Cilt 12, Sayı 1, Sayfa 37-43, 2023 Volume 12, Issue 1, Page 37-43, 2023

Araştırma Makalesi https://doi.org/10.46810/tdfd.1225756 Research Article

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Classification of Scenes in Aerial Images with Deep Learning Models

Özkan İNİK[1*]

[1] Department of Computer Engineering, Tokat Gaziosmanpasa University, Tokat, Türkiye
ORCID No: 0000-0003-4728-8438

**Corresponding author: ozkan.inik@gop.edu.tr*

(Received: 28.12.2022, Accepted: 08.02.2023, Online Publication: 27.03.2023)

Keywords
Aerial images classification, Deep learning, CNN pruning, VGG19

Abstract: Automatic classification of aerial images has become one of the topics studied in recent years. Especially for the use of drones in different fields such as agricultural applications, smart city applications, surveillance and security applications, it is necessary to automatically classify the images obtained with the camera during autonomous mission execution. For this purpose, researchers have created new data sets and some computer vision methods have been developed to achieve high accuracy. However, in addition to increasing the accuracy of the developed methods, the computational complexity should also be reduced. Because the methods to be used in devices such as drones where energy consumption is important should have low computational complexity. In this study, firstly, five different state-of-the-art deep learning models were used to obtain high accuracy values in the classification of aerial images. Among these models, the VGG19 model achieved the highest accuracy with 94.21%. In the second part of the study, the parameters of this model were analyzed and the model was reconstructed. The number of 143.6 million parameters of the VGG19 model was reduced to 34 million. The accuracy of the model obtained by reducing the number of parameters is 93.56% on the same test data. Thus, despite the 66.5% decrease in the parameter ratio, there was only a 0.7% decrease in the accuracy value. When compared to previous studies, the results show improved performance.

Havasal Görüntülerdeki Sahnelerin Derin Öğrenme Modelleri ile Sınıflandırılması

Anahtar Kelimeler
Havasal görüntü sınıflandırma, Derin Öğrenme, ESA budama, VGG19

Öz: Havadan alınan görüntülerin otomatik olarak sınıflandırılması son yıllarda üzerinde yoğun çalışılan konulardan biri haline gelmiştir. Özellikle dronların tarımsal uygulamalar, akıllı şehir uygulamaları, gözetleme ve güvenlik uygulamaları gibi farklı alanlarda kullanımında otonom görev icrası sırasında kamera ile elde edilen görüntülerin otomatik olarak sınıflandırılması gerekmektedir. Bu amaçla araştırmacılar yeni veri setleri oluşturmuş ve yüksek doğruluk elde etmek için bazı bilgisayarlı görü yöntemleri geliştirilmiştir. Ancak geliştirilen yöntemlerin doğruluğunun artırılmasının yanı sıra hesaplama karmaşıklığının da azaltılması gerekmektedir. Çünkü dron gibi enerji tüketiminin önemli olduğu cihazlarda kullanılacak yöntemlerin düşük hesaplama karmaşıklığına sahip olması gerekmektedir. Bu çalışmada, öncelikle hava görüntülerinin sınıflandırılmasında yüksek doğruluk değerleri elde etmek için beş farklı derin öğrenme modeli kullanılmıştır. Bu modeller arasında en yüksek doğruluğu %94.21 ile VGG19 modeli elde etmiştir. Çalışmanın ikinci bölümünde bu modelin parametreleri analiz edilerek model yeniden yapılandırılmıştır. VGG19 modelinin 143.6 milyon olan parametre sayısı 34 milyona düşürülmüştür. Parametre sayısının azaltılmasıyla elde edilen modelin doğruluğu aynı test verileri üzerinde %93.56'dır. Böylece parametre oranındaki %66.5'lik azalmaya karşın doğruluk değerinde sadece %0.7'lik bir azalma olmuştur. Elde edilen sonuçlar önceki çalışmalarla karşılaştırıldığında, daha iyi sonuçların elde edildiği görülmüştür.

