Eurasian Journal of Science Engineering and Technology

Research Article

CONVCAT: A NEW CLASSIFICATION APPROACH USING UC MERCED AND RESISC45 DATASETS

Selim SÜRÜCÜ[1,*], Esma DEMİRKIRAN[2]

[1] *Çankırı Karatekin University, Faculty of Engineering, Department of Computer Engineering, 18100, Çankırı, Türkiye*
[2] *Yeşilova Anatolian High School, Hatay, Türkiye*

ABSTRACT

With advances in Earth observation systems, the importance of remote sensing data is increasing daily. These data are used in various fields ranging from image segmentation to terrain classification, from disaster impact assessment to climate change analysis. The use of remotely sensed images for terrain classification has been the subject of a number of studies. This study proposes a new method for terrain classification in the UC Merced Land Use Dataset and RESISC45 remote sensing images. This method is called ConvCat model, which is a combination of classical convolutional layer and CatBoost models. The performance of this model is measured in terms of accuracy, the Matthews Correlation Coefficient (MCC) and the Cohen's Kappa metrics. The results are compared with ensemble models (XGBoost, CatBoost), with ConvXGB, a combination of convolutional learning and XGBoost, and with ResNet50, one of the most widely used transfer learning models. The developed ConvCat model outperformed the other models, achieving an accuracy of 97.44% on the UC Merced data set and an accuracy of 96.89% on the Resisc45 data set. This study shows that our newly developed model provides the best results for the classification problem based on remote sensing images.

Keywords: Remote Sensing, Scene Classification, Ensemble Learning, ConvCat, Ground Observation Classification

1. INTRODUCTION

Remote sensing (RS) refers to the process of remote observation, recording and measurement of the properties of a surface or object and the technology that makes it possible. Remote sensing can be performed with many different tools and devices such as satellites, aircrafts (aircraft, drones, etc.). Today, with the increase in the number of satellites of states and companies, there is a high increase in UA data. Researchers are carried out on these data in many different fields from agriculture to defense.

Information extraction from UA data can usually be done by image classification techniques. UA image classification attempts to assign an area or pixel to a class (label) using certain features. These features can be spectral characteristics, colour or building materials. These features can be used for classification processes in many different subjects and areas, from crops in agricultural lands to the condition of forested areas, from damaged areas to the detection of landslide areas.

Scientists and researchers have made great efforts to develop advanced classification approaches and techniques to improve the accuracy of remote sensing image classification. With the advances in artificial intelligence technologies, there has been a great increase in the number of studies in the field of UA.

Recent years have seen significant advances in the application of Artificial Intelligence (AI) and in particular Deep Learning (DL) techniques to remote sensing imagery. These techniques allow useful information to be extracted from satellite imagery. This information can be used in a variety of applications. For example, it can be used in agricultural activities such as monitoring of the condition of agricultural land, identification of crop species and environmental monitoring through forest tracking. It can also be used for damage assessment in the event of disasters, while the tracking of objects (vehicles, ships, aircraft, etc.) can be used for many different reasons, such as defense and security.

The contributions of this paper are;

- This paper proposes a new convolution-based model for classifying UA data. This proposed method is run on UC Merced and RESISC45 datasets.
- The performance of this model is compared with another convolution-based model, ConvXGB, as well as classical deep learning methods.
- Cohen cappa and MCC metrics, which are rarely used in remote sensing imagery, were used to measure model success.
- The proposed model seems to give the most successful results.

* Corresponding author, e-mail: selimsurucu@karatekin.edu.tr, (S. Sürücü)
Recieved: 09.01.2024 Accepted: 06.05.2024
doi: 10.55696/ejset.1417171

After the introduction, the article continues with a literature section including recent studies. Section 3 is the methodology where the datasets, models and metrics are explained. Section 4, the results section, presents the results obtained from the developed model and other models. The discussion section, which includes other studies and general comparisons, takes the 5th place in the paper. The last section is conclusion and future work section.

