Journal of Information Systems Engineering and Management
2025, 10(12s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Enhancing Remote Sensing Image Classification and Interpretability: A Multi-Stage Feature Extraction Approach and Grad-CAM

Dr S Jayanthi[1*], Vempati Krishna[2], Dr. Khalid Mohiuddin[3], Dr. Osman A. Nasr[4], Dr.Ijteba Sultana[5], Pannangi Rajyalakshmi[6]

[1*]*Senior Assistant Professor, Department of AI & DS, Faculty of Science and Technology (IcfaiTech), The ICFAI Foundation for Higher Education (IFHE), Hyderabad, Telangana – 501503, India, drsjayanthicse@gmail.com*
[2] *Professor, Department of CSE-DS, TKR College of Engineering & Technology, Hyderabad, India, vempati.k@gmail.com*
[3]*King Khalid University, Department of Business Informatics, College of Business, KSA, kalden@kku.edu.sa*
[4]*King Khalid University, Department of Business Informatics, College of Business, KSA, oanassr@kku.edu.sa*
[5]*Associate Professor, Department of Computer science and Engineering, ISL Engineering college, Hyderabad, India, drijtebasultana@gmail.com*
[6] *Assistant professor, Department of CSE, TKR College of Engineering & Technology, Hyderabad, India, pannangiraji@gmail.com*

ARTICLE INFO

Received: 21 Nov 2024

Revised: 05 Jan 2025

Accepted: 25 Jan 2025

ABSTRACT

Relevant and current land use data remain critical for effective spatial layout, environmental stewardship, and bestow resources. This study assesses the classification accuracy of machine learning classifiers developed for land use/ classification (LUC) and how effective are they when combined with feature engineering and augmentation strategies on a benchmark dataset. The processing pipeline uses a Convolutional Neural Network (CNN) model called ResNet-50 for image feature extraction, which captures spatial patterns and also performs well in classification tasks. Model performance is enhanced by using a combination of the pre-trained ResNet-50 as an input for self-compiled image classifiers and applying Principal Component Analysis (PCA) for dimensionality reduction. Advanced-Data augmentation(ADA) techniques further improve the generalization of the models. Of all tested classifiers (Logistic Regression, Random Forest, SVM, Gradient Boosting and XGBoost), SVM outperformed the rest with AUC-ROC at 0.993 and MCC at 0.843. In addition, Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations applied to enable an understanding of the decision processes of the model, thus increasing its interpretability. This work demonstrates the combining of deep feature extraction with classical machine learning and makes land use classification scalable and robust.

Keywords: Land Use Classification, ResNet-50, Feature Extraction, PCA, Dimensionality Reduction, Support Vector Machine, Advanced-Data Augmentation, AutoAugment, Machine Learning, AUC-ROC, Empirical Study.

I. INTRODUCTION

LUC information is critical for many activities, such as disaster mitigation and environmental management. Traditional methods of LUC mapping are useful, but they have many limitations, such as being labor-intensive, subjective, and difficult to scale. Ground surveys and aerial images are both time-consuming methods of collecting data. Automated LUC mapping has recently come into the picture due to advancements in remote technology and machine learning, allowing for LUC mapping methods that are more accurate, efficient, and easy to scale [1- 4].

The benefits of automation may pose overwhelming challenges, but it can provide solutions in areas such as agriculture, urban development, and social activity in any region. However, utilizing deep learning systems poses a different challenge as many requirements need to be considered. Among other things, the selection of methods and tools for feature extraction, data augmentation, and dimensionality reduction are crucial when trying to optimize deep learning-based LUC systems [5-9].

This research delved into assessing the rendition of machine learning algorithms for land use classification tasks, about the state-of-the-art feature engineering and data augmentation. In particular, a pre-trained ResNet-50 was

applied as it is a high-level, discriminative feature extractor from the imagery. In doing so, all the pre-trained layers were freezed, but the last fully connected layer was replaced with an identity mapping. Thereby, the architecture of the ResNet-50 is slightly modified to serve as a feature extractor. This allowed us to gain access to deeply embedded spatial and contextual features of the land use data. To further curtail the chances of overfitting while optimizing computational resources, PCA [10] was applied to eliminate the less informative dimensions in the extracted feature space.