1. INTRODUCTION

The development of aerial vehicles and remote sensing methods has led to the studies for the autonomous implementation of many new applications. In particular, studies have been carried out on aerial control security systems, agricultural applications, control of smart cities, and the identification of scenes in aerial images in the movie industry. For example, an autonomous tracking

drone is used to shoot any scene on a movie set. In another example, aerial imagery is used to detect an environmental disaster from satellite photos. For this reason, researchers in the literature have initially created benchmark data sets for artificial intelligence-based vision systems to be developed in this field [1-5]. Deep learning-based methods were developed on these datasets [6-12]. Deep learning is a sub-branch of artificial intelligence and first attracted attention with the ImageNet competition in 2012. Deep learning, which was recognized for its high accuracy in this competition, has been used in many studies such as classification [13, 14], detection [15, 16], segmentation [17-20] and prediction [21]. The details of some of the deep learning-based studies used for the classification of scenes in aerial images are given below.

In [6], a deep learning method for aerial image classification using Residual Network v2 (IRV2) with Inception and a multilayer perceptron (MLP) model is proposed. The method includes preprocessing, feature extraction and classification of the images acquired by the UAV. The proposed method uses IRV2 for feature extraction and MLP for classification. The effectiveness of this method, called DLIRV2-MLP, was tested on a comparative aerial imagery dataset and was found to have an accuracy and precision of 90% and above. In [7] a semi-supervised center-based discriminative adversarial learning framework called SCDAL is proposed to address the labeling problem, which is the biggest time consuming issue in supervised classification of aerial images. The SCDAL framework is tested on two large aerial image datasets and shown to be superior to most existing domain adaptation methods with at least a 3% improvement in overall accuracy. Work [8] proposes a prototype-based memory network for recognizing multiple scenes in a single aerial image. The network consists of a prototype learning module, a prototype-hosting external memory, and a multi-head attention-based memory retrieval module. In the study, a puplic dataset is also created and the effectiveness of the proposed method on other datasets is demonstrated in experimental results. In [9] a deep learning system is proposed to classify objects and facilities into 63 different classes from high-resolution, multi-spectral satellite images. The proposed method achieved an overall accuracy of 83%, an F1 score of 0.797, and classified 15 of the classes with 95% or higher accuracy. Study [10] proposes a transfer learning based technique for aerial scene classification using a layer selection strategy called ReLu Based Feature Fusion (RBFF). In RBFF, a pre-trained MobileNetV2 deep learning model is used for feature discovery. The extracted features are dimensionally reduced with dimension reduction algorithms and classification is performed with support vector machine. With the result obtained, it is stated that it outperforms the studies conducted in recent years. In [11] proposed a method that all grains, one scheme (AGOS). The method is a method for extracting discriminative information from multi-granular representations of data proposed. It consists of three components: MGP, which preserves extended convolutional features from the backbone used for feature extraction; MBMIR, which highlights key examples in the multi-granular representation using multiple examples learning; and SSF, which allows the method to learn the same scene diagram from multi-granular instance representations and combine them for optimization. The AGOS method has been tested on three public datasets and has been reported to perform well. In [12], it is stated that objects in birds eye view images are more complex than objects in natural view, and therefore the discriminative features of scenes are difficult. To overcome this problem, a solution was sought by designing a new representation set called instance representation bank (IRB). The performance of the proposed method on three different datasets is very competitive compared to its competitors. Although new methods have been developed for classifying scenes in aerial images, these methods need to have low computational complexity. Because they need to be used on hardware with less energy-consuming processors. Especially deep learning-based methods need high capacity computing resources due to the high number of parameters. In order to overcome this situation, some studies [22-25] should be carried out to reduce the number of parameters of deep learning architectures. In this study firstly, the state of the art deep learning models Googlenet [26], Alexnet [27], Vgg19 [28], resnet-50, and Resnet-101 [29] were used to achieve high accuracy in the classification of scenes in aerial images. Applications were made on the popular Aerial Image data set (AID) [5]. Secondly, the effect of the feature maps in the VGG19 model, which gives the highest accuracy among these models, was investigated and the features with little effect were deleted. In this way, the VGG19 model was able to achieve results close to the same success with fewer parameters.