2. LITERATURE REVIEW

In their study using resisc45 and UCMerced datasets, Tombe et al. proposed a method called deep coherence feature learning (ADCFL) [1]. The ADCFL method uses a convolutional neural network to extract spatial feature information from an image in combination with filters, which is then fed into a multigranular forest for feature learning and classification through majority voting with ensemble classifiers. Their method combined with VGG16 gave the best result with an accuracy of 91.05% on the Resisc45 dataset.

In a study on agricultural lands in Ukraine, an unsupervised neural network and a supervised neural network model were proposed for land cover and crop type classification from multi-temporal multi-source satellite images. The images used in the study were obtained from Landsat-8 and Sentinel-1A satellites, and an accuracy of 85% was achieved [2].

In another study, a peak-wave sparse cooperative representation convolutional neural network (RRSCRN) algorithm is proposed for remote sensing image classification. RRSCRN enables to learn the relationship between competition and co-operation of the remote sensing image to be sparsely classified, and to better learn the features of the remote sensing image. This algorithm is tested on Pavia University and Indian Pines datasets. This algorithm has shown better performance compared to traditional remote sensing image classification methods [3].

Yu et al. (2018) tried to classify remote sensing images based on the improved KNN (k nearest neighbour) algorithm. They improved the KNN algorithm with two methods. Firstly, they used the weighted distance to improve the locality of the KNN algorithm. Secondly, they used support vector machine to improve the accuracy. For the University of Pavia dataset, the accuracy of the improved KNN algorithm is 95.6%, while the accuracy of the traditional KNN algorithm is 94.8%. For the Pines dataset, accuracy values of 92.7% and 92.1% were obtained, respectively [4].

In another study, multi-label remote sensing image classification was studied using capsule networks, which can better capture the features of objects than traditional neural networks. For the University of Pavia dataset, the accuracy value of the capsule networks was 96.9%, while the accuracy value of the traditional neural networks was 95.2%. For Indian Pines dataset, 94.8% and 93.9% accuracy values were calculated respectively. In this study, it is shown that capsule networks achieve a higher classification performance than traditional neural networks. In the study in which success measurement was performed using RESISC45 dataset with ANN by feature extraction with enhanced attention module, an accuracy value of 94.29% was obtained. It was observed that feature extraction with the enhanced attention module achieved good performance compared to classical methods [5].

Dai et al. proposed an improved remote sensing scene image classification model to improve the overall accuracy (OA) of the AlexNet model [6]. They used Layer Normalization (LN) instead of Local Response Normalization (LRN) in AlexNet and modified the model by changing the convolution kernel to 7 × 7. They also used Block Attention Module (CBAM) and Compression and Excitation Module (SEM) to reduce the effect of backgrounds. An accuracy of 96.29% was achieved on the Resisc45 dataset.

In another study, Zhao et al. proposed a model called CGINet. They developed this model by combining global context and part-level discriminative features in a unified framework [7]. In another study, a hierarchical feature fusion with patch expansion of the transformer (HFFT-PD) was developed. The proposed model consists of a hierarchical transformer fusion (HTM) block and a lightweight adaptive channel compression (LACC) module. The most important contribution of this work is the newly proposed Patch Expansion strategy. In this strategy, it acts as a reassembly operator based on patch features [8]. In the artificial intelligence model developed by Sharma and Gupta with a hybrid approach, they used three attention mechanisms as channel, spatial and local relation attention with EfficientNet model. They dimensionally reduced the feature vector with PCA and performed classification with SVM. They tested this model on UC Merced, EuroSAT, MLRSNet and RSI-CB256 datasets and obtained an accuracy of 97.11% in UC Merced dataset [9].

3. MATERIAL AND METHOD

3.1 Dataset

Two datasets were used in this study.

