

Article

Distinguishing Malicious Drones Using Vision Transformer

Sonain Jamil ^{1,*} , Muhammad Sohail Abbas ² and Arunabha M. Roy ^{3,*} 

¹ Department of Electronics Engineering, Sejong University, Seoul 05006, Korea

² School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; mabbas.mscs20seecs@seecs.edu.pk

³ Aerospace Engineering Department, University of Michigan, Ann Arbor, MI 48109, USA

* Correspondence: sonainjamil@sju.ac.kr (S.J.); arunabhr@umich.edu or arunabhr.umich@gmail.com (A.M.R.)

Abstract: Drones are commonly used in numerous applications, such as surveillance, navigation, spraying pesticides in autonomous agricultural systems, various military services, etc., due to their variable sizes and workloads. However, malicious drones that carry harmful objects are often adversely used to intrude restricted areas and attack critical public places. Thus, the timely detection of malicious drones can prevent potential harm. This article proposes a vision transformer (ViT) based framework to distinguish between drones and malicious drones. In the proposed ViT based model, drone images are split into fixed-size patches; then, linearly embeddings and position embeddings are applied, and the resulting sequence of vectors is finally fed to a standard ViT encoder. During classification, an additional learnable classification token associated to the sequence is used. The proposed framework is compared with several handcrafted and deep convolutional neural networks (D-CNN), which reveal that the proposed model has achieved an accuracy of 98.3%, outperforming various handcrafted and D-CNNs models. Additionally, the superiority of the proposed model is illustrated by comparing it with the existing state-of-the-art drone-detection methods.

Keywords: vision transformer; deep convolutional neural networks; deep learning; malicious drones; classification; drones



Citation: Jamil, S.; Abbas, M.S.; Roy, A.M. Distinguishing Malicious Drones Using Vision Transformer. *AI* **2022**, *3*, 260–273. <https://doi.org/10.3390/ai3020016>

Academic Editor: Rafał Dreżewski

Received: 6 March 2022

Accepted: 29 March 2022

Published: 31 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

With recent development in remote sensing technology, drones play an essential role in developing smart cities and innovative industries due to their numerous applications, including automated irrigation, spraying pesticides and fertilizers in agriculture [1], water management [2], food services [3], UAS-based image velocimetry [4], flying base stations [5], etc. Drones of variable sizes and different shapes have been deployed in the military for navigation and surveillance purposes [6].

In spite of numerous useful applications, drones are often used for spying and carrying dangerous loads. Such drones are termed malicious drones, which enter in the restricted non-fly zones avoiding radar detection due to their low-altitude flight path. The schematic in Figure 1a,b shows the normal use cases of drones, and Figure 1c depicts the intrusion of malicious drones in the restricted zones.

Hence, it is critical to develop an autonomous system that can efficiently detect the intrusion of malicious drones to avoid any potential damage. In that regard, machine learning (ML) and computer vision (CV) can allow us to develop automated systems that can detect malicious drones. The existing techniques in the literature usually rely on audios, images, videos, and radio frequency signals to detect drones. In [7], the authors proposed a DL-based hybrid audio and integrated visual framework for detecting malicious drones, which achieved an accuracy of 98.5% for the combined audio and visual dataset. However, the main drawback of the model was that it was limited to drone detection, and the model was unable to differentiate between drones with loads and without loads. Along similar lines, in [8], the authors proposed a mel frequency cepstral coefficient (MFCC)

with a SVM-based model for detecting malicious drones; however, the performance model deteriorates when detecting amateur drones in adverse weather conditions and noisy environments. Moreover, in [9], the authors proposed a handcrafted feature extraction-based technique to detect drones using audios and images. The method achieved 81% accuracy, but deteriorates when detecting drones in adverse weather conditions. In [10], Dumitrescu et al. designed a DL-based system for drone detection by employing acoustic signals. However, the authors did not consider malicious drones as a separate class and the article only addressed drone detection. In [11], Digulescu et al. investigated a radio frequency signal-based advanced signal processing model to detect the movement of drones. The model performed relatively well in the controlled environment. In [12], Singha et al. proposed a YOLOv4-based model for detecting drones. The model achieved mean average precision (mAP) of 74.36%. Furthermore, in [13], the authors proposed a DL-based detection and identification of the drones using audio signals. The technique achieved the highest accuracy of 85.26%; however, the model showed limited performance in adverse weather conditions. In [14], the authors distinguished drones from birds using the laser. The framework detected drones with less than five kilograms of mass. However, the technique was not used to detect drones with loads. In a subsequent study [15], the authors proposed a YOLOv3 based model for detecting drones and birds. The model performance varies with the variation in the shape of drones and the visibility of the drones. In [16], Swinney et al. analyzed the impact of real-world interference in the classification of drones, using CNNs and radio frequency signals.