In the remaining sections of this paper is Section 2, the architecture and parameters of the VGG19 model are described. Also, the AID dataset is described in this section. Finally, the proposed method is presented. Section 3 describes the experimental studies, the results obtained by testing the models and comparing them with each other. Section 4 presents the conclusions of the study

2. MATERIAL AND METHOD

2.1. VGG19

The VGG19 model was developed by Karen Simonyan and Andrew Zisserman for the ImageNet competition in 2014. The goal of this model is to use convolution filters of smaller size for deeper model design. The architecture of the model and its values at each layer are given in Figure 1. When we look at the number of parameters of the model, it is seen that the most parameters are in the last convolution layer. The number of parameters here constitutes approximately 86% of the total number of parameters of the model. The aim of this study is to find the best features in the feature maps in this layer and achieve high accuracy with fewer features

Figure 1. The layer architecture of the VGG19 model and parameter values in each layer [30]

2.2. Data Set

The dataset was created by Xia et al. from google earth images [5]. The dataset consists of 30 classes in total. It consists of a total of 1000 images, with an average of 333 images for each class. A sample image of the dataset is given in Figure 2. The number of images in each class and class names are given in Table 1.

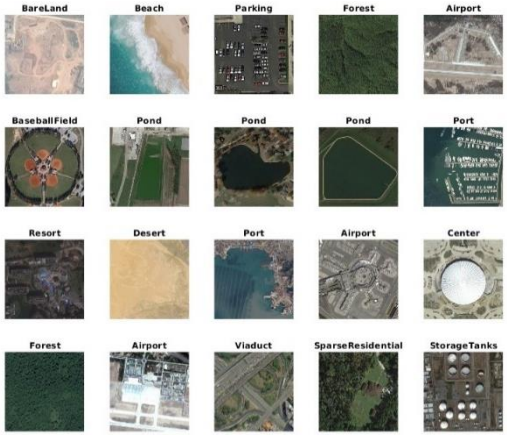

Figure 2. An example representation of the images in the AID dataset

2.3. Proposed Method

In this study, the VGG19 model was used along with other deep learning models to classify scenes in aerial images. Although the VGG19 model achieves high accuracy values, it has high parameter values. This increases the computational complexity of this model during the training and testing phase. To overcome this situation, the number of features in the flattening layer, which is the bottleneck of the model, has been reduced. The flow diagram of the proposed method is given in Figure 3. First, the dataset is divided into two parts, 50% training and 50% testing. The VGG19 model was trained with 50% of the dataset. After the training process, the feature map in the trained model was analyzed. In the analysis, the features obtained by the last convolution layer of the model were tested one by one on the test data and their effect on the result was calculated. The features with low impact were deleted from the VGG19 model. The VGG19 model was recoded according to the deleted features and retrained with the training data. After the training process, it was tested with test data and its final success was measured.

Table 1. Classes in the AID dataset and the number of images in each class

Class name	Number of images	Class name	Number of images
Airport	360	Mountain	340
BareLand	310	Park	350
BaseballField	220	Parking	390
Beach	400	Playground	370
Bridge	360	Pond	420
Center	260	Port	380
Church	240	RailwayStation	260
Commercial	350	Resort	290
DenseResidential	410	River	410
Desert	300	School	300
Farmland	370	SparseResidential	300
Forest	250	Square	330
Industrial	390	Stadium	290
Meadow	280	StorageTanks	360
MediumResidential	290	Viaduct	420

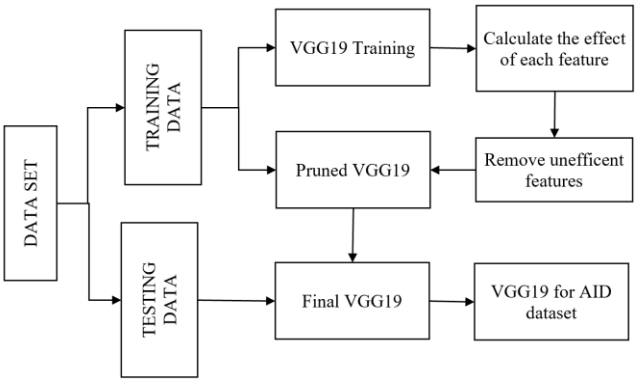

Figure 3. Flow diagram of the proposed method for reducing the number of parameters of the VGG19 model.