UC Merced Land Use Dataset = UC Merced is a dataset created from 21 different urban areas. These classes are agriculture, aircraft, baseball diamond, baseball diamond, beach, buildings, chaparral, chaparral, denserresidential, dense residential, forest, motorway, golf course, golfcourse, port, intersection, mediumresidential, mobilehome park, mobilehomepark, overpass, parking lot, river, runway, sparseresidential, storage tanks, tennis court, tenniscourt. In this dataset, which contains 2100 images in total, each class consists of 100 images. In addition, these images are RGB images with 256x256 pixel dimensions.

RESISC45 = It is a dataset used for Remote Sensing Image Scene Classification. The RESISC45 dataset consists of a set of satellite images containing 45 different classes of natural and human scenes. These classes include forests, farmland, motorways, beaches, lakes, buildings and other landscape elements. Each image consists of multi-spectral and/or hyper-spectral data from satellite sensors. The images contain high resolution pixels collected at different wavelengths.

To compare these two datasets;

- The UC Merced Land Use Dataset was launched in 2010, while RESISC45 was launched in 2017. As can be understood from this point, RESISC45 is a new dataset compared to the other dataset.
- While the UC Merced Land Use Dataset contains 2100 images in 21 different categories, the RESISC45 dataset contains 31500 images in 45 different categories. As can be seen, the RESISC45 dataset contains a much broader categorisation and more images, while the UC Merced Land Use Dataset focuses on specific land use and land cover types.
- Both datasets are 256x256 pixels in size and RGB.

3.2 Metrics

In this study, 3 different metrics were used to calculate the measurement of the success of the models.

Accuracy = It is a measure of success obtained by dividing the correctly predicted data (TP+TN) by the total data and gives a value between 0%-100% [10].

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

Matthews Correlation Coefficient (MCC) = MCC measures how well the classification results match the actual class labels and is used to calculate the classification performance [11]. A result between -1 and +1 is produced. If +1, this means a perfect classification.

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP+FP)(TP+FN)(TN+FN)(TN+FN)}} \quad (2)$$

Cohen’s Cappa = Cohen's Kappa is calculated using a table based on the probability of observers being placed in their randomly selected category. The coefficient value usually ranges between -1 and 1. 1 means perfect agreement, 0 means randomness and -1 means complete disagreement [12].

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (3)$$

Meanings of the expressions in the equation;

- p_e = Overall accuracy of the model
- p_0 = Agreement between model predictions and actual class values

3.3 Models

3.3.1 Convolutional neural networks (CNNs)

CNNs are a type of deep learning model designed to mimic the image processing ability of the human brain and are used in many fields such as image processing and natural language processing. CNNs consist of multiple convolution layers to extract features of the image and one or more fully connected layers to classify the image. In addition to these layers, there are many different layers such as the pooling layer, which is used to reduce the output of the convolution layers, the input layer to receive the input data to the neural network, and the dropout layer to reduce overfitting during training.

CNNs constitute one of the most basic processes of machine learning. It is widely used as a basic model for various tasks from image classification to object detection.

3.3.2 XGBoost

XGboost is a kind of gradient boosting method. Gradient Boosting is an ensemble learning method that builds a stronger model by combining the predictions of multiple models [13]. Gradient Boosting builds a powerful learning model by combining simple models whose predictions are often inaccurate [14]. Gradient boosting methods are optimized for high performance, scalability and accuracy. The XGBoost algorithm is a high-performance version of the Gradient boosting algorithm optimized with various modifications. This algorithm is among the most preferred algorithms among decision tree algorithms for many reasons such as the best performance, high prediction power, and the ability to manage empty data.

3.3.3 CatBoost

CatBoost is based on an algorithm similar to XGBoost. CatBoost differs from XGBoost by using a more efficient technique to process categorical data. In addition to this difference, it can work with larger data sets than the XGboost algorithm. As a powerful algorithm in machine learning studies, it can be used for many tasks such as classification, regression and time series prediction.

