Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Remote Sensing Image Classification by Integrating Transfer Learning and Random Forests

Jiyan Suleiman Dahir
[1] *Pathology and Microbiology .Dept. Veterinary Medicine Collage, University of Duhok, Duhok, Iraq*
jeean.daher@uod.ac

ARTICLE INFO

Received: 22 Dec 2024

Revised: 12 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

Aerial ground images contain multi-level patterns and features, and the classification accuracy of the model for optical remote sensing scene images is limited due to the complex spatial patterns, similarity between classes, and high-class diversity. In this paper, an optical remote sensing scene classification algorithm based on random forests and transfer learning is proposed. Firstly, multi-level feature information is extracted from the MobileNetV2 model, Vgg19, and then the features are combined. The features extracted from both algorithms are filtered based on the importance of the feature to the random forest model chosen for classification, and finally the model is trained on the final features. The proposed method is verified by experiments on a public remote sensing dataset, UCM. Compared with other advanced scene classification methods, this method achieves better classification performance, achieving an accuracy of 97.14%.

Keywords: Remote sensing; Scene classification; Transfer learning; Random forests; Feature fusion.

INTRODUCTION

With the great development of remote sensing technologies, satellite and drone images have become widely used in various fields such as agriculture, urban planning, environmental management, and even in the face of natural disasters. These images provide knowledge of the Earth's surface, but they are difficult to analyze and classify.[1], [2], [3] The biggest problem researchers face is to classify scenes correctly, especially with the difficulty of spatial patterns and similarity between classes[4], [5]. For example, agricultural land may match residential or industrial land in images and it may be difficult to distinguish between them[6], [3]. The diversity of classes in remote sensing images makes the task even more difficult, which requires the development of new techniques to improve classification accuracy[7], [3], [8]. The past decade has seen the emergence of deep learning techniques, such as CNN, as one of the innovative methods to improve image classification. However, these methods tend to use large amounts of labeled data and high computational resources, which is not always possible. Convolutional neural networks consist of two stages: the first stage is feature extraction and the second stage is classification. [9], [10]Several pre-trained convolutional neural network models have emerged, such as MobileNet VGG ResNet, which have shown their effectiveness in image classification and feature extraction in Earth scene[11], [12], [10]. In this paper, a new algorithm for remote sensing scene classification is presented based on two techniques: random forests and transfer learning. Where pre-trained deep learning models like MobileNet and VGG19 are used to obtain multi-level features from images. The features are then combined and filtered according to their importance using random forests.

2. RELATED WORKS

Several researchers have presented methods and models for aerial scene classification where machine learning techniques were combined with deep learning to enhance classification accuracy. In a study on the effect of learning methods on remote sensing image classification [13], the author used boosting image classification with pre-trained models and deep neural networks such as ResNet and VGG to enhance classification accuracy over traditional methods. In [14], the authors used neural networks on the University of California Merced land dataset as a basis. The results showed that using GoogLeNet network with support vector machine (SVM) achieved a classification accuracy of 93%. This demonstrates the effectiveness of combining networks with

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

machine learning techniques to support classification processes. Also, in [15], a model called ConvCat was proposed, which combines the inner network layers with the CatBoost model for scene classification. The results showed that the new model performed better than existing models with a classification accuracy of 97.44 percent, on the UC Merced dataset and 96.89 percent, on the RESISC45 dataset. In [16], the authors used deep neural networks to process remote sensing images and classify land use. The experimental results showed that by combining different models such as CaffeNet and VGG-F, the accuracy is improved. A light convolutional neural network model was presented in [17], which can effectively classify remote sensing scenes without increasing computational complexity. The model achieved an accuracy of 88.29% on the University of California Merced dataset. In [18], the current developments in satellite image analysis using deep learning algorithms were highlighted. The paper discussed in detail how spectral analysis and machine learning can be used in the pursuit of more accurate classification. It was noted that the limited labeling of the data is a major obstacle to this pursuit. The paper [19] proposed deep learning plus classification fusion methods towards improving remote sensing scene classification. The developed model showed potential to accommodate variable data patterns and outperformed traditional methods in terms of classification. Researchers have proposed several methods but have not focused on the fact that this type of image has varying levels of detail and requires the extraction and classification of different features.