3. EXPARIMENTAL STUDIES

In the experimental studies, training and testing of the Googlenet, VGG19, Alexnet, Resnet-50, and Resnet-101 models were carried out. In addition, the VGG19 model, which has the highest number of parameters, was pruned and the training and testing of the adapted VGG19 model was carried out. Deep Learning toolbox of Matlab 2022b software was used in all experimental studies. The applications were performed on Nvidia 2080TI GeForce graphics cards.

3.1. Performance Criteria

In this study, the confusion matrix shown in Figure 4 was used to measure the performance of the models. The accuracy value obtained using the confusion matrix and given in Equation-1 was used to compare the models.

		PREDICT CLASS	
		Positives	Negatives
TRUE CLASS	Positives	True Positive **(TP)**	False Positive **(FP)**
	Negatives	False Negative **(FN)**	True Negative **(TN)**

Figure 4. Confusion matrix [18]

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

3.2. Training and Testing Results

The training parameters of the models are given in Table 2. All models are trained according to the same parameters.

Table 2. Parameters used in training the models

Parameter Name	Parameter Value
Optimization algorithm	SGDM
Initial learning rate	0.001
Epoch	50
Batch size	64
Learning rate drop factor	0.8
Learning rate drop period	10

The performances of the models were performed on the AID dataset, 50% of which was used for training and the rest for testing. In the training phase, the test data was used as validation data. In this way, it was examined whether the model overfitting in training. Some statistical information about the dataset is given in Table3.

Table 3. Some statistical information calculated on the AID dataset

Information	Value
Class number	30
Total number of images	10000
Average number of images per class	333
Number of images of the class with the fewest images	220
Number of images of the class with the most images	420
Standard deviation of the number of images in classes	57.79
Number of images reserved for training	5000
Number of images reserved for testing	5000

The convergence graphs obtained by the VGG19 model when trained with the training dataset are given in Figure 5. It reached 100% accuracy in the training phase with a total of 30 classes and 5000 images. In the validation data, the accuracy values after the 12th epoch were 93-94%.

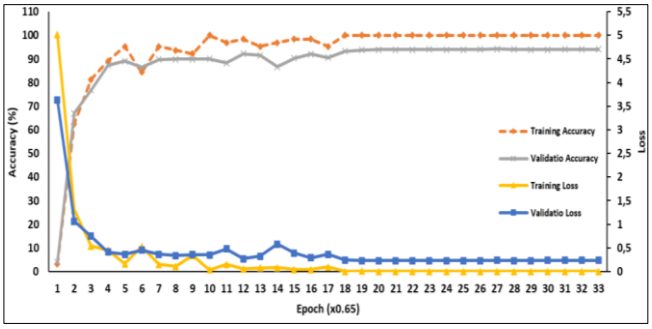

Figure 5. Convergence graphs obtained by the original VGG19 model in the training phase

The convergence graphs obtained by the pruned VGG19 model on the training data are given in Figure 6. Looking at the figure, it is expected that the accuracy and loss values are close to the previous result since the model was previously trained with the same data set. The most important reason for retraining the model is to rediscover the features in the last convolution layer. Because there were 512 filters in this layer, 449 of which were deleted. The features discovered by the remaining 63 filters were used to determine the model success.