Additionally, several crucial data augmentation strategies were employed, such as AutoAugment (ImageNet policy) [11], Random Affine transformations, Color Jittering, Random Horizontal and Vertical Flipping, Random Cropping with paddings too, which all assist in augmenting the model's expectation and accuracy toward changes of the images.

The efficacy of five widely used machine learning models, namely Logistic Regression (LR) [12], Random Forests (RF) [13], Gradient Boosted Trees (GBst) [14], Support Vector Machine (SVM) [15], and eXtreme Gradient Boosted (XGBst)[16] classifiers was thoroughly assessed using a comprehensive suite of metrics. This study is set out to meet the following objectives:

- Make scholarly investigations, to assess the readily available ResNet-50 model in terms of its performance in the LUC.
- Test the influence of ADAs on the efficiency and robustness of the classifiers.
- Compare the performance of various machine learning classifiers, using features extracted through overhead methods, and recognize the model with the best performance focused only on the given task.
- Apply feature engineering, dimensionality reduction, and data augmentation to have a balanced impact on the overall accuracy and sustainability of the model.

II. RELATED WORK

This section expounds on the history of the LUC classification techniques from the conventional methods, through statistical learning, the rise of deep learning, big data and cloud computing, and their ethical implications.

The initial LUC classification was based greatly on manual photointerpretation by skilled personnel. While informative, it was a subjective process that was very slow and inconsistent in its application. More objective approaches became possible with the introduction of statistical classifiers such as Maximum Likelihood Classification and clustering algorithms like K-means and ISODATA, which allowed for automated image processing. However, these methods had limitations on spectral variability, mixed pixels, and complicated spatial patterns that confounded efficient classification of landscapes with high heterogeneity.

The incorporation of ML for LUC classification was revolutionary since it was more efficient in the classification of high-dimensional data. The use of SVM [17] and RF [18] algorithms resolved class imbalance and noise problems with AutoML frameworks. Further improvement of the classification was achieved by GBst through iterative prediction adjustments.

GBst has refined classification by improving the predictions in each subsequent iteration. Nevertheless, these models relied on handcrafted features, which limited their ability to capture sophisticated spatial and contextual relationships available in high-resolution imagery.

The development of deep learning models, especially CNNs, changed the face of LULC classification through automated feature extraction and hierarchical learning from raw images. Higher-order models like AlexNet, VGGNet, ResNet, and EfficientNet pinpointed complex spatial arrangements and relationships at unprecedented accuracy [19-21]. For example: Alem and Kumar focus on the effectiveness of the pre-trained and fine-tuned CNNs for the UCMLU dataset with great classification efficiency [22]. Hamza et al. blended deep learning with Bayesian Optimization to reach the highest accuracy levels on many benchmark datasets [23]. Papoutsis et al. created a comparative benchmark of CNNs, Vision Transformers, and EfficientNets on BigEarthNet to find suitable architectures for large-scale LULC tasks [24]. Ma et al. developed and took a look at the Feature Enhancement Network, which was capable of capturing fine granular features such as spatial buildings and water [25]. These studies highlight the promise deep learning holds for the challenges posed by high-resolution remote sensing images.

The introduction of high-resolution remote sensors like Landsat, Sentinel, and LiDAR has significantly increased the scale and intricacy of remote sensing data. The availability of cloud-based platforms such as Google Earth Engine and Amazon Web Services offers new possibilities for fast processing of massive datasets and minimization of computing constraints [26]. There have also been attempts to achieve improved classification with AutoAugment(Cubuk), including Aljebreen et al. [27], who expounded deep learned optimized River Formation Dynamics Algorithm, and Huang et al. [28], who used a modified ResNet-50 with K-mean clustering for rural land-use classification. This research expands from what the earlier learning methods provided while filling the gaps caused by deep learning and classical problem-solving techniques. Key contributions of this study are:

- Using a ResNet-50 model already trained for other contexts, employing it as a feature extractor enabled the capturing of complex spatial and contextual relationships in the imagery.
- Implementing a broad range set of advanced data augmentation methods to include AutoAugment in order to improve model performance and generalization.
- Using PCA-based dimensionality reduction to reduce overfitting, thus improving computation efficiency and pointing out the most discriminative features to classify.
- Evaluating several ML classifiers with a rigorous evaluation, thereby giving an exhaustive assessment of how these classifiers would perform on the extracted features.