3 MATERIALS AND METHODS

The suggested method is explained in this section. A remote earth scene is classified using a multi-step process. To attain balance, the input photos are first processed and enhanced. Using feature fusion and transfer learning on the obtained deep feature maps, the second step uses the refined deep CNN models, MobileNet and VGG19, to extract deep features. Redundant features can be found in the concatenated feature space. The last step involves classifying aerial terrestrial photos by applying a random forest algorithm to enhance the fused feature vector's characteristics and eliminate redundant features. Our suggested methodology for classifying distant earth scenes is explained in full in Figure 1.

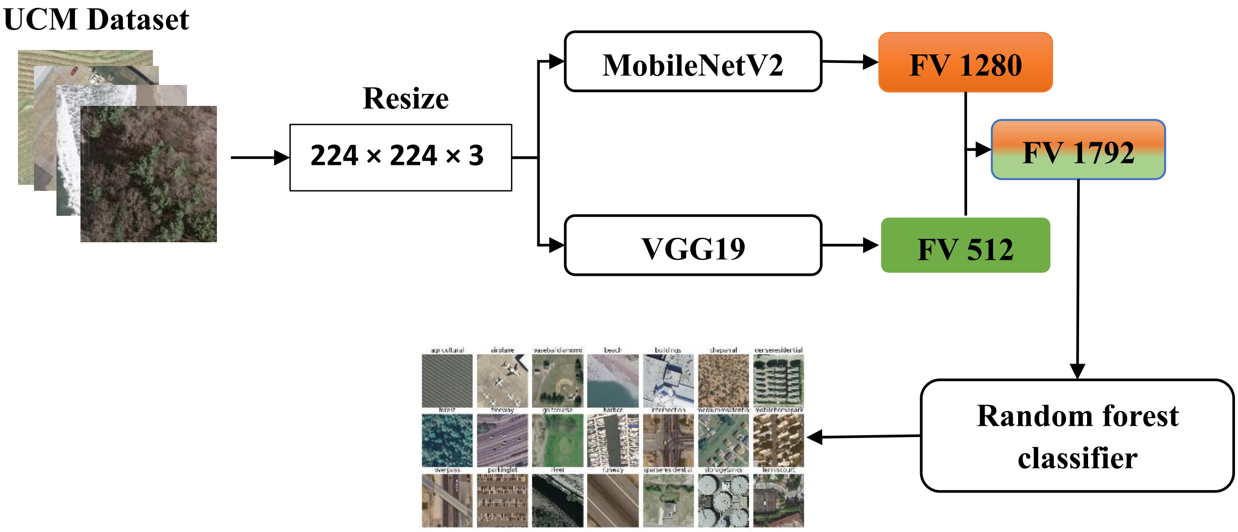

Fig 1. Proposed architecture of Feature fusion model.

This study made use of the UC Merced Land-Use dataset [20] for the purpose of examining and comparing various classification methods with the proposed model. This data is made up of high-resolution optical images that are taken in different locations in the USA and belong to the USGS (United States Geological Survey). This catalog of images includes several types of lands that display various patterns such as agricultural, as displayed in Fig 2.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

