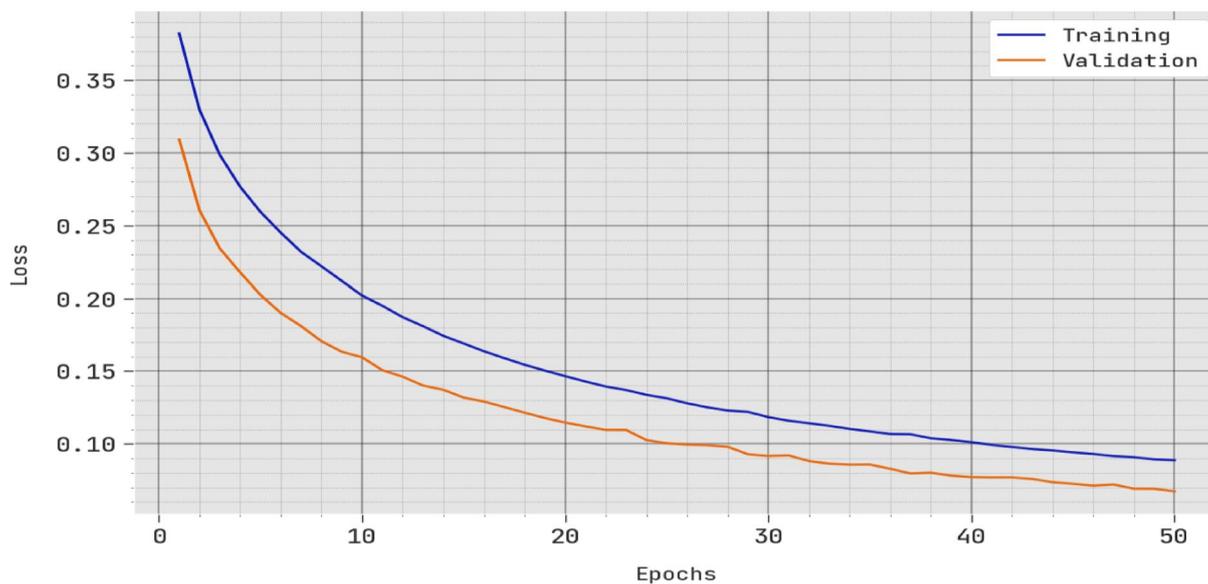
The TFC and AMS-DAT techniques have gained considerably decreased CT of 5.31s and 6.69s, respectively. But the VDTC-CEOADL technique reaches superior results with a CT of less than 3.93<sup>31–33</sup>. These results confirmed the supremacy of the VDTC-CEOADL technique in the vehicle detection and classification process.

The Chaotic Equilibrium Optimization Algorithm (CEOA) creates an expensive computational challenge that affects its hyperparameter tuning stage. This single offline operation requires notable GPU resources and processing time especially when handling big high-resolution imagery. The successful implementation of the system depends heavily on extensive amounts of well-annotated training data because they enable accurate detection and classification performance. The process of manually labeling remote sensing imagery requires extensive human effort and time because it becomes especially difficult to label small and hidden vehicle objects. The challenges with these limitations require future research on both semi-supervised learning and weakly supervised annotation methods to decrease dependency on complete labeled datasets. We are exploring the implementation of lightweight versions of the proposed model to handle resource-limited environments during actual time deployment processes.