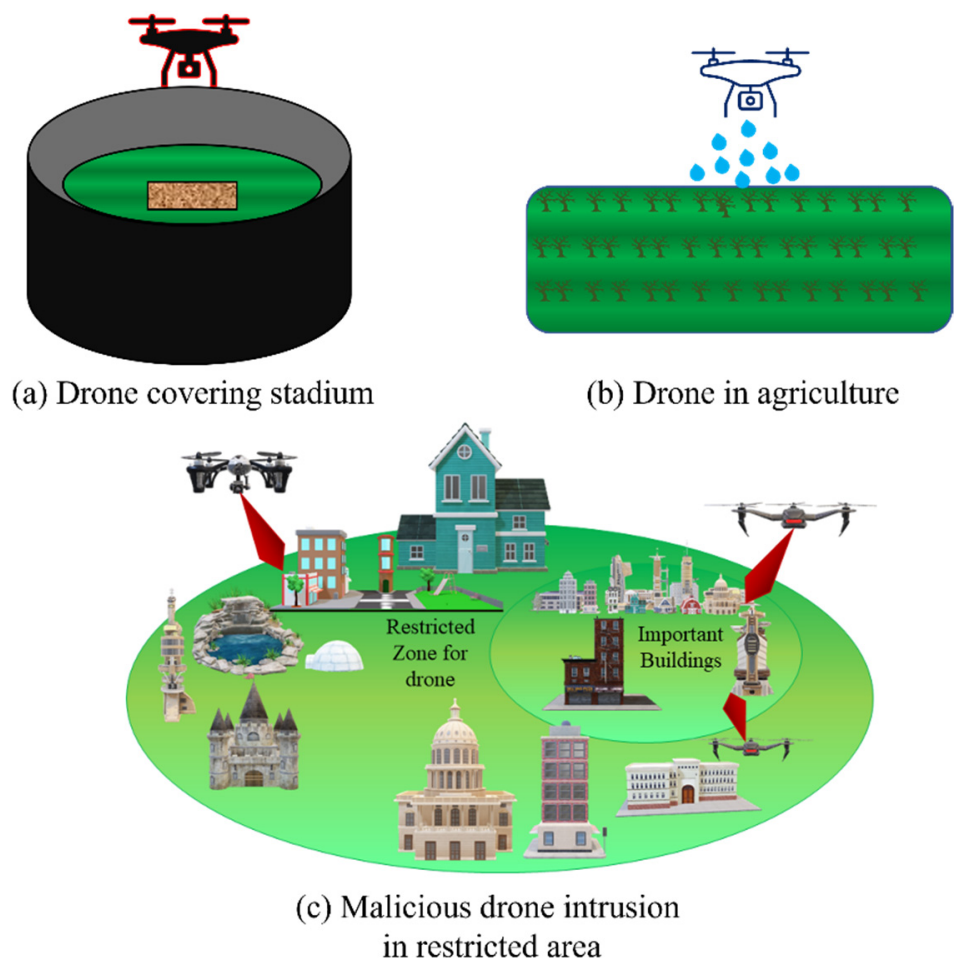


Figure 1. (a,b) Normal use cases of drones; (c) malicious drone intrusion in the restricted areas.

From the aforementioned discussion, it is clear that although several existing DL models can classify and detect drones based on acoustic, radio frequency, and visual signals, they may not be useful in challenging scenarios of distinguishing between several subject classes, such as drones, malicious drones, birds, airplanes, helicopters, etc. Furthermore, none of the existing ML and deep learning (DL) models address the issue of drone detection with loads.

In order to address the aforementioned shortcomings and drawbacks, the current article proposes a vision transformer (ViT) based framework for classifying drones, malicious drones, airplanes, birds, and helicopters. The idea of ViT was introduced by [17]. We compare the proposed framework with various handcrafted feature extraction such as histogram of oriented gradient (HOG) [18], locally encoded transform feature histogram (LETRIST) [19], local binary pattern (LBP) [20], gray level co-occurrence matrix (GLCM) [21], non-redundant local binary pattern (NRLBP) [22], completed joint-scale local binary pattern (CJLBP) [23], local tetra pattern (LTrP) [24], and D-CNN models, such as AlexNet [25], ShuffleNet [26], ResNet-50 [27], SqueezeNet [28], MobileNet-v2 [29], Inceptionv3 [30], GoogleNet [31], EfficientNetb0 [32], Inception-ResNet-v2 [33], DarkNet-53 [34] and Xception [35]. We also compare the feature extractions' performance with several classifiers, such as support vector machine [36,37], decision tree, k-nearest neighbors, ensemble, Naive Bayes, multi-layer perceptron (MLP) [38,39], radial basis function (RBF) and group method of data handling (GMDH). The comparisons demonstrate that the proposed model can significantly outperform existing state-of-the-art models in terms of classification accuracy and can be employed as a robust classification model for malicious drones' detection. The remainder of the paper is organized as follows: Section 2 describes the proposed methodology, different handcrafted descriptor models, and dataset description; Section 3 deals with the relevant finding and discussion of the proposed classifier. Finally, the conclusions and prospects of the current work are discussed in Section 4.

2. Proposed Methodology

Drones have different visual characteristics, such as color, shape, load, and size. Thus, the images are useful for distinguishing malicious drones from the other classes, such as drones without load, aeroplanes, helicopters, and birds. The images are fed into the handcrafted descriptors, D-CNNs, and ViT-classifier. Handcrafted descriptors and D-CNNs are used to extract features that are used to train the classifier. The schematic in Figure 2 shows the flow diagram of the framework.

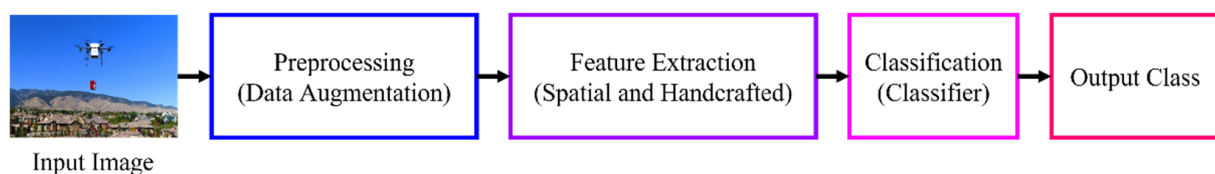


Figure 2. Flow diagram of the proposed methodology.

2.1. Handcrafted Descriptors

The images are resized to 224×224 and after that, features are extracted with the help of HOG, LETRIST, LBP, GLCM, NRLBP, CJLBP, and LTrP. The features are stored in the feature vectors, which are used to train ML classifiers.

2.2. D-CNN Models

The images are resized to the input size of each D-CNNs and after that, features are extracted with the help of AlexNet, ShuffleNet, ResNet-50, SqueezeNet, MobileNet-v2, Inceptionv3, GoogleNet, EfficientNetb0, Inception-ResNet-v2, DarkNet-53, and Xception. The features are saved in the feature vectors, which are used to train ML classifiers.

2.3. ViT-Based Classification

Initially, the images are resized to 224×224 and then fed into ViT. ViT splits images into 14×14 vectors with patches of 16×16 . These patch embedding vectors are followed by adding learnable position embedding vectors. These embedded vectors are further fed into the transformer encoder (TE), which is proposed in [40]. In TE, the embedded vectors are divided into a query (a), key (b), and value (c) after being expanded by a fully connected (fc) layer. Then, a, b, and c are further divided and fed to the parallel attention heads (AH). Outputs from AHs are concatenated to form the vectors whose shape is the same as the encoder input. The vectors go through an fc, a layer normalization, and a multi-layer perceptron MLP block with two fc layers. TE encodes the embedding vector and outputs a vector of the same size. The output vector of the TE is fed into the MLP head to make the final classification. The complete schematic diagram of the ViT is shown in Figure 3.