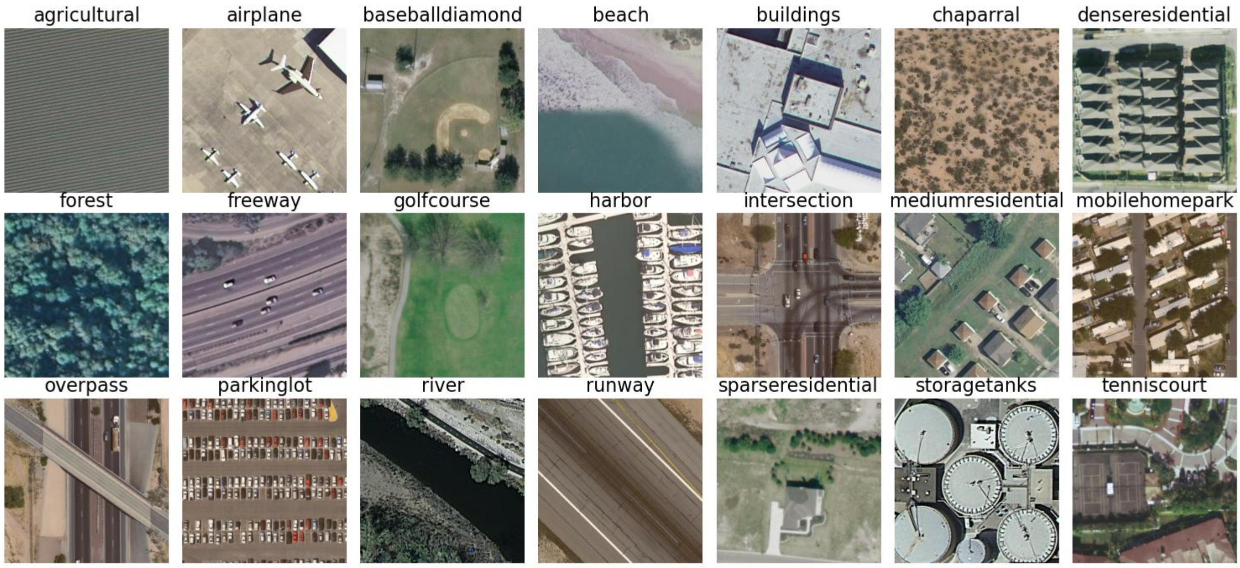

Fig. 2. 21 Classes representative of the UC-Merced Land-Use dataset

The UC Merced collection includes 21 different land-use categories, each with 100 pictures, with the resolution of 256 × 256 pixels. These categories represent diverse landscapes and urban settlements.

Remote sensing image analysis to discover patterns has a profound reliance on the application of deep learning features to also extract the deep ones within the images. These are the ones that correspond to the spatial and the geometric characteristics of objects studied, and thus they contribute to the improvement of the classification accuracy and efficiency. The suggested method is to employ both MobileNet and VGG19 to extract deep features from aerial photos, where these features are assembled and studied to spot the most crucial ones in the classification.

3.1. MobileNetV2 Model

The model uses a MobileNet architecture based on Deeply Separable Convolutions (DSCs), which in turn reduces the number of computations required. Images of size 224×224×3 are fed into the model, where they pass through multiple layers to extract the basic features, and then the dimensions are reduced using an Average Pooling layer to obtain a final feature map of size FV1×1024. Start The process By extension The feature dimensions are expanded by 1×1 convolution, where the number of channels is increased to represent the data more detailed. Batch normalization is applied to speed up training and stabilize the model, followed by the ReLU6 activation function [21][22][23].

$$Y = \sigma\big(BN(W_e * X)\big)$$

Where, Y denotes the output of the layer after the expansion is applied, while X represents the input matrix containing the features extracted from the previous layer. W_e are the weights of the expansion matrix that determine how the features are transformed. Batch normalization (BN) is applied to improve the stability of the model during training, and the activation function σ, which represents ReLU6, is used to apply nonlinear activation to the extracted features.

used This is amazing Stage Separate convolutions for each channel (Depthwise Convolutions), which reduce computation while maintaining feature quality.

$$Y_d = \sigma\big(BN(W_d * Y)\big)$$

Where, Y_d represents the output of deep convolution. W_d refers to the weights of the deep convolution matrix. After convolution, BN batch normalization is applied to improve the data distribution, while the activation function σ (ReLU6) is used to introduce a nonlinear response and improve the model performance.

After Extract Features With deep convolution, the dimensionality is reduced using 1×1 convolution (Pointwise Convolution). This process aims to reduce the number of channels while retaining the most important information. Batch normalization is applied after this step to ensure the stability of the training.

$$Y_p = BN\big(W_p * Y_d\big)$$

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Where, Y_p refers to the output of the projection layer, where the number of channels is reduced to the desired value. This is done using the projection weights W_p, which determine how to reduce the features extracted from the previous layers. The batch normalization process BN ensures improved model stability and prevents oscillations during training.

if He was number The channels in the input equal the number of channels in the output, a residual connection is applied to enable deep training without losing information.

$$Y_{\{out\}} = Y_p + X$$

Where, $Y_{\{out\}}$represents the final output after applying residual correlation, where the extracted features Y_p are combined with the original input X. This correlation allows the data to flow through the layers without losing information, which helps improve training efficiency and prevents the vanishing gradient problem.

The following equation represents the complete process of feature extraction in MobileNetV2:

$$Y\_\{out\} = BN(W\_p * (\sigma(BN(W\_d * (\sigma(BN(W\_e * X)))))))$$

In this equation, the following sequence is applied to extract features from the input image: First, the input X is dilated using weights W_e, then it is passed to the deep convolution stage using weights W_d. Then, the projection operation is applied using W_p, and finally the batch normalization BN and ReLU6 activation function are applied to obtain the final features $Y_{\{out\}}$. If there is a dimensional match, the original input X is added to preserve the information through the residual connection.