## Conclusion

In this study, we have introduced a new VDTC-CEOADL approach for the identification and classification of vehicles on high-resolution RSIs. The VDTC-CEOADL technique presented examined high-quality RSIs to precisely detect and classify different types of vehicle. The VDTC-CEOADL technique comprises several subprocesses namely YOLO-HR based object detector, CEOA based hyperparameter tuning, and ALSTM based vehicle classification. The utilization of the CEOA technique helps in the optimal hyperparameter tuning of the YOLO-HR model, which resulted in enhanced detection performance. The performance validation of the

### Training and Validation Loss (ISPRS Potsdam Dataset)

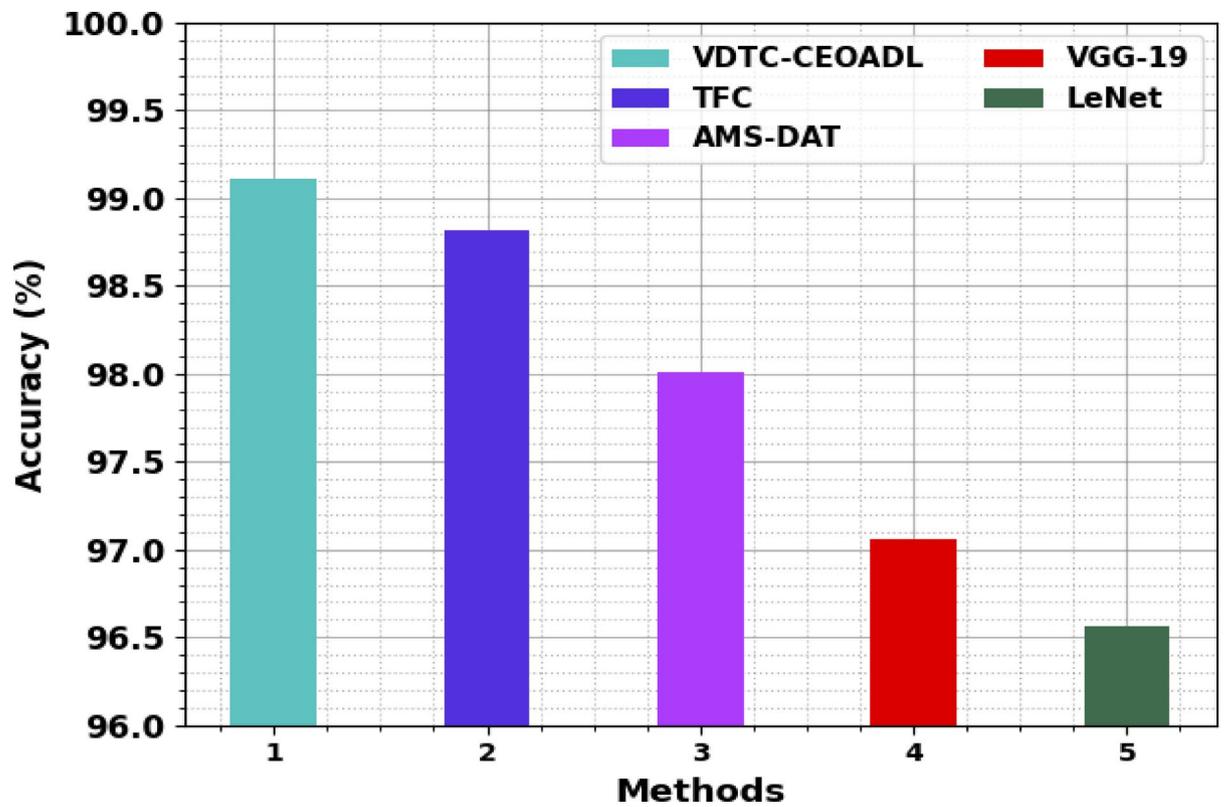


**Fig. 8.** Loss curve of the VDT-C-CEOADL algorithm in the ISPRS Potsdam dataset.

Methods	Accuracy (%)
VDT-C-CEOADL	99.11
TFC	98.82
AMS-DAT	98.01
VGG-19	97.06
LeNet	96.56