This research fills the gap between traditional approaches and deep learning by providing a robust and scalable solution for LUC classification that results in overcoming data scarcity, computational complexity, as well as responsible AI needs.

III. METHODS

A. Dataset

UCMerced LandUse (UCMLU) Dataset [29] was utilized in the study, encompassing 2,100 images grouped into 21 land classes. Each category comprises 100 images, thus resulting in a balanced dataset. The collection includes a variety of land cover types, namely, agricultural areas, urban buildings, roads, natural environments - forests and chaparral, and water bodies-The National Map Urban Area Imagery comprises high-resolution images, which enable investigation of the land surface over the smallest and finest details, thus favoring the superior isolation of land cover types.

B. Data Preprocessing

Comprehensive data preprocessing pipeline were applied for the optimization of the model performance. The data set was normalized in the first place to avoid variations in color or brightness. Therefore, each image was resized according to the size of 224x224 using bicubic interpolation to meet the pre-trained ResNet-50 model's entrance conditions. Second, ImageNet mean ([0.485, 0.456, 0.406]) was subtracted; subsequently, ImageNet standard deviation ([0.229, 0.224, 0.225]) was divided into pictures to make model training effective and help in enhancing good feature extraction. Fig. 1 depicts the image data augmentation techniques applied on remote sensing images which pertain specifically to agriculture and riverine landscapes. The transformations include cropping, flipping the image both ways, changing the colors, and normalizing the image which improves the model by adding different spatial and spectral properties while maintaining the core properties.

To achieve enhanced generalization and robustness of the model, data augmentations were gradually applied to all the labeled data along with some jittering, such as random horizontal and vertical shifts, random rotations within ±15 degrees, random cropping having 4-pixel padding, and color jittering to adjust brightness, contrast, saturation, and hue. Moreover, AutoAugment, a policy pre-trained using ImageNet, was utilized to dynamically learn and perform a sequence of complicated augmentation operations providing diverse visual representations to immeasurably enrich the training data.

Stratified sampling was applied to divide the dataset into training (80%) and testing (20%) sets, ensure proper distribution of data in both the sets, and thereby minimize bias during evaluation. Random seed, 42, was used to assure the reproducibility of the data split in experiments.

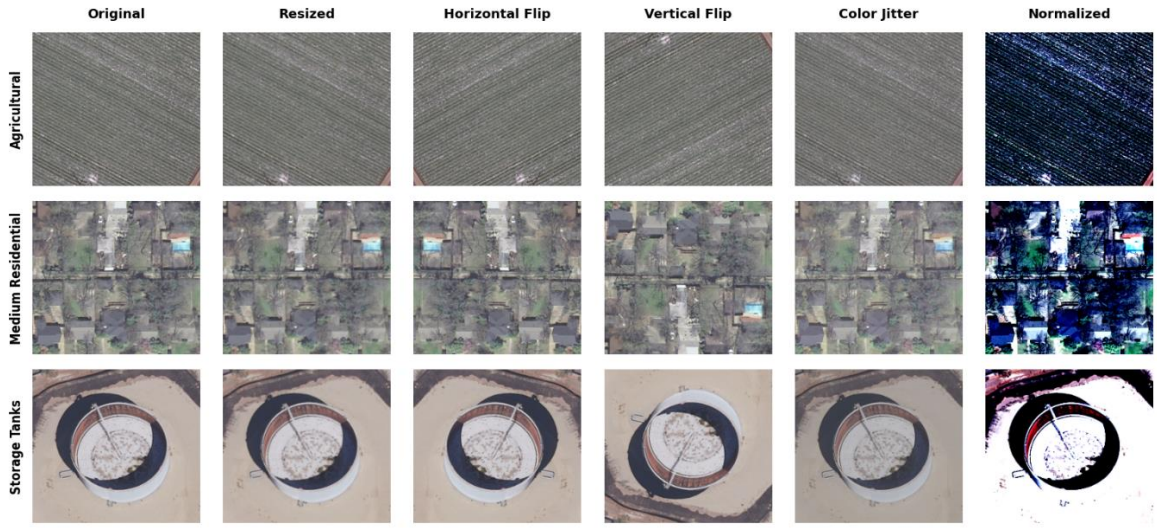

Figure 1: Illustration of Image Data Augmentation Techniques Applied to Remote Sensing Imagery

C. Feature Extraction and Dimensionality Reduction

The pretrained ResNet-50 CNN was used to extract high-level feature representations from the preprocessed images. The convolutional layers of the pretrained model were held frozen, and the penultimate convolutional layer's output was used as the feature vector. This approach capitalizes on the ability of that model to capture detailed spatial and contextual information in images. The PCA was applied to reduce the dimensions of the extracted features and avoid overfitting. It finds principal components or linear combinations of the original features that capture the maximum variance. By projecting the data onto these principal components, the most salient information is retained while the dimensionality is reduced. ResNet-50 Architecture is used for feature extraction is illustrated in Fig.2.