3.2. VGG19 Model

The VGG19 model has 3×3 convolutions. Images of size 224×224×3 are fed into the model, and passed through 19 layers of convolutional and pooling layers, which helps in extracting multi-level features from aerial images. SoftMax is then used in the final layer to classify the extracted features. The features are extracted across layers using the following mathematical equation [24][25]:

$$A_m = ReLU\left(W_m * A_{\{m-1\}} + b_m\right)$$

In this equation, $\mathrm{A_m}$ represents the output of layer m after applying convolution and activation, which is calculated using the convolution weights W_m, the features extracted from the previous layer $A_{\{m-1\}}$, and the bias b_m. ReLU activation function is applied.

$$P_m = MaxPool(A_m)$$

where P_m represents the pooling output after applying Max Pooling, which reduces the size of extracted features.

3.3. Feature Fusion

Feature Fusion The technique that is proposed is a feature fusion-based classification approach with MobileNet and VGG19 features. The features are combined through a hierarchical feature fusion, in which features of different levels are combined to obtain a representation of the aerial images. Feature selection methods are applied after feature fusion to identify the most important features in classification. This is achieved based on feature importance analysis for all the trees in the random forest model, such that non-significant features are eliminated so that the computational efficiency of the model is enhanced without compromising its classification accuracy.

3.4. Random Forest Classification

Random Forest is one of the bagging algorithms. This model is based on creating an ensemble of decision trees, where each tree is trained on a random set of features. The final class is determined based on the votes of the trees, which reduces the possibility of overfitting and improves classification accuracy. In this paper, a random forest model is trained using the extracted final features [26][27].

4. RESULTS

The performance of different feature extraction models was evaluated using four classification algorithms, namely SVM, Random Forest, XGBoost, and Neural Network. Table 1 shows the classification accuracy of each model with each classifier.

Table 1: Performance of feature models with different classifiers

Feature Model	SVM	Random Forest	XGBoost	Neural Network
MobileNetV2_VGG19	91.43	97.14	92.62	96.67

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

MobileNetV2	85.71	93.33	86.90	93.33
VGG19	63.81	72.38	64.29	73.57
ResNet50	60.48	27.14	65.95	32.14

The average performance of each model across all classification algorithms was calculated, based on the criteria of Accuracy, Precision, Recall, and F1-Score. Table 2 shows the calculated results.

Table 2: Average model performance comparison

Evaluation Criteria	MobileNetV2_VGG19	MobileNetV2	VGG19	ResNet50
Average Accuracy	94.46%	89.31%	68.51%	46.93%
Average Precision	94.32%	88.95%	68.23%	46.52%
Average Recall	94.48%	89.35%	68.61%	47.12%
Average F1-Score	94.39%	89.13%	68.41%	46.82%

The performance of each model was analyzed using Precision, Recall, and F1-Score for each classifier. Table 3 shows the detailed results.

Table 3: Analysis of confusion matrices for each model with different classifiers

Feature Model	Classification Model	Precision	Recall	F1-Score
MobileNetV2_VGG19	SVM	0.91	0.91	0.91
MobileNetV2_VGG19	Random Forest	0.97	0.97	0.97
MobileNetV2_VGG19	XGBoost	0.92	0.92	0.92
MobileNetV2_VGG19	Neural Network	0.96	0.96	0.96
MobileNetV2	Random Forest	0.85	0.86	0.85
MobileNetV2	SVM	0.93	0.93	0.93
MobileNetV2	XGBoost	0.87	0.87	0.87
MobileNetV2	Neural Network	0.93	0.93	0.93
VGG19	Random Forest	0.64	0.63	0.62
VGG19	SVM	0.72	0.72	0.71
VGG19	XGBoost	0.64	0.64	0.63
VGG19	Neural Network	0.74	0.73	0.73
ResNet50	Random Forest	0.60	0.60	0.59
ResNet50	SVM	0.42	0.27	0.25
ResNet50	XGBoost	0.66	0.65	0.64
ResNet50	Neural Network	0.30	0.32	0.27

The results showed that combining features using MobileNetV2 and VGG19 achieved the highest classification accuracy of 97.14% when using Random Forest, followed by Neural Network at 96.67%.