**Table 4.** Accuracy analysis of the VDT-C-CEOADL approach with recent DL systems.

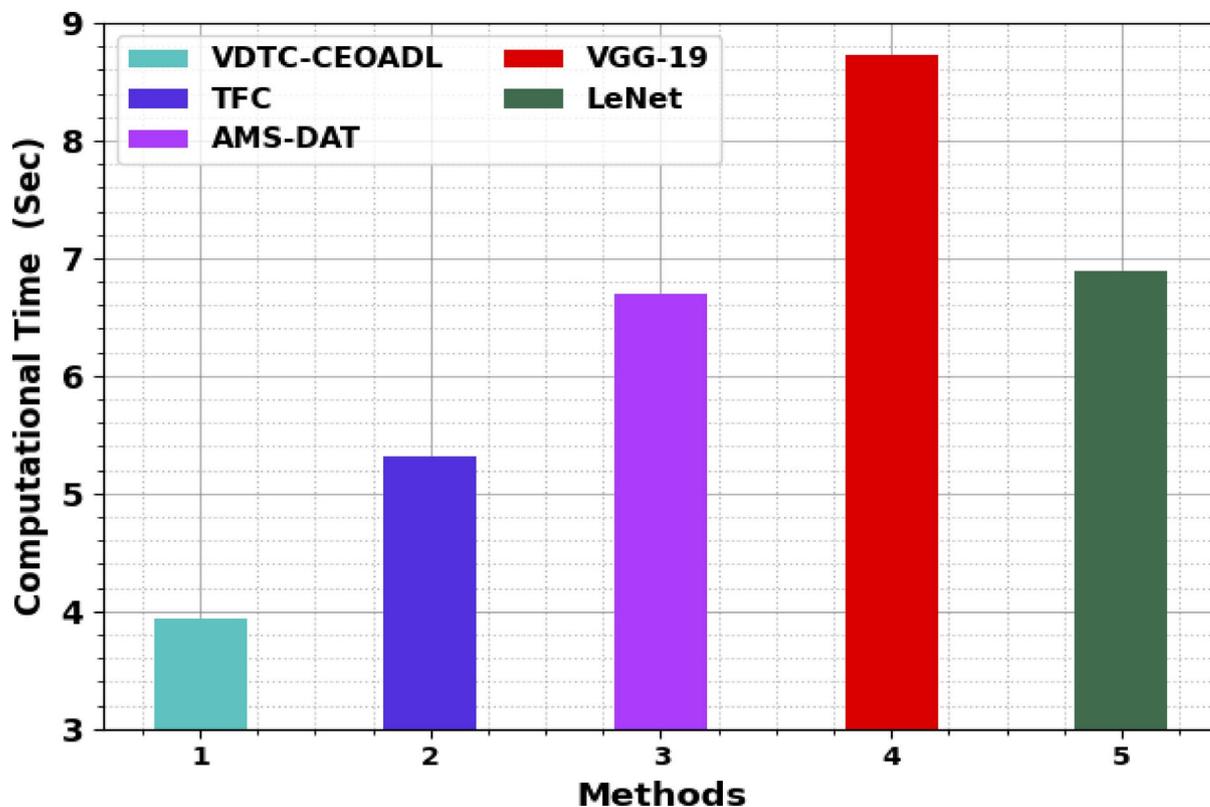
VDT-C-CEOADL technique is validated on a high-resolution RSI dataset and the results portrayed the supremacy of the VDT-C-CEOADL technique in terms of different measures.



**Fig. 9.** Accuracy Analysis of VDTc-CEOADL Approach with recent DL systems.

Methods	Computational time (s)
VDTc-CEOADL	3.93
TFC	5.31
AMS-DAT	6.69
VGG-19	8.73
LeNet	6.89

**Table 5.** CT analysis of the VDTc-CEOADL approach with recent DL systems.



**Fig. 10.** CT analysis of the VDT-C-EOADL approach with recent DL systems.

### Data availability

Data Availability Statement: Available Based on Request. The datasets generated and/or analyzed during the current study are not publicly available due to the extension of the submitted research work. They are available from the corresponding author upon reasonable request.