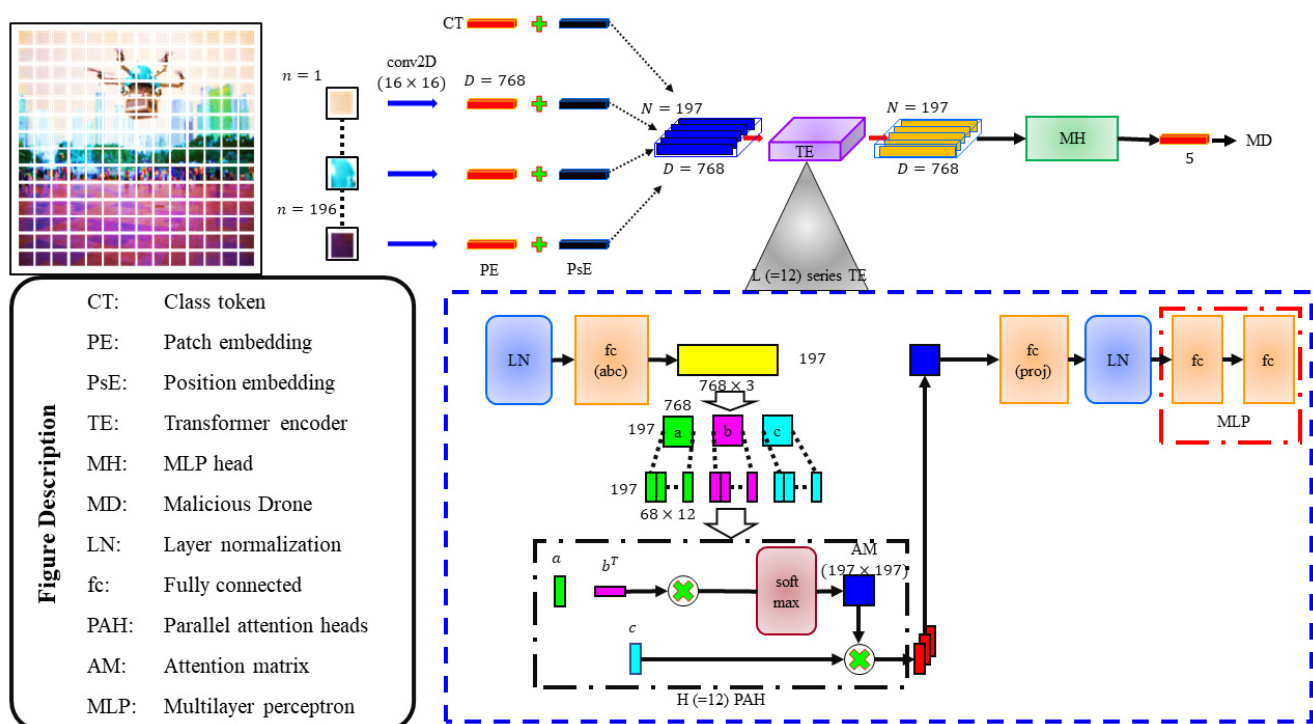


Figure 3. Schematic diagram of the ViT classifier.

2.4. Dataset

In the present study, a customized dataset consisting of five different classes (i.e., aeroplanes, birds, drones, helicopters, and malicious drones) is utilized. The dataset is challenging due to the presence of occluded images, night images, low visibility of object images, and adverse weather condition images. The dataset has a total of 776 images. The aeroplane and bird classes have 105 images in each class. Similarly, the drone, helicopter, and malicious drone classes have 200, 167, and 199 images, respectively. All the images are resized to 224×224 . The dataset is publicly available on Kaggle, and the link can be found in the data availability section. The dataset is divided into a train set with 70% images and a test set with 30% images. Some of the typical images from the dataset are shown in Figure 4.

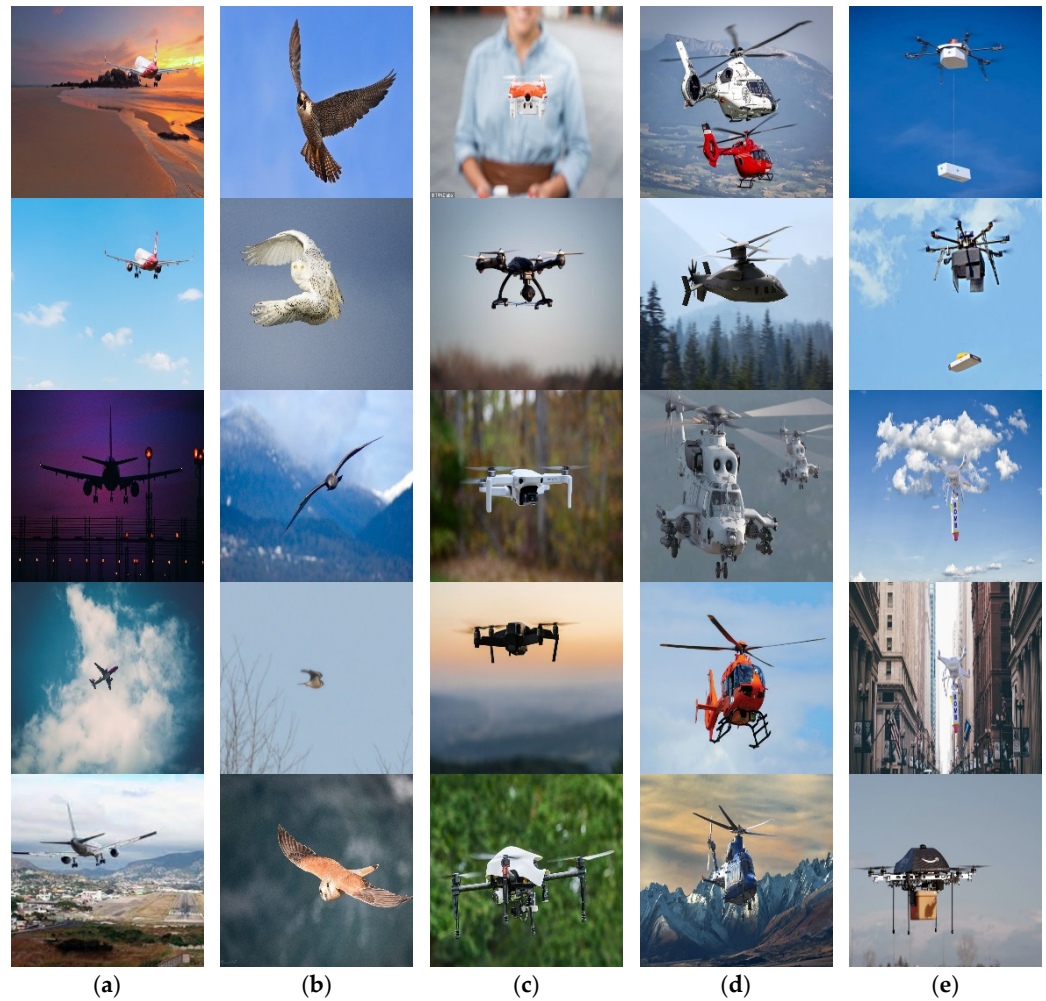


Figure 4. Sample images from the custom dataset for five different classes: (a) aeroplane; (b) bird; (c) drone; (d) helicopter; and (e) malicious drone.