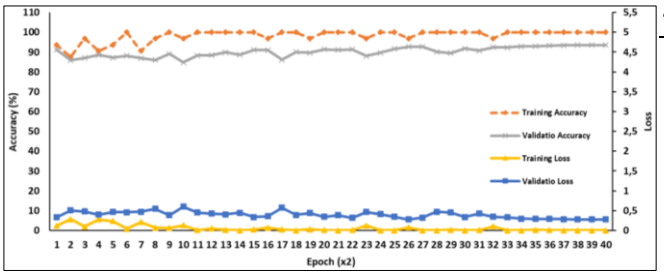

Figure 6. Convergence plots obtained by the pruned VGG19 model in the training phase

The confusion matrix obtained by the original VGG19 model in the test data is given in Figure 7. The confusion matrix of the pruned VGG19 model obtained by modifying the model with the proposed method is given in Figure 8. The model achieved an accuracy of 90.10% at this stage. After the modification, there were 63 features in the last feature layer of the Pruned VGG19 model. When the model was retrained and tested with the training dataset to retrieve these features, the confusion matrix in Figure 9 was obtained. When Figure 7-9 is analyzed, it is seen that the class most affected by the deleted 449 features is "BaseballField". With the retraining of the model, it was seen that this class was classified with the desired accuracy.

In addition to the VGG19 model, Alexnet, Googlenet, Resnet-50, Resnet-101 models, which are state-of-the-art models in the field of deep learning, were also tested on the AID dataset. Confusion matrices obtained for these

models are given in Figure 10-13 respectively. The comparison of all models is given in Table 4.

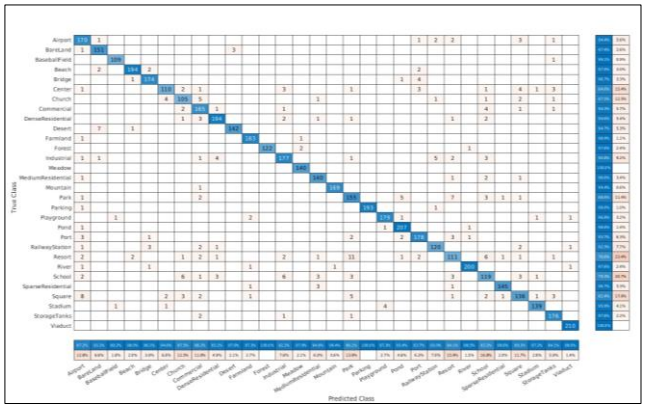

Figure 7. Confusion matrix obtained by the original VGG19 on test data

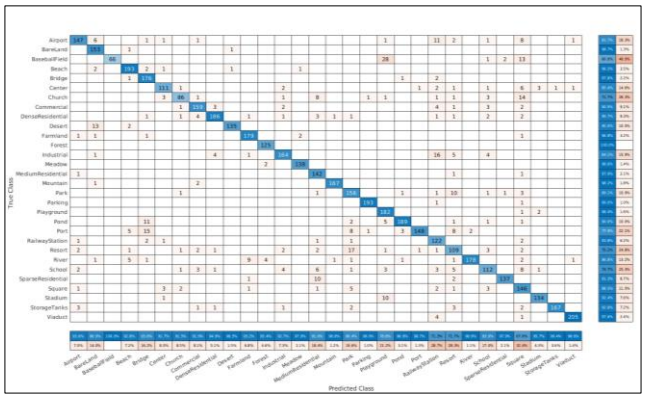

Figure 8. Confusion matrix obtained by the pruned VGG19 on test data

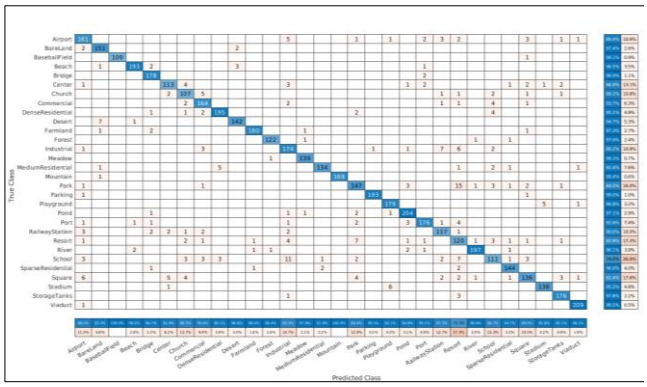

Figure 9. Confusion matrix obtained by the pruned VGG19 model on test data after retraining with training data