Received: 24 August 2024; Accepted: 13 May 2025

Published online: 23 May 2025

### References

- Mu, C. Y., Kung, P., Chen, C. F. & Chuang, S. C. Enhancing Front-Vehicle detection in large vehicle fleet management. *Remote Sens.* **14** (7), 1544 (2022).
- Shen, J., Liu, N., Sun, H., Tao, X. & Li, Q. Vehicle detection in aerial images based on hyper feature map in deep convolutional network. *KSII Trans. Internet Inform. Syst. (TIIS)*. **13** (4), 1989–2011 (2019).
- Javadi, S., Dahl, M. & Pettersson, M. I 2021 vehicle detection in aerial images based on 3D depth maps and deep neural networks. *IEEE Access.* **9**: 8381–8391.
- Kim, K. J., Kim, P. K., Chung, Y. S. & Choi, D. H. Multi-scale detector for accurate vehicle detection in traffic surveillance data. *IEEE Access.* **7**, 78311–78319 (2019).
- Mo, W. et al. Y 2023 PVDet: towards pedestrian and vehicle detection on gigapixel-level images. *Eng. Appl. Artif. Intell.* **118**: 105705.
- Zhang, R. et al. D 2021 Multi-scale adversarial network for vehicle detection in UAV imagery. *ISPRS J. Photogrammetry Remote Sens.* **180**: 283–295.
- Tak, S., Lee, J. D., Song, J. & Kim, S. Development of AI-based vehicle detection and tracking system for C-ITS application. *J. Adv. Transp.* 1–15. (2021).
- Djenouri, Y., Belhadi, A., Srivastava, G., Djenouri, D. & Lin, J. C. W. Vehicle detection using improved region Convolution neural network for accident prevention in smart roads. *Pattern Recognit. Lett.* **158**, 42–47 (2022).
- Shen, J., Liu, N., Sun, H. & Zhou, H. Vehicle detection in aerial images based on lightweight deep convolutional network and generative adversarial network. *IEEE Access.* **7**, 148119–148130 (2019).
- Yu, L. et al. Small-Sized vehicle detection in remote sensing image based on key point detection. *Remote Sens.* **13** (21), 4442 (2021).
- Song, H., Liang, H., Li, H., Dai, Z. & Yun X 2019 Vision-based vehicle detection and counting system using deep learning in highway scenes. *Eur. Transp. Res. Rev.* **11**: 1–16.
- Rafique, A. A. et al. J 2023 Smart Traffic Monitoring Through Pyramid Pooling Vehicle Detection and Filter-based Tracking on Aerial Images. *IEEE Access.*
- Singh, G. H., Ibrahim Khalaf, O., Alotaibi, Y., Alghamdi, S. & Alassery, F. Multi-model CNN-RNN-LSTM based fruit recognition and classification. *Intell. Autom. Soft Comput.* **33** (1), 637–650 (2022).
- Charouh, Z., Ezzouhri, A., Ghogho, M. & Guennoun Z 2022 A resource-efficient CNN-based method for moving vehicle detection. *Sensors* **22**(3): 1193.
- Qiu, Z., Bai, H. & Chen, T. Special vehicle detection from UAV perspective via YOLO-GNS based deep learning network. *Drones* **7** (2), 117 (2023).