3. Results

In order to evaluate the performance of the proposed classifier, various performance metrics, including accuracy, specificity, sensitivity, and F_1 – score are considered. The accuracy of the classifier can be obtained as follows:

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (1)$$

where, in Equation (1), t_n and t_p denote true negative and true positive, respectively, while f_n and f_p represent false negative and false positive, respectively. The accuracy of the classifier indicates the ability to distinguish malicious drone classes correctly. Sensitivity (s_e) is the proportion of actual positives that are correctly predicted as positives and is determined as

$$s_e = \frac{t_p}{t_p + f_n} \quad (2)$$

Precision or specificity (s_p) is the proportion of actual positives that are correctly predicted as negatives and is calculated as follows:

$$s_p = \frac{t_n}{t_n + f_p} \quad (3)$$

From the definition of s_e and s_p in Equations (2) and (3), the F_1 – score can be obtained as

$$F_1 - score = 2 \times \left[\frac{(s_e * s_p)}{(s_e + s_p)} \right] \quad (4)$$

Additionally, Cohen’s kappa (κ) is considered to further evaluate the performance of the proposed model, which can be calculated as

$$\kappa = 2 \times \left[\frac{(t_p \cdot t_n - f_p \cdot f_n)}{(t_p + f_p) \cdot (f_p + t_n) + (t_p + f_n) \cdot (f_n + t_n)} \right] \quad (5)$$

The experiments are conducted on the local system with 12 GB RAM and Tesla T4 GPU. The model complexity and hyperparameters of the model are shown in Table 1.

Table 1. Model complexity and hyperparameters.

Parameter	Value
Trainable Parameters	85.8 M
Model Parameters size	171.605 MB
Learning rate	2×10^{-5}
Optimizer	Adam
Mini Batch Size	8

From the classification result, it is found that the proposed ViT classifier has achieved 98.28% overall accuracy. The accuracy values for aeroplanes, birds, and helicopters are 100%, 100%, and 100%, respectively, indicating excellent robustness of the model for these classes. However, the accuracy values for the drone and malicious drone classes slightly drop to 96.8% and 96.8%, respectively. The confusion matrix of the ViT classifier is shown in Figure 5.

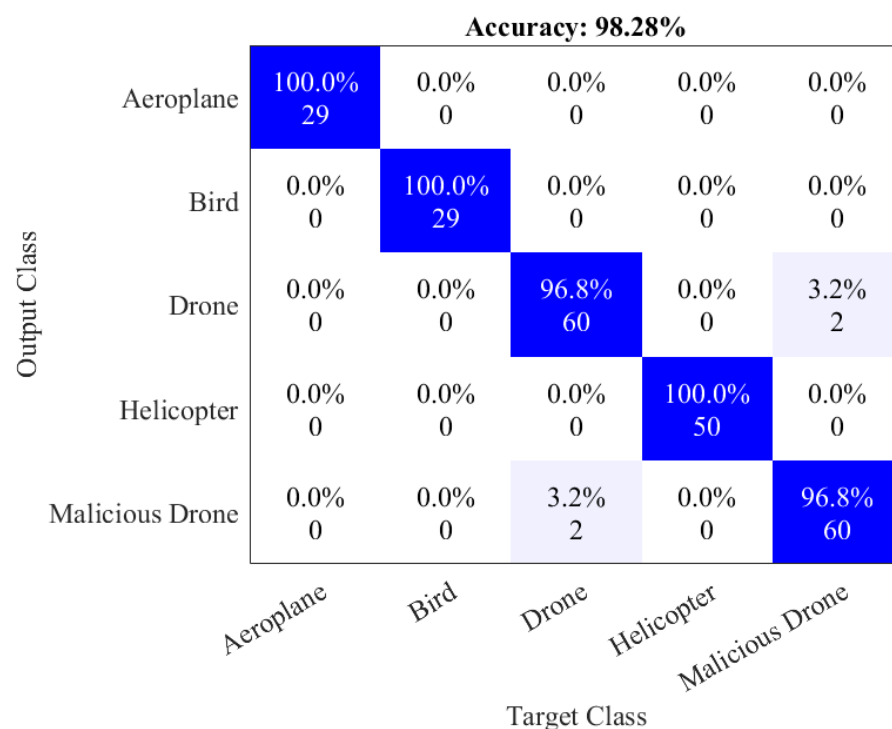


Figure 5. Confusion matrix of the ViT classifier.

The ViT classifier achieves the overall s_e , s_p , F_1 - score, and κ values of 99.00%, 99.00%, 99.00%, and 99.00%, respectively. The s_e , s_p and F_1 - score of aeroplane, bird, and helicopter classes are 100%, 100%, and 100% respectively. The s_e , s_p , and F_1 - score for drone and malicious drone classes are 97.0%, 97.0%, and 97.0%, respectively. Figure 6 shows the comparison bar chart of various classification metrics obtained from the ViT classifier for different classes.