3.3.4 ConvXGB (CNN + XGBoost)

Thongsuwan et al. developed a new deep learning approach for the classification problem that combines the performances of a convolutional neural network and Extreme Gradient Boosting (XGBoost) models. Figure 1 shows the architecture of this proposed model [15].

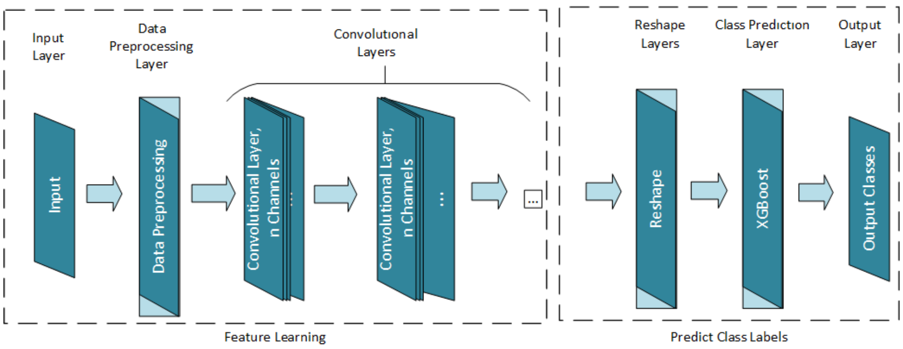

Figure 1. Architecture of the ConvXGB model

3.3.5 ResNet50

In 2015, He et al. announced in a paper titled "Deep Residual Learning for Image Recognition". The expression 50 in the nomenclature refers to the total number of convolution layers. It was developed as a solution to the performance degradation experienced in very deep networks [16]. As a solution method, shortcuts have been added between network layers. ResNet50 has been a turning point for deep learning networks. Good results can be obtained for many problems compared to other models. The architecture of this model is given in Figure 2.

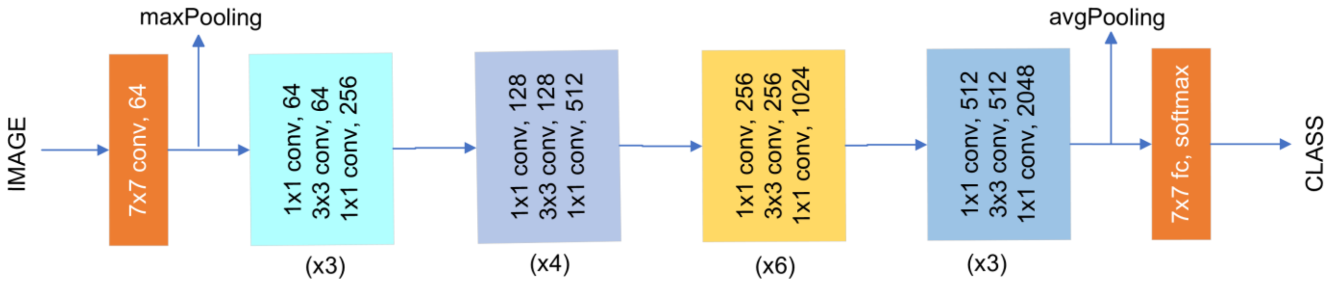

Figure 2. Architecture of the ResNet50 (Layers are successively interconnected.)

3.3.6 ConvCat (CNN + CatBoost)

In this study, we developed the ConvCat model inspired by the ConvXGB model. In our architecture, CatBoost algorithm is used for the class prediction task after the convolution layers. The architecture of this proposed model is shown in Figure 3.

Figure 3 shows that the classification layer, which is the last layer of the classical cnn structure, has been removed. In its new form, the classical CNN structure was used for feature extraction. After feature extraction, the extracted features were given to the catboost part of the model and classification was performed. By combining these two models, the feature extraction of the CNN structure and the power of the catboost model in classification are combined.