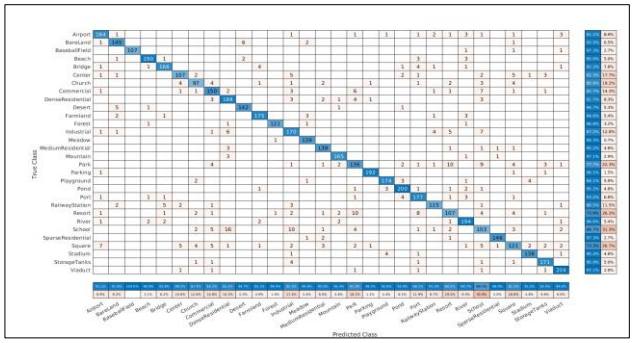

Figure 10. Confusion matrix obtained by Alexnet on test data

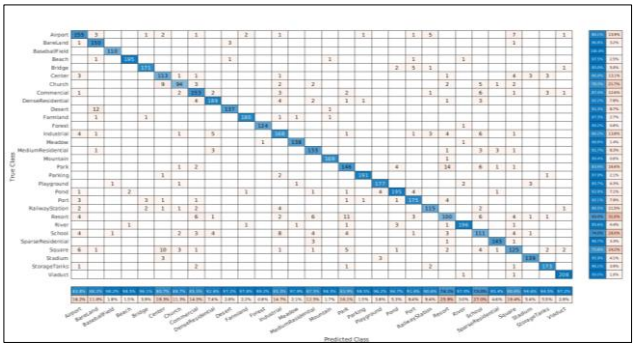

Figure 11. Confusion matrix obtained by Googlenet on the test data

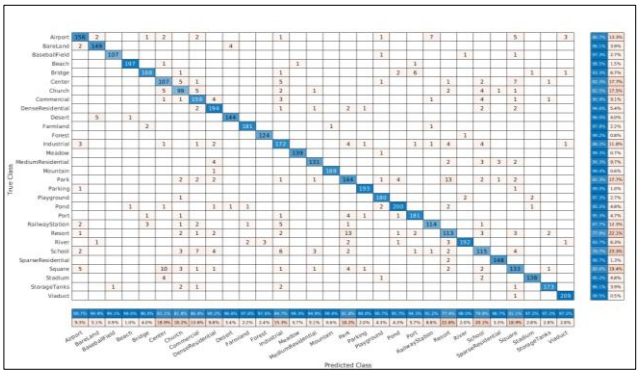

Figure 12. Confusion matrix obtained by Resnet-50 on the test data

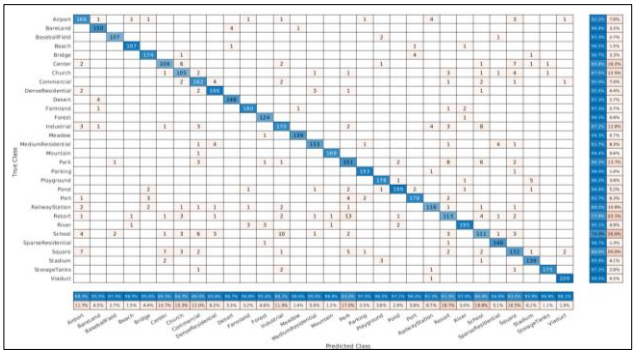

Figure 13. Confusion matrix obtained by ResNet-101 on the test data.

Table 4. Number of parameters, layer depth, and accuracy of the deep learning models on the AID dataset.