16. Yang, B., Zhang, S., Tian, Y. & Li, B. Front-vehicle detection in video images based on Temporal and Spatial characteristics. *Sensors* **19** (7), 1728 (2019).
17. Zhou, L. et al. Y 2021 Vehicle Detection in Remote Sensing Image Based on Machine Vision. Computational intelligence and neuroscience.
18. Gomaa, A. & Saad, O. M. Residual Channel-attention (RCA) network for remote sensing image scene classification. *Multimedia Tools Appl.*, 1–25. (2025).
19. Gomaa, A. Advanced domain adaptation technique for object detection leveraging semi-automated dataset construction and enhanced yolov8. In 2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES) (pp. 211–214). IEEE. (2024), October.
20. Gomaa, A., Abdelwahab, M. M. & Abo-Zahhad, M. Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. *Multimedia Tools Appl.* **79** (35), 26023–26043 (2020).
21. Gomaa, A. & Abdalrazik, A. Novel deep learning domain adaptation approach for object detection using semi-self Building dataset and modified yolov4. *World Electr. Veh. J.* **15** (6), 255 (2024).
22. Alotaibi, Y., Rajendran, B. & Rajendran, S. Dipper throated optimization with deep convolutional neural network-based crop classification for remote sensing image analysis. *PeerJ Comput. Sci.* **10**, e1828 (2024).
23. Wan, D. et al. X 2023 YOLO-HR: improved YOLOv5 for object detection in High-Resolution optical remote sensing images. *Remote Sens.* **15**(3): 614.
24. Mohamed, B., Mohaisen, L. & Amin, M. Binary Equilibrium Optimization Algorithm for Computing Connected Domination Metric Dimension Problem. *Scientific Programming.* (2022).
25. Sundarapandi, A. M. S., Alotaibi, Y., Thanarajan, T. & Rajendran, S. Archimedes optimisation algorithm quantum dilated convolutional neural network for road extraction in remote sensing images. *Heliyon*, **10**(5). (2024).
26. Chen, H., Gao, P., Tan, S., Yuan, H. & Guan, M. Prediction of Automatic Scram during Abnormal Conditions of Nuclear Power Plants Based on Long Short-Term Memory (LSTM) and Dropout. *Science and Technology of Nuclear Installations*, 2023. (2023).
27. Razakarivony, S. & Jurie, F. Vehicle detection in aerial imagery: A small target detection benchmark. *J. Vis. Common. Image Represent.* **34**, 187–203 (2016).
28. Rottensteiner, F. et al. The ISPRS benchmark on urban object classification and 3D Building reconstruction. *ISPRS Ann. Photogramm Remote Sens. Spat. Inf. Sci.* **1**, 293–298 (2012).
29. Tamilvizhi, T., Surendran, R. & Krishnaraj, N. Cloud based smart vehicle tracking system. In 2021 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, United Kingdom, 1–6. IEEE, (2021).
30. Chen, R., Ferreira, V. G. & Li, X. Detecting moving vehicles from Satellite-Based videos by tracklet feature classification. *Remote Sens.* **15**, 34. <https://doi.org/10.3390/rs15010034> (2023).
31. Purnachand, K. et al. A 2022 A new method for scene classification from the remote sensing images. *CMC-Comput Mater. Contin.* **72**: 1339–1355.
32. Chen, H., Gao, P., Tan, S., Yuan, H. & Guan, M. *Prediction of Automatic Scram during Abnormal Conditions of Nuclear Power Plants Based on long Short-Term Memory (LSTM) and Dropout* (Science and Technology of Nuclear Installations, 2023).
33. Xiaoling, L., Chen, X. & Duansong, W. Adaptive neural network control for marine surface vehicles platoon with input saturation and output constraints. *AIMS Math.* **5** (1), 587–602. <https://doi.org/10.3934/math.2020039> (2020).

## Acknowledgements

The authors extend their appreciation to Umm Al-Qura University, Saudi Arabia for funding this research work through grant number: 25UQU4281768GSSR01.

## Author contributions

“Y.A. and K.N. wrote the main manuscript text and T.T. and S.R prepared all figures and Table. All authors reviewed the manuscript.”

## Funding

This research work was funded by Umm Al-Qura University, Saudi Arabia under grant number: 25UQU4281768GSSR01.

## Declarations

## Competing interests

The authors declare no competing interests.

## Research involving human participants and/or animals

No human or animals are involved in this study.

## Additional information

**Correspondence** and requests for materials should be addressed to T.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025