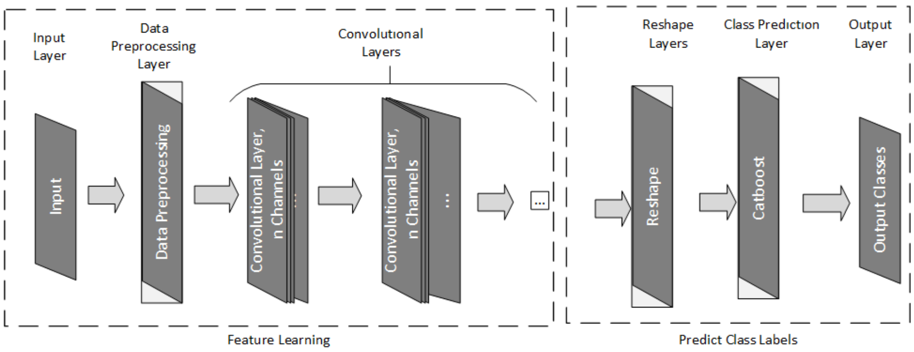

Figure 3. Architecture of our proposed ConvCat model

4. RESULTS AND DISCUSSION

In this study, 30 epoch trainings were performed on a machine with v100 GPU as the working environment and results were obtained.

When Table 1 is analysed, the ConvCat model we developed gave better results than the ConvXGB model developed by Thongsuwan et al. The ConvCat model achieved about 2 times more success than the classical ensemble (XGBoost, CatBoost) models. In addition, ConvCat and ConvXGB showed high success compared to the ResNet model, one of the most common transfer learning models. In addition to the accuracy metric, it was also observed that it gave the best results according to the MCC and Cohen's Cappa metrics.

Table 1. Results Obtained From models according to UC Merced and Resics45 Datasets

	UC Merced			RESISC45		
	Accuracy	**MCC**	**Cohen's Cappa**	**Accuracy**	**MCC**	**Cohen's Cappa**
XGBoost	42%	0.478	0.4425	70%	0.707	0.7001
CatBoost	55.6%	0.554	0.5236	66.6%	0.65	0.677
ResNet50	80.88%	0.80	0.7789	75%	0.755	0.781
ConvXGB	87%	0.857	0.866	82.75%	0.838	0.8155
ConvCat	**97.44%**	0.971	0.962	**96.89%**	0.9645	0.9505

Dai et al. obtained one of the best results with the RESISC45 dataset used in this study. Dai et al. obtained a good result with 96.29% with the improved version of the AlexNet model. With our proposed model, ConvCat, we obtained 96.89% accuracy. Compared to other studies in the literature, we obtained the best accuracy on the Resisc45 dataset. Table 2 gives an overview of the results.

Table 2. Results of studies with the Resics45 dataset in the literature

Paper	Method	Result (Accuracy - %)	Year
[1]	VGG16-Classifier-fusion(proposed)	91.05	2021
[4]	ANN	94.29	2021
[6]	Improved AlexNet (proposed)	96.29	2022
Our Proposed	ConvCat	96.89	

When the studies in the literature using the UC Merced dataset are examined, it is seen that the accuracy values are very close to each other. Using the same dataset, we obtained an accuracy of 97.44% with our ConvCat model. Although we obtained a better result than the studies in the literature, there is a very small difference. Table 3 shows the information about the studies in the literature.

Table 3. Results of studies with the UC Merced dataset in the literature

Paper	Method	Result (Accuracy - %)	Year
[1]	VGG16-Classifier-fusion(proposed)	96.55	2021
[6]	Improved AlexNet (proposed)	96.57	2022
[9]	EfficientNet + PCA+SVM	97.11	2023
Our Proposed	ConvCat	97.44	

5. CONCLUSION

This study shows that instead of using classical deep learning models for image classification problems, better results can be obtained by combining several models. In the combination models, the choice of models can be random as long as the inputs and outputs are taken into account. However, it is recommended to prefer transfer learning models in the feature extraction phase, as transfer learning models are more successful in feature extraction than other models. For classification, ensemble models can be preferred. This study shows that for the problem of terrain classification in remote sensing images, the combination of classical convolutional layers in the feature extraction phase and the CatBoost model, which is an ensemble model used as a classifier, gives the best result in the literature. In addition, it may be preferable to combine new models, such as Transformar, which has been developed in the last few years, with the classical models.