5. DISCUSSION

The results of the experiments showed that the combination of features between MobileNetV2 and VGG19 resulted in the highest classification accuracy compared to each model used individually. The Random Forest performed the best when the features were combined with a classification accuracy of 914%, followed by the Neural Network with 96.67%. This indicates that combining features across different models improves performance by leveraging the diverse features learned across different models. On the other hand, the worst performer was ResNet50 with a model accuracy of 27.14% when using the Random Forest and 32.14% when using the Neural Network. The comparison shows that MobileNetV2 and VGG19 together provide better feature representation than either model used in isolation. This is because MobileNetV2 focuses on extracting low-complexity features, while VGG19 extracts more detailed features. When combining the features extracted from both models, performance was improved, confirming that combining models can enhance classification accuracy. For classifiers, Random Forest performed better than the rest of the classifiers, especially when using feature fusion, indicating that it is very suitable for classifying aerial images. In contrast, XGBoost and SVM performed relatively less but still achieved good accuracy. The deep neural network performed well, especially with MobileNetV2_VGG19, where it achieved an accuracy of 96.67%.

When analyzing the confusion matrices, we found that the Precision, Recall, and F1-Score were higher when using MobileNetV2_VGG19, which means that the model was not only more accurate, but was also able to reduce errors in class prediction and balance at the same time.

Despite the study's results, there are still some challenges that can be improved in the future, such as increasing the complexity of the models: it may be useful to test other models such as Vision Transformers (ViTs) or use Recursive Convolutional Networks (RCNNs) to improve performance. Data quality can also be improved using data augmentation techniques to improve model performance and reduce the need for large amounts of labeled data. Finally, although feature fusion has proven effective, it increases computational complexity, so feature pruning techniques can be tried to reduce dimensionality without losing important information.

6. CONCLUSIONS

This paper argues the necessity of integrating deep knowledge with traditional machine learning methods in remote sensing scene recognition. Through the combination of MobileNetV2 and VGG19 features and the analysis of them using Random Forest, it showed better results than the individual models. Consequently, it can be deduced from the fact that integrating different types of models satisfies the desired feature representation, which in turn contributes to the better classification model's operation. Outcomes reported that Random Forest was the one that exhibited the maximum proficiency among the mentioned classifiers, which, in turn, surpassed methods such as SVM and XGBoost, and besides pushed aside the artificial neural networks methodologies. The addition of the characteristics fusion technique was reported to have aided in the attainment of a balance among both the classification accuracy and the elimination of the information deficit, which consequently increased the performance of the system. Nevertheless, there are a number of issues that could be addressed in the future, such as the complexity of the models by trying new methods, increasing data quality, and reducing computational complexity. As this study conveys, combining models and deep learning technologies, on the one hand, can provide for more efficient solutions in remote sensing-image classification and, on the other hand, is the first step for further development, which is accurate classification of ground scenes.

REFERENCES:

[1] L. Khelifi and M. Mignotte, "Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis," Jan. 01, 2020, Cornell University. doi: 10.48550/arxiv.2006.05612.

[2] R. Hong, J. Park, S. Jang, H.-J. Shin, H. Kim, and I. Song, "Development of a Parcel-Level Land Boundary Extraction Algorithm for Aerial Imagery of Regularly Arranged Agricultural Areas," Mar. 18, 2021, Multidisciplinary Digital Publishing Institute. doi: 10.3390/rs13061167.

[3] S. Mahyoub, H. Rhinane, M. Mansour, A. Fadil, Y. Akensous, and A. S. Ahmed, "THE USE OF DEEP LEARNING IN REMOTE SENSING FOR MAPPING IMPERVIOUS SURFACE: A REVIEW PAPER," ☐The ☐international archives of the photogrammetry, remote sensing and spatial information sciences/International archives of the photogrammetry, remote sensing and spatial information sciences. Copernicus Publications, p. 199, Jan. 11, 2022. doi: 10.5194/isprs-archives-xlvi-4-w3-2021-199-2022.

[4] S. Ahuja, M. Michael, M. Selvan, and V. Mantri, "Natural disaster detection in social media and satellite imagery," Jan. 01, 2022, EDP Sciences. doi: 10.1051/itmconf/20224403010.

[5] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities," Jan. 01, 2020, Institute of Electrical and Electronics Engineers. doi: 10.1109/jstars.2020.3005403.

[6] G. Câmara, "On the semantics of big Earth observation data for land classification," Jun. 25, 2020, University of Maine. doi: 10.5311/josis.2020.20.645.

[7] N. Horning, "Land cover mapping with ultra-high-resolution aerial imagery," Dec. 01, 2020, Wiley. doi: 10.1002/rse2.189.

[8] C. Zhang et al., "Improved Remote Sensing Image Classification Based on Multi-Scale Feature Fusion," Jan. 08, 2020, Multidisciplinary Digital Publishing Institute. doi: 10.3390/rs12020213.