Model Name	Accuracy (%)	Number of Parameters (Million)	Number of Layers
Resnet-50	92.58	25.5	177
Resnet-101	93.28	42.6	347
Googlenet	91.50	6.9	144
Alexnet	90.84	60.9	25
VGG19	94.26	143.6	47
Pruned VGG19	93.56	48.1	47

Looking at Table 4, it is seen that the best accuracy value among the models is obtained by the VGG19 model with a value of 94.26%. However, the parameter value of this model is higher than all other models. On the contrary, the Pruned VGG19 model has 95.5 million fewer parameters than its original version and gave the second best result.

3.3. Comparison with Other Methods

Many studies have been carried out in the literature on the AID dataset. The results obtained in these studies and the result obtained with the proposed VGG19 model are given in Table 5. In the table, while the baseline is 68.96% in the AID data set, an accuracy value of 93.56% was obtained with Pruned VGG19. It has been observed that the results obtained with the proposed method have significantly improved the accuracy values compared to other studies.

Table 5. Comparison of the results of previous studies in the AID dataset with the proposed study.

Authors/Reference	Method	Accuracy (%)
Xia, G.-S., et al.[5]	PLSA (SIFT)	63.07
Xia, G.-S., et al.[5]	BoVW (SIFT)	68.37
Xia, G.-S., et al.[5]	LDA (SIFT)	68.96
Han, X., et al [31]	SPP with Alexnet	91.45
Anwer, R.M., et al [32]	TEX-Net with VGG	90.00
Ilse, M., et al.[33]	Gated Attention	92.01
Bi, Q., et al.[34]	MIDC-Net CS	92.95
Bi, Q., et al.[35]	RADC-Net	92.35
Cao, R., et al.[36]	VGG VD16 + SAFF	93.83
Arefeen, M.A., et al.[10]	RBFF (3 6 13) + PCA (600) + LDA + SVM	93.73
Cheng, G., et al.[37]	D-CNN with AlexNet	94.47
Proposed	VGG19	94.26
Proposed	Pruned VGG19	93.56

4. CONCLUSION

In this study, deep learning models are used to classify scenes from aerial images. First, the state-of-the-art deep learning models AlexNet, VGG19, Googlenet, Resnet-50, and Resnet-101 were used on the AID dataset. The accuracy values obtained by each model are 90.84%, 94.26%, 91.50%, 92.58%, and 93.28% respectively. Secondly, the model was pruned by selecting the effective features among the features discovered in the VGG19 model and deleting the remaining features. The total number of parameters of the original model is 143.6 million and the accuracy is 94.26%, while the total number of parameters of the pruned version is 48.1 million and the accuracy is 93.56%. Thus, although the number of parameters of the VGG19 model is reduced by 66.5%, the loss in accuracy is only 0.7%. The pruned VGG19 model was found to achieve better results than the other four models. Future studies are planned to use multi-objective optimization algorithms to prune CNN models according to accuracy and number of parameters.

REFERENCES

[1] Zou, Q., et al., Deep learning based feature selection for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters, 2015. 12(11): p. 2321-2325.

[2] Xia, G.-S., et al. Structural high-resolution satellite image indexing. in ISPRS TC VII Symposium-100 Years ISPRS. 2010.

[3] Yang, Y. and S. Newsam. Bag-of-visual-words and spatial extensions for land-use classification. in Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. 2010.

[4] Cheng, G., J. Han, and X. Lu, Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE, 2017. 105(10): p. 1865-1883.

[5] Xia, G.-S., et al., AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 2017. 55(7): p. 3965-3981.

[6] Minu, M. and R.A. Canessane, Deep learning-based aerial image classification model using inception with residual network and multilayer perceptron. Microprocessors and Microsystems, 2022. 95: p. 104652.

[7] Zhu, R., et al., Semi-supervised center-based discriminative adversarial learning for cross-domain scene-level land-cover classification of aerial images. ISPRS Journal of Photogrammetry and Remote Sensing, 2019. 155: p. 72-89.

[8] Hua, Y., et al., Aerial scene understanding in the wild: Multi-scene recognition via prototype-based memory networks. ISPRS Journal of Photogrammetry and Remote Sensing, 2021. 177: p. 89-102.