Furthermore, most of the studies on terrain classification in the literature have measured their success in terms of the accuracy metric. In this study, the MCC and Cohen's Cappa metrics were used in addition to the accuracy metric. These metrics have been used in land classification studies and have been shown to be useful.

In future studies, the ConvCat model proposed in this study can be adapted and used for many problems other than the classification of land. It can be preferred as a new approach to classifying problems.

SIMILARITY RATE: 14%

AUTHOR CONTRIBUTION

First Author: initiated the research idea, developed, analyzed and interpreted the data and wrote the manuscript
Second Author: suggested the research methods, structured the paper and edited the manuscript.

CONFLICT of INTEREST

The authors declared that they have no known conflict of interest.

REFERENCES

[1] R. Tombe and S. Viriri, "Adaptive Deep Co-Occurrence Feature Learning Based on Classifier-Fusion for Remote Sensing Scene Classification," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 155-164, 2021, doi: 10.1109/JSTARS.2020.3044264.

[2] N. Kussul, M. Lavreniuk, S. Skakun and A. Shelestov, "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778-782, May 2017, doi: 10.1109/LGRS.2017.2681128.

[3] F. An and J. Liu, “Remote Sensing Image Classification algorithm based on ridge wave sparse collaborative representation convolutional neural network,” Multimedia Tools and Applications, vol. 80, no. 21–23, pp. 33099–33114, 2021. doi:10.1007/s11042-021-11406-w

[4] Z. Yu, "Research on Remote Sensing Image Terrain Classification Algorithm Based on Improved KNN," 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 2020, pp. 569-573, doi: 10.1109/ICISCAE51034.2020.9236884.
[5] Z. Zhao, J. Li, Z. Luo, J. Li and C. Chen, "Remote Sensing Image Scene Classification Based on an Enhanced Attention Module," in IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 11, pp. 1926-1930, Nov. 2021, doi: 10.1109/LGRS.2020.3011405.
[6] D. Dai, Xu, W., and Huang, S., “Improved AlexNet and embedded multi-attention for remote sensing scene image classification”, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2022, vol. 12506. doi:10.1117/12.2661951.
[7] Zhao, Y., Chen, Y., Xiong, S., Lu, X., Zhu, X. X., & Mou, L., “Co-Enhanced Global-Part Integration for Remote-Sensing Scene Classification.” IEEE Transactions on Geoscience and Remote Sensing. 62, 1-14, 2024.
[8] Chen, X., Ma, M., Li, Y., Mei, S., Han, Z., Zhao, J., & Cheng, W., “Hierarchical Feature Fusion of Transformer With Patch Dilating for Remote Sensing Scene Classification.” IEEE Transactions on Geoscience and Remote Sensing. 61, 1-16. 2023.
[9] Sharma, I., and Savita G.. "A Hybrid Machine Learning and Deep Learning Approach for Remote Sensing Scene Classification." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.
[10] M. Hossin, Sulaiman M.N, “A Review on Evaluation Metrics for Data Classification Evaluations”, in International Journal of Data Mining & Knowledge Management Process, 2015, 10.5121/ijdkp.2015.5201.
[11] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, ‘Assessing the accuracy of prediction algorithms for classification: an overview’, Bioinformatics, vol. 16, no. 5, pp. 412–424, 05 2000.
[12] J. Cohen. “A Coefficient of Agreement for Nominal Scales.” Educational and Psychological Measurement 20 (1960): 37 - 46.
[13] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’, 08 2016, pp. 785–794.
[14] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.
[15] S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal, ‘ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost’, Nuclear Engineering and Technology, vol. 53, no. 2, pp. 522–531, 2021.
[16] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2015.