[9] W. A. Ezat, M. M. Dessouky, and N. A. Ismail, “Multi-class Image Classification Using Deep Learning Algorithm,” Jan. 01, 2020, IOP Publishing. doi: 10.1088/1742-6596/1447/1/012021.
[10] S. Ghosh and A. Chatterjee, “T-Fusion Net: A Novel Deep Neural Network Augmented with Multiple Localizations Based Spatial Attention Mechanisms for Covid-19 Detection,” in Communications in computer and information science, Springer Science+Business Media, 2024, p. 213. doi: 10.1007/978-3-031-58174-8_19.
[11] X. Zhang, Y. Guo, and X. Zhang, “Deep convolutional neural network structure design for remote sensing image scene classification based on transfer learning,” Sep. 01, 2020, IOP Publishing. doi: 10.1088/1755-1315/569/1/012046.
[12] W. V. der Vaart and K. Lambers, “Applying automated object detection in archaeological practice: A case study from the southern Netherlands,” Jun. 18, 2021, Wiley. doi: 10.1002/arp.1833.
[13] H. Lei, "Deep learning techniques for remote sensing image analysis," in Deep Learning for Remote Sensing Images with Open-Source Software, 1st ed., Wiley, 2024, ch. 4. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781394234271.ch4.
[14] F. G. Yaşar and S. Utku, "Performance Comparison of CNN Based Hybrid Systems Using UC Merced Land-Use Dataset," Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, vol. 25, no. 75, pp. 725-737, 2023. [Online]. Available: https://avesis.deu.edu.tr/yayin/dd5cd71a-61df-44c6-b6e5-a4d6c6daf8f6/performance-comparison-of-cnn-based-hybrid-systems-using-uc-merced-land-use-dataset.
[15] S. Sürücü and E. Demirkıran, "CONVCAT: A New Classification Approach Using UC Merced and RESISC45 Datasets," Eurasian Journal of Science Engineering and Technology, pp. 9-15, 2024. [Online]. Available: https://dergipark.org.tr/en/pub/ejset/article/1417172.
[16] Z. Zhang, X. Cui, Q. Zheng, and J. Cao, "Land Use Classification of Remote Sensing Images Based on Convolution Neural Network," Arabian Journal of Geosciences, vol. 14, article no. 267, 2021. [Online]. Available: https://link.springer.com/article/10.1007/12517-021-06587-5.
[17] M. S. Alam and R. Kim, "A Hybrid Deep Learning Framework for Satellite Image Classification," in Lecture Notes in Networks and Systems, vol. 830, Springer, 2024, pp. 241-255. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-97-9124-8_15.
[18] A. F. O'Leary, "Advancements in Remote Sensing Image Processing: A Review," Computers & Geosciences, vol. 167, pp. 104-113, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0141938222000592.
[19] R. Tombe and S. Viriri, "Adaptive Deep Co-Occurrence Feature Learning Based on Classifier-Fusion for Remote Sensing Scene Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 155-164, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9173451.
[20] “UC Merced Land Use Dataset.” Jan. 2024. Accessed: Mar. 01, 2025. [Online]. Available: https://academictorrents.com/details/e9ac5edf285a43309e57e1289e8816a4e78a937c
[21] Y. Wang et al., “A survey on deploying mobile deep learning applications: A systemic and technical perspective,” Jun. 16, 2021, KeAi. doi: 10.1016/j.dcan.2021.06.001.
[22] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, "A novel image classification approach via dense-MobileNet models," Mobile Information Systems, vol. 2020, no. 1, p. 7602384, 2020.
[23] D. Sinha and M. El-Sharkawy, "Thin mobilenet: An enhanced mobilenet architecture," in 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2019, pp. 0280-0285.
[24] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," in 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), vol. 1, Nov. 2021, pp. 96-99, IEEE.
[25] M. Elhousni, X. Huang, and Z. Zhang, “Aerial Height Prediction and Refinement Neural Networks with Semantic and Geometric Guidance,” Jan. 01, 2020, Cornell University. doi: 10.48550/arxiv.2011.10697.
[26] C. Komusiewicz, P. Kunz, F. Sommer, and M. Sorge, “On Computing Optimal Tree Ensembles,” Jan. 01, 2023, Cornell University. doi: 10.48550/arXiv.2306.
[27] I. D. Mienye and N. Jere, “A Survey of Decision Trees: Concepts, Algorithms, and Applications,” Jan. 01, 2024, Institute of Electrical and Electronics Engineers. doi: 10.1109/access.2024.3416838.