The first step directly involved reducing dimension representative of the extracted features through PCA, with the retained variance amounting to 95%. In the PCA implementation, a seed was utilized (random_state=42) to ascertain at a later time the consistency of the dimension reduction in different experimental replications. The process flow of the proposed model is illustrated in Fig. 3.

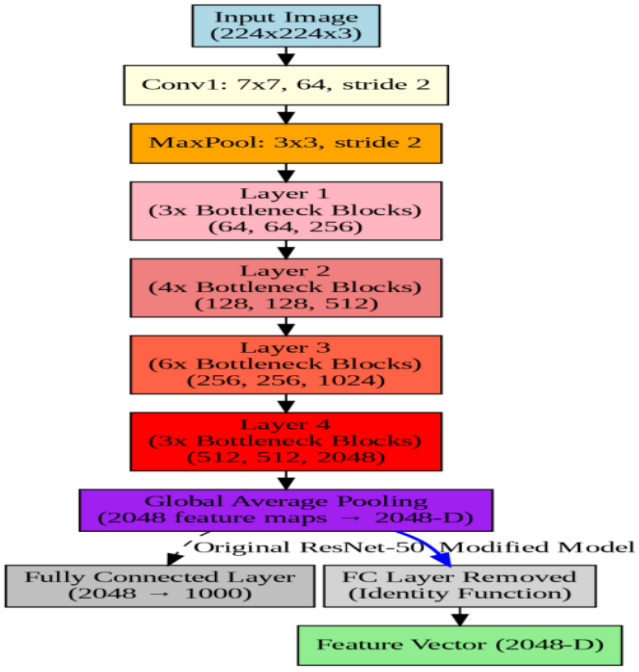

Figure 2 : Modified Resnet50 architecture for Feature Extraction in LUC

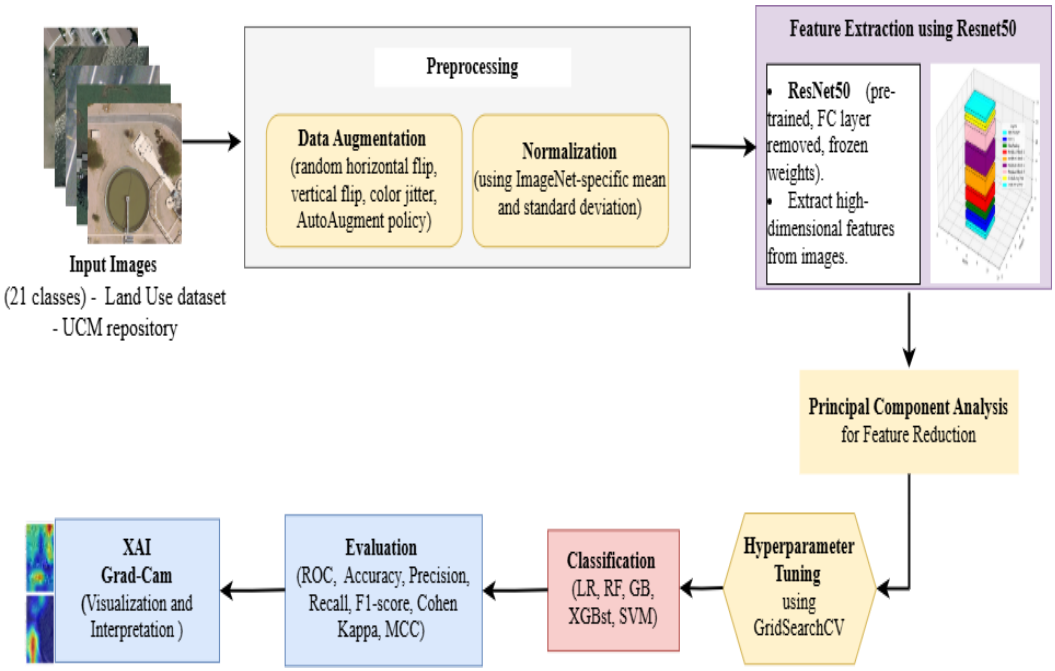

Figure 3: Process flow of the Proposed model

D. Model Selection

The aim in this study was to bridge the gap between deep feature extraction and ML (LR, RF, GB, SVM, and XGBst) approaches based on their capability to cope with the complexity and heterogeneity of aerial imagery. For this purpose, five different models were chosen.