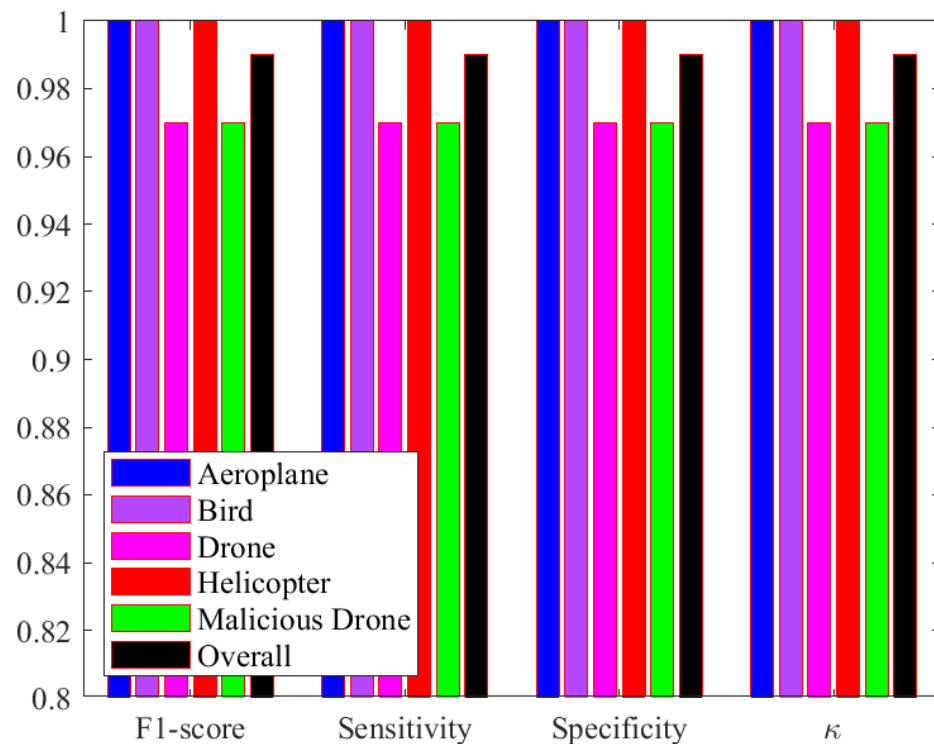


Figure 6. The classification performance metrics for different individual class and overall classes of the ViT classifier.

This section reports the performance comparison of various handcrafted descriptors considering different classifiers. The accuracy of the HOG, LETRIST, LBP, GLCM, NRLBP, CjLBP, and LTrP with different classifiers such as SVM with linear kernel, kNN, DT, Ensemble, NB, MLP, RBF, and GMDH are shown in Table 2.

Analyzing Table 2, it is evident that the performance of handcrafted descriptors is quite low compared to ViT classifier, as the highest accuracy is 78.90% using HOG and ensemble classifier. The accuracy of HOG with the SVM classifier is 76.70% whereas, with kNN, NB and DT, it is 37.90%, 57.30%, and 58.60%, respectively. Similarly, Table 3 shows the test accuracy of the AlexNet, ShuffleNet, ResNet-50, SqueezeNet, MobileNet-v2, Inceptionv3, GoogleNet, EfficientNetb0, Inception-ResNet-v2, DarkNet-53, and Xception models with different classifiers. All the D-CNNs are trained with 1000 epochs and the best number of the epochs are achieved by monitoring the validation accuracy of the models and adding early stopping.

Table 2. Performance comparison of various handcrafted descriptors considering different classifiers.

Descriptor	Classifier	Accuracy
HOG	SVM (Linear Kernel) ¹	76.70%
	kNN ²	37.90%
	DT ³	58.60%
	NB ⁴	57.30%
	Ensemble	78.90%
	MLP	70.50%
	RBF	75.60%
	GMDH	74.50%
LETRIST	SVM (Linear Kernel) ¹	31.90%
	kNN ²	39.70%
	DT ³	36.60%
	NB ⁴	43.50%
	Ensemble	52.20%
	MLP	30.90%
	RBF	32.30%
	GMDH	30.40%
LBP	SVM (Linear Kernel) ¹	45.30%
	kNN ²	38.40%
	DT ³	34.90%
	NB ⁴	39.70%
	Ensemble	45.70%
	MLP	39.10%
	RBF	44.20%
	GMDH	43.10%
GLCM	SVM (Linear Kernel) ¹	49.60%
	kNN ²	36.60%
	DT ³	39.20%
	NB ⁴	34.50%
	Ensemble	44.40%
	MLP	43.40%
	RBF	48.50%
	GMDH	47.40%
NRLBP	SVM (Linear Kernel) ¹	28.00%
	kNN ²	16.80%
	DT ³	30.60%
	Ensemble	30.60%
	MLP	22.00%
	RBF	27.00%
	GMDH	26.00%
CJLBP	SVM (Linear Kernel) ¹	36.20%
	kNN ²	30.20%
	DT ³	38.40%
	NB ⁴	36.60%
	Ensemble	50.90%
	MLP	30.00%
	RBF	35.10%
	GMDH	34.00%
LTrP	SVM (Linear Kernel) ¹	29.70%
	kNN ²	34.10%
	DT ³	37.90%
	NB ⁴	44.80%
	Ensemble	47.80%
	MLP	23.50%
	RBF	28.60%
	GMDH	27.50%

¹ SVM = support vector machine, ² kNN = k nearest neighbor, ³ DT = decision tree, ⁴ NB = naïve Bayes.

Table 3. Performance values in terms of accuracy obtained from different D-CNN models.

D-CNN Model	Classifier	Accuracy
AlexNet	SVM ¹	88.80%
	kNN ²	75.90%
	DT ³	59.50%
	NB ⁴	71.10%
	Ensemble	83.30%
	MLP	82.60%
	RBF	87.70%
	GMDH	86.60%
ShuffleNet	SVM ¹	86.20%
	kNN ²	77.60%
	DT ³	63.80%
	NB ⁴	76.30%
	Ensemble	86.20%
	MLP	80.00%
	RBF	85.10%
	GMDH	84.00%
ResNet-50	SVM ¹	89.20%
	kNN ²	77.20%
	DT ³	73.70%
	NB ⁴	72.80%
	Ensemble	86.60%
	MLP	83.00%
	RBF	88.10%
	GMDH	87.00%
SqueezeNet	SVM ¹	61.60%
	kNN ²	64.20%
	DT ³	66.40%
	NB ⁴	63.80%
	Ensemble	82.80%
	MLP	55.40%
	RBF	60.50%
	GMDH	59.40%
MobileNet-v2	SVM ¹	91.80%
	kNN ²	84.50%
	DT ³	62.90%
	NB ⁴	83.60%
	Ensemble	85.30%
	MLP	85.60%
	RBF	90.70%
	GMDH	89.60%
Inceptionv3	SVM ¹	90.90%
	kNN ²	88.40%
	DT ³	70.70%
	NB ⁴	85.30%
	Ensemble	88.40%
	MLP	84.70%
	RBF	89.80%
	GMDH	88.70%