[9] Pritt, M. and G. Chern. Satellite image classification with deep learning. in 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). 2017. IEEE.

[10] Arefeen, M.A., et al. A lightweight relu-based feature fusion for aerial scene classification. in 2021 IEEE International Conference on Image Processing (ICIP). 2021. IEEE.

[11] Bi, Q., et al., All Grains, One Scheme (AGOS): Learning Multi-grain Instance Representation for Aerial Scene Classification. arXiv preprint arXiv:2205.03371, 2022.

[12] Yi, J. and B. Zhou, Learning Instance Representation Banks for Aerial Scene Classification. arXiv preprint arXiv:2205.13744, 2022.

[13] İnik, Ö., CNN hyper-parameter optimization for environmental sound classification. Applied Acoustics, 2023. 202: p. 109168.

[14] 14. Falaschetti, L., et al., A CNN-based image detector for plant leaf diseases classification. HardwareX, 2022. 12: p. e00363.

[15] Girshick, R. Fast r-cnn. in Proceedings of the IEEE international conference on computer vision. 2015.

[16] İnik, Ö., et al., A new method for automatic counting of ovarian follicles on whole slide histological images based on convolutional neural network. Computers in biology and medicine, 2019. 112: p. 103350.

[17] İnik, Ö., et al., MODE-CNN: A fast converging multi-objective optimization algorithm for CNN-based models. Applied Soft Computing, 2021. 109: p. 107582.

[18] Inik, Ö. and E. Ülker, Optimization of deep learning based segmentation method. Soft Computing, 2022. 26(7): p. 3329-3344.
[19] Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. 2015. Springer.
[20] Genze, N., et al., Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. Computers and Electronics in Agriculture, 2022. 202: p. 107388.
[21] Orhan, İ., et al., Soil Temperature Prediction with Long Short Term Memory (LSTM). Türk Tarım ve Doğa Bilimleri Dergisi. 9(3): p. 779-785.
[22] Mondal, M., et al., Adaptive CNN filter pruning using global importance metric. Computer Vision and Image Understanding, 2022. 222: p. 103511.
[23] Pattanayak, S., S. Nag, and S. Mittal, CURATING: A multi-objective based pruning technique for CNNs. Journal of Systems Architecture, 2021. 116: p. 102031.
[24] Ide, H., et al., Robust pruning for efficient CNNs. Pattern Recognition Letters, 2020. 135: p. 90-98.
[25] Yang, C. and H. Liu, Channel pruning based on convolutional neural network sensitivity. Neurocomputing, 2022. 507: p. 97-106.
[26] Szegedy, C., et al. Going deeper with convolutions. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[27] Krizhevsky, A., I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks. Communications of the ACM, 2017. 60(6): p. 84-90.
[28] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[29] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[30] Matlab_2022b. Get Started with Deep Network Designer. 2022 [cited 2022 23.12.2022]; Available from: https://www.mathworks.com/help/deeplearning/gs/get-started-with-deep-network-designer.html.
[31] Han, X., et al., Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. Remote Sensing, 2017. 9(8): p. 848.
[32] Anwer, R.M., et al., Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS journal of photogrammetry and remote sensing, 2018. 138: p. 74-85.
[33] Ilse, M., J. Tomczak, and M. Welling. Attention-based deep multiple instance learning. in International conference on machine learning. 2018. PMLR.
[34] Bi, Q., et al., A multiple-instance densely-connected ConvNet for aerial scene classification. IEEE Transactions on Image Processing, 2020. 29: p. 4911-4926.
[35] Bi, Q., et al., RADC-Net: A residual attention based convolution network for aerial scene classification. Neurocomputing, 2020. 377: p. 345-359.
[36] Cao, R., et al., Self-attention-based deep feature fusion for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters, 2020. 18(1): p. 43-47.
[37] Cheng, G., et al., When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs. IEEE transactions on geoscience and remote sensing, 2018. 56(5): p. 2811-2821.