LR is a simple and highly interpretable model used here as a baseline: From its essential encumbrance of linear condition results derived benchmark performance from which the more complicated models could be compared. Finally, here the solver parameters were set to 'lbfgs'—specifically useful for high-dimensional spaces—and iteration was extended to 1000 to ensure convergence.

The ensemble-based RF constructs multiple decision trees and then aggregates their outcomes to improve classification accuracy. It effectively helps with the high dimensionality data through inherent feature selection capabilities. Key parameters considered were 100 estimators, max_features='sqrt', criterion='gini'.

The most important feature in SVM is the kernel whose job is to handle these decision boundaries. For effective performance of selections of land use at multiple scales and rapid land-use classification, we used the RBF kernel, with C = 1.0 for an appropriate level of regularization and gamma set to 'scale' for automatic adaptation to the data.

They were used instead for modeling XGBst, an improved version of the boosting process that had improved computational efficiency and strong regularizing potential to some extent, key to translating complex patterns from aerial imagery. The model was trained with 100 estimators, a learning rate of 0.1, and a maximum tree depth of 3. The penalty terms — alpha (L1) and lambda (L2) — were fine-tuned to improve the regularization, with an eye on generalizing well. Stopping was implemented in cross-validation in order to ensure that no change in performance was brought about while the overall performance condition was still being met.

The selection of these models now turns to fulfilling the purpose in hand, i.e., managing issues related to high dimensions, spatial heterogeneity, and different landscape classes, typical of all aerial imagery. The ensembling methods (RF, GB, XGBst) had been chosen for unleashing their strengths in reducing both variance and bias, in making these models complement and justify themselves well. In combination, kernel-based (SVM) and linear models significantly contribute to the interdisciplinary dynamics regarding the connection of the deep-forged features with the leveraging methods of the various models.

The current method of model selection was well based on previous studies: with those profitably applied in diverse remote sensing tasks, matching paradigm of algorithms have to face thorough scrutiny. The discussion in-detail adds

a deeper grasp to the way that classical machine learning models support deep-features-extraction techniques in aerial-image classification.

The hyperparameter tuning settings of the examined models are contained in table 1. Each model adjusted the most important hyperparameters with the aim of achieving optimal results. For example, Regularization (C) and Iteration limits (max_iter) were set for tuning LR models. RF and GBst models were tuned using tree depth (max_depth), number of estimators (n_estimators), and the algorithm for splitting. SVM models were evaluated using different kernel types, regularization (C), and gamma parameters. XGBst was tuned using n_estimators, learning_rate, max_depth, and subsample.

This tuning process made sure that the evaluation of the model was done for varying hyperparameter values.

Table.1. Explored Hyperparameter Configurations for Model Optimization

Model	Hyperparameter	Explored Values
LR	max_iter	100, 1000, 2000
	C (Regularization)	0.01, 1, 10
RF	n_estimators	50, 100, 200
	max_depth	10, None, 50
	min_samples_split	2, 5, 10
GBst	n_estimators	50, 100, 200
	learning_rate	0.01, 0.1, 0.2
	max_depth	3, 5, 7
SVM	kernel	linear, rbf, poly
	C (Regularization)	0.1, 1, 10
	gamma	scale, auto, 0.01
XGBst	n_estimators	50, 100, 200
	learning_rate	0.01, 0.1, 0.2
	max_depth	3, 6, 9
	subsample	0.5, 1.0

IV. MODEL EVALAUTION METRICS

The performance of these classification models was analyzed with a variety of parameters that provide holistic understanding of the capability of these models to discriminate between different classes of aerial images.

Accuracy(Refer Eqn.1) gives a measure of number of instances correctly classified:

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

In the above expression, TP and TN indicate the number of true positives and true negatives while FP and FN are the false positives and false negatives counts respectively.