Table 3. *Cont.*

D-CNN Model	Classifier	Accuracy
GoogleNet	SVM ¹	87.90%
	kNN ²	82.30%
	DT ³	64.20%
	NB ⁴	84.50%
	Ensemble	87.50%
	MLP	83.70%
	RBF	86.80%
	GMDH	85.70%
EfficientNetb0	SVM ¹	92.20%
	kNN ²	84.50%
	DT ³	66.40%
	NB ⁴	86.20%
	Ensemble	89.20%
	MLP	86.00%
	RBF	91.10%
	GMDH	90.00%
Inception-ResNet-v2	SVM ¹	91.80%
	kNN ²	87.90%
	DT ³	72.00%
	NB ⁴	80.20%
	Ensemble	89.70%
	MLP	85.60%
	RBF	90.70%
	GMDH	89.60%
DarkNet-53	SVM ¹	68.50%
	kNN ²	62.50%
	DT ³	75.00%
	NB ⁴	74.60%
	Ensemble	91.40%
	MLP	62.30%
	RBF	67.40%
	GMDH	66.30%
Xception	SVM ¹	93.50%
	kNN ²	87.90%
	DT ³	72.40%
	NB ⁴	87.50%
	Ensemble	88.80%
	MLP	87.30%
	RBF	92.40%
	GMDH	91.30%
Proposed	ViT classifier	98.28%

¹ SVM = support vector machine, ² kNN = k nearest neighbor, ³ DT = decision tree, ⁴ NB = naïve Bayes.

The results in Table 3 indicate that the performance of the D-CNN models is better than handcrafted descriptors. However, the highest accuracy of 93.50% is achieved by Xception with multiclass SVM. The accuracy values for the Xception with kNN, DT, NB, and Ensemble are 87.90%, 72.40%, 87.50%, and 88.80%, respectively. The highest accuracy achieved by AlexNet is 88.80% with SVM. Similarly, ResNet-50 achieves the maximum accuracy of 89.20% with SVM. ShuffleNet achieves the highest accuracy of 86.20% with SVM and ensemble. SqueezeNet achieves the highest accuracy of the 82.80% with the ensemble. MobileNet-v2, Inceptionv3, GoogleNet, EfficientNetb0 and Inception-ResNet-v2 has 91.80%, 90.90%, 87.90%, 92.20% and 91.80% accuracy with the SVM classifier, respectively. However, DarkNet-53 achieves the highest accuracy of 91.40% with the ensemble. The proposed framework with ViT classifier achieves an accuracy of 98.28%, which is a 4.78% increase

in accuracy compared to Xception with SVM. The comparisons demonstrate that the proposed model can significantly outperform existing D-CNN models by achieving the highest classification accuracy.

We also visualized the hot maps of the Grad CAM to visualize the portion of the image, which helps in classification of the images with 10% to 90% background. The visual results are shown in Figure 7.

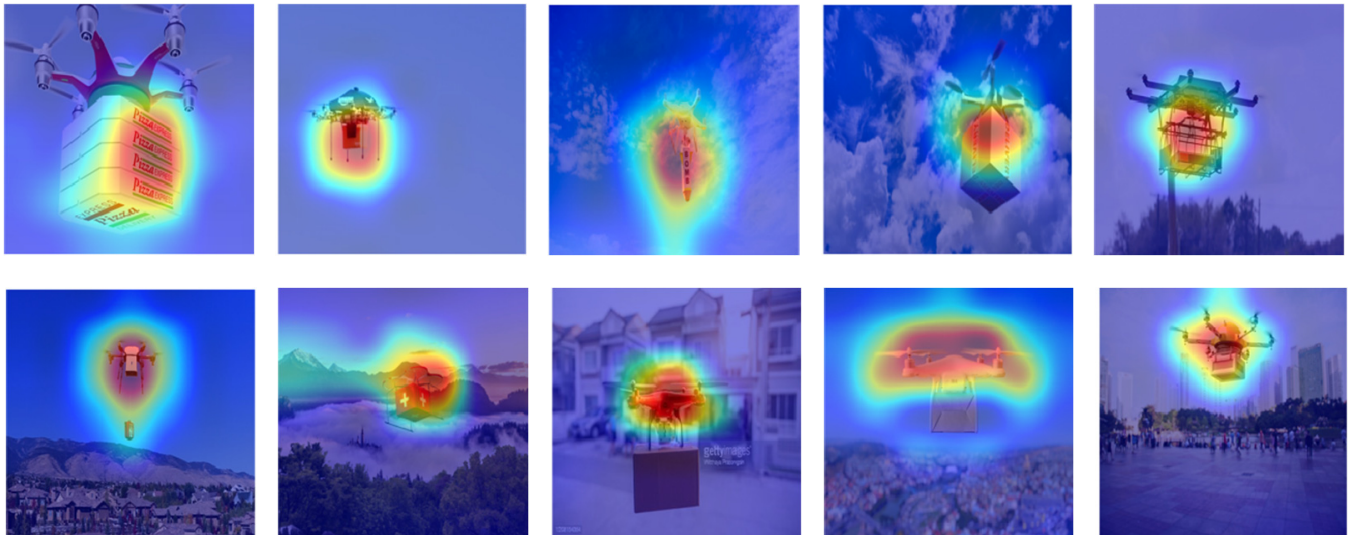


Figure 7. Hot map visualization of malicious drone images with 10% to 90% background using Grad CAM.

From Figure 7, it can be observed that when the load is near the drone, even in the 90% background images, it contributes to the classification. However, when the load is tied with string or relatively far away from the drone body, then only the drone contributes to the classification. From the performance comparison, it is evident that the proposed framework can be employed as a robust and efficient classification model for malicious drone detection. The current framework can be extended for the image compression [41], classification, and other computer vision tasks, such as object detection [42–44], and motor imagery classification in the brain–computer interface (BCI) [45–47]. The work can further be extended to classify malicious drones using selected features with nature and bio inspired algorithms [48–50], such as particle swarm optimization (PSO), genetic algorithm (GA), artificial bee colony (ABC), etc.

4. Conclusions

Drones are widely used due to their numerous applications. However, malicious drones which carry harmful material can cause destruction and bomb blasts. Thus, it is critical to distinguish between malicious drones and other flying objects. In this article, several ML and DL techniques are analyzed, which reveal that the performance of the handcrafted descriptors with ML classifiers is relatively low. Furthermore, the performance of various D-CNN ML classifiers is also evaluated. Our study indicates that the highest accuracy achieved by D-CNN models is 93.50%. However, the overall classification accuracy of the ViT classifier is 98.3%, which is the highest among all models. The ViT classifier achieves the overall recall, precision, and F_1 – score of 99.0%, 99.0%, 99.0%, and 99.0%, respectively. The precision, recall, F_1 – score, and Cohen’s kappa for malicious drone class are 97.0%, 97.0%, 97.0%, and 97.0%, respectively. The current study illustrates that the proposed ViT-based approach can help to classify malicious drones more efficiently than state-of-the-art D-CNN models. Training with a large dataset can further enhance the performance of the ViT-based framework. Nevertheless, the current framework can also

be extended to various classification and computer vision tasks, such as object detection, motor imagery classification in the brain–computer interface, etc.

Author Contributions: Conceptualization, S.J. and M.S.A.; methodology, S.J.; software, S.J.; validation, S.J., M.S.A. and A.M.R.; formal analysis, S.J.; writing—original draft preparation, S.J. and M.S.A.; writing—review and editing, A.M.R.; visualization, S.J.; supervision, A.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset link: <https://www.kaggle.com/sonainjamil/malicious-drones> (accessed on 28 February 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ayamga, M.; Tekinerdogan, B.; Kassahun, A. Exploring the Challenges Posed by Regulations for the Use of Drones in Agriculture in the African Context. *Land* **2021**, *10*, 164. [CrossRef]
2. Cancela, J.J.; González, X.P.; Vilanova, M.; Mirás-Avalos, J.M. Water Management Using Drones and Satellites in Agriculture. *Water* **2019**, *11*, 874. [CrossRef]
3. Hwang, J.; Kim, I.; Gulzar, M.A. Understanding the Eco-Friendly Role of Drone Food Delivery Services: Deepening the Theory of Planned Behavior. *Sustainability* **2020**, *12*, 1440. [CrossRef]
4. Dal Sasso, S.F.; Pizarro, A.; Manfreda, S. Recent Advancements and Perspectives in UAS-Based Image Velocimetry. *Drones* **2021**, *5*, 81. [CrossRef]
5. Amponis, G.; Lagkas, T.; Zevgara, M.; Katsikas, G.; Xirofotos, T.; Moscholios, I.; Sarigiannidis, P. Drones in B5G/6G Networks as Flying Base Stations. *Drones* **2022**, *6*, 39. [CrossRef]
6. Verdiesen, I.; Aler Tubella, A.; Dignum, V. Integrating Comprehensive Human Oversight in Drone Deployment: A Conceptual Framework Applied to the Case of Military Surveillance Drones. *Information* **2021**, *12*, 385. [CrossRef]
7. Jamil, S.; Fawad; Rahman, M.; Ullah, A.; Badnava, S.; Forsat, M.; Mirjavadi, S.S. Malicious UAV Detection Using Integrated Audio and Visual Features for Public Safety Applications. *Sensors* **2020**, *20*, 3923. [CrossRef] [PubMed]
8. Anwar, M.Z.; Kaleem, Z.; Jamalipour, A. Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 2526–2534. [CrossRef]
9. Liu, H.; Wei, Z.; Chen, Y.; Pan, J.; Lin, L.; Ren, Y. Drone detection based on an audio-assisted camera array. In Proceedings of the 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), Laguna Hills, CA, USA, 19–21 April 2017; pp. 402–406.
10. Dumitrescu, C.; Minea, M.; Costea, I.M.; Cosmin Chiva, I.; Semenescu, A. Development of an Acoustic System for UAV Detection. *Sensors* **2020**, *20*, 4870. [CrossRef]
11. Digulescu, A.; Despina-Stoian, C.; Stănescu, D.; Popescu, F.; Enache, F.; Ioana, C.; Rădoi, E.; Rîncu, I.; Șerbănescu, A. New Approach of UAV Movement Detection and Characterization Using Advanced Signal Processing Methods Based on UWB Sensing. *Sensors* **2020**, *20*, 5904. [CrossRef] [PubMed]
12. Singha, S.; Aydin, B. Automated Drone Detection Using YOLOv4. *Drones* **2021**, *5*, 95. [CrossRef]
13. Al-Emadi, S.; Al-Ali, A.; Al-Ali, A. Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks. *Sensors* **2021**, *21*, 4953. [CrossRef] [PubMed]
14. Wojtanowski, J.; Zygmunt, M.; Drozd, T.; Jakubaszek, M.; Życzkowski, M.; Muzal, M. Distinguishing Drones from Birds in a UAV Searching Laser Scanner Based on Echo Depolarization Measurement. *Sensors* **2021**, *21*, 5597. [CrossRef]
15. Coluccia, A.; Fascista, A.; Schumann, A.; Sommer, L.; Dimou, A.; Zarpalas, D.; Méndez, M.; de la Iglesia, D.; González, I.; Mercier, J.-P.; et al. Drone vs. Bird Detection: Deep Learning Algorithms and Results from a Grand Challenge. *Sensors* **2021**, *21*, 2824. [CrossRef] [PubMed]
16. Swinney, C.J.; Woods, J.C. The Effect of Real-World Interference on CNN Feature Extraction and Machine Learning Classification of Unmanned Aerial Systems. *Aerospace* **2021**, *8*, 179. [CrossRef]
17. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Houlsby, N. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
18. Patel, C.I.; Labana, D.; Pandya, S.; Modi, K.; Ghayvat, H.; Awais, M. Histogram of Oriented Gradient-Based Fusion of Features for Human Action Recognition in Action Video Sequences. *Sensors* **2020**, *20*, 7299. [CrossRef] [PubMed]
19. Song, T.; Li, H.; Meng, F.; Wu, Q.; Cai, J. LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 1565–1579. [CrossRef]

20. Yasmin, S.; Pathan, R.K.; Biswas, M.; Khandaker, M.U.; Faruque, M.R.I. Development of a Robust Multi-Scale Featured Local Binary Pattern for Improved Facial Expression Recognition. *Sensors* **2020**, *20*, 5391. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Fanizzi, A.; Basile, T.M.; Losurdo, L.; Bellotti, R.; Bottigli, U.; Campobasso, F.; Didonna, V.; Fausto, A.; Massafra, R.; Tagliafico, A.; et al. Ensemble Discrete Wavelet Transform and Gray-Level Co-Occurrence Matrix for Microcalcification Cluster Classification in Digital Mammography. *Appl. Sci.* **2019**, *9*, 5388. [\[CrossRef\]](#)
22. Nguyen, D.T.; Zong, Z.; Ogunbona, P.; Li, W. Object detection using non-redundant local binary patterns. In Proceedings of the 17th IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 4609–4612.
23. Wu, X.; Sun, J. Joint-scale LBP: A new feature descriptor for texture classification. *Vis. Comput.* **2017**, *33*, 317–329. [\[CrossRef\]](#)
24. Murala, S.; Maheshwari, R.P.; Balasubramanian, R. Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval. *IEEE Trans. Image Process.* **2012**, *21*, 2874–2886. [\[CrossRef\]](#)
25. Minhas, R.A.; Javed, A.; Irtaza, A.; Mahmood, M.T.; Joo, Y.B. Shot Classification of Field Sports Videos Using AlexNet Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 483. [\[CrossRef\]](#)
26. Liu, G.; Zhang, C.; Xu, Q.; Cheng, R.; Song, Y.; Yuan, X.; Sun, J. I3D-Shufflenet Based Human Action Recognition. *Algorithms* **2020**, *13*, 301. [\[CrossRef\]](#)
27. Fulton, L.V.; Dolezel, D.; Harrop, J.; Yan, Y.; Fulton, C.P. Classification of Alzheimer’s Disease with and without Imagery Using Gradient Boosted Machines and ResNet-50. *Brain Sci.* **2019**, *9*, 212. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Wang, A.; Wang, M.; Jiang, K.; Cao, M.; Iwahori, Y. A Dual Neural Architecture Combined SqueezeNet with OctConv for LiDAR Data Classification. *Sensors* **2019**, *19*, 4927. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Li, W.; Liu, K. Confidence-Aware Object Detection Based on MobileNetv2 for Autonomous Driving. *Sensors* **2021**, *21*, 2380. [\[CrossRef\]](#)
30. Sun, X.; Li, Z.; Zhu, T.; Ni, C. Four-Dimension Deep Learning Method for Flower Quality Grading with Depth Information. *Electronics* **2021**, *10*, 2353. [\[CrossRef\]](#)
31. Lee, Y.; Nam, S. Performance Comparisons of AlexNet and GoogLeNet in Cell Growth Inhibition IC50 Prediction. *Int. J. Mol. Sci.* **2021**, *22*, 7721. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Jamil, S.; Rahman, M.; Haider, A. Bag of Features (BoF) Based Deep Learning Framework for Bleached Corals Detection. *Big Data Cogn. Comput.* **2021**, *5*, 53. [\[CrossRef\]](#)
33. Ananda, A.; Ngan, K.H.; Karabağ, C.; Ter-Sarkisov, A.; Alonso, E.; Reyes-Aldasoro, C.C. Classification and Visualisation of Normal and Abnormal Radiographs; A Comparison between Eleven Convolutional Neural Network Architectures. *Sensors* **2021**, *21*, 5381. [\[CrossRef\]](#)
34. Demertzis, K.; Tsiknas, K.; Takezis, D.; Skianis, C.; Iliadis, L. Darknet Traffic Big-Data Analysis and Network Management for Real-Time Automating of the Malicious Intent Detection Process by a Weight Agnostic Neural Networks Framework. *Electronics* **2021**, *10*, 781. [\[CrossRef\]](#)
35. Chao, X.; Hu, X.; Feng, J.; Zhang, Z.; Wang, M.; He, D. Construction of Apple Leaf Diseases Identification Networks Based on Xception Fused by SE Module. *Appl. Sci.* **2021**, *11*, 4614. [\[CrossRef\]](#)
36. Guo, Y.; Fu, Y.; Hao, F.; Zhang, X.; Wu, W.; Jin, X.; Bryant, C.R.; Senthilnath, J. Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol. Indic.* **2021**, *120*, 106935. [\[CrossRef\]](#)
37. Joachims, T. 11 Making Large-Scale Support Vector Machine Learning Practical. In *Advances in Kernel Methods: Support Vector Learning*; The MIT Press: Cambridge, MA, USA, 1999; p. 169.
38. Roshani, M.; Phan, G.T.T.; Ali, P.J.M.; Roshani, G.H.; Hanus, R.; Duong, T.; Corniani, E.; Nazemi, E.; Kalmoun, E.M. Evaluation of flow pattern recognition and void fraction measurement in two phase flow independent of oil pipeline’s scale layer thickness. *Alex. Eng. J.* **2021**, *6*, 1955–1966. [\[CrossRef\]](#)
39. Sattari, M.A.; Roshani, G.H.; Hanus, R.; Nazemi, E. Applicability of time-domain feature extraction methods and artificial intelligence in two-phase flow meters based on gamma-ray absorption technique. *Measurement* **2021**, *168*, 108474. [\[CrossRef\]](#)
40. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; p. 30.
41. Jamil, S.; Piran, M.J.; Rahman, M. Learning-Driven Lossy Image Compression; A Comprehensive Survey. *arXiv* **2022**, arXiv:2201.09240.
42. Roy, A.M.; Bhaduri, J. A Deep Learning Enabled Multi-Class Plant Disease Detection Model Based on Computer Vision. *AI* **2021**, *2*, 413–428. [\[CrossRef\]](#)
43. Roy, A.M.; Bhaduri, J. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. *Comput. Electron. Agric.* **2022**, *193*, 106694. [\[CrossRef\]](#)
44. Roy, A.M.; Bose, R.; Bhaduri, J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. *Neural Comput. Appl.* **2022**, *34*, 3895–3921.
45. Roy, A.M. An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces. *Biomed. Signal Process. Control* **2022**, *74*, 103496. [\[CrossRef\]](#)
46. Roy, A.M. A multi-scale fusion CNN model based on adaptive transfer learning for multi-class MI-classification in BCI system. *BioRxiv* **2022**. [\[CrossRef\]](#)
47. Jamil, S.; Rahman, M. A Novel Deep-Learning-Based Framework for the Classification of Cardiac Arrhythmia. *J. Imaging* **2022**, *8*, 70. [\[CrossRef\]](#)

-
48. Jamil, S.; Rahman, M.; Tanveer, J.; Haider, A. Energy Efficiency and Throughput Maximization Using Millimeter Waves–Microwaves HetNets. *Electronics* **2022**, *11*, 474. [[CrossRef](#)]
 49. Too, J.; Abdullah, A.R.; Mohd Saad, N.; Tee, W. EMG Feature Selection and Classification Using a Pbest-Guide Binary Particle Swarm Optimization. *Computation* **2019**, *7*, 12. [[CrossRef](#)]
 50. Jamil, S.; Rahman, M.; Abbas, M.S.; Fawad. Resource Allocation Using Reconfigurable Intelligent Surface (RIS)-Assisted Wireless Networks in Industry 5.0 Scenario. *Telecom* **2022**, *3*, 163–173. [[CrossRef](#)]