Precision(Refer Eqn.2) is the percentage of positive predictions that are correct:

$$Precision = \frac{TP}{(TP+FP)} \qquad (2)$$

High precision means high confidence on real positive instances and low false positives which is highly desired in the domain where classifying specific types of instances wrongly is detrimental.

Recall(Refer Eqn.3) is the percentage of positive instances that were predicted as positive by the model:

$$Recall = \frac{TP}{(TP+FN)} \qquad (3)$$

Lastly, recall is important in aerial image classification because it maintains that if critical land-use classes are overlooked, it will create bias and ultimately paint inaccurate picture.

The F1-score(Refer Eqn.4)is one such metric that considers both precision and recall, thus being a neutral parameter to judge the model's performance:

$$F1_{Score} = 2 * \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

This is more informative than evaluating only precision or recall, considering the fact that many real-world datasets contain classes with different number of instances.

The formula for Cohen's Kappa(Refer Eqn.5)seeks to evaluate the degree of correlation between predictions and actual labels whilst also allowing for chance agreement:

$$k = \frac{\text{observed accuracy} - \text{expected accuracy}}{1 - \text{expected accuracy}} \quad (5)$$

This formula entails observed accuracy as the actual model accuracy while expected accuracy is the agreement expected by chance. For the performance of the model especially with imbalanced sets, kappa gives a more reliable evaluation of the model.

By evaluating all elements of the confusion matrix, especially the four, MCC (Refer Eqn.6) can also be derived from a much simpler piece of information.

$$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

Such alterations make monitoring model performance particularly simple since the description is MCC of +1 is perfect prediction, while 0 suggests random prediction.

In its most basic form AUC-ROC (Refer Eqn.7) captures the ability of a machines model to differentiate between classes. It analyzes the TPR and FPR for diverse classification processes as a function of a specific classification mechanism.

$$AUC\_ROC = \int_0^1 TPR(FPR)\,d(FPR) \quad (7)$$

Where:

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN}$$

Notice that AUC-ROC is particularly great for tasks that have to do with comparison of model performance with different types of operating thresholds.

The performance of all the models was analyzed in a more rigorous with this particular metric suite and it's ensured that the metrics generated are robust and reliable. Such evaluation gives multi-dimensional understanding for model behavior towards different, preserving their relevancy for the classification of aerial images.

V. RESULTS AND IMPLEMENTATIONS

The present study discusses the relative performance of the selected classification models working in a competitive atmosphere on the dataset with layer-wise performance criteria.

A. Overall Model Performance

Given the dissimilar accuracies, models also performed with different performance metrics. SVM yielded the highest accuracy of 85.00%, immediately followed by 80.71% of LR. Although Logistic Regression performed better in the measures of Precision, Recall, F1, and AUC-ROC, its accuracy score was slightly lower than that of SVM. The overall model performance across different metrics is illustrated in Table.2.

SVM had remarkable results with 85.46% Precision and 85.00% Recall. This gives heart to the model's balanced skill between Precision and Recall, with its F1-score at 84.89%. Internally, the AUC-ROC value is stratospheric at 0.993, meaning the model does incredibly well separating the classes.

Meanwhile, the LR model showed Precision (81.11%), Recall (80.71%), and F1 (80.61%) with a performance close to the SVM model. Its Cohen Value Kappa (0.7975) and MCC (0.7978) were indicative of a strong agreement between predicted and the true class labels. Metric wise comparison of all the models are shown in Fig.4.

RF and XGBst posted closely similar overall accuracy scores of 74.05%; now the big difference here is that RF edged out (73.82%) on Precision and Recall at 74.05%. Although always dependable performance-wise, RF and XGBst were not able to generalize well across all classes, unlike SVM and LR.

GBst had the lowest accuracy of 68.57%, but although it showed reasonable Precision, Recall, and F1 for particular classes, all were nullified by those opposite tendencies in others, leading to mixed result performances. The confusion matrix for LR, GBst, RF, SVM, XGBst, are given in Fig. 5-9. And the Roc-curve of the comparative models are given Fig. 10.

Table.2. Overall Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Cohen Kappa	MCC
Logistic Regression	0.807143	0.811134	0.807143	0.806110	0.991642	0.7975	0.797818
Random Forest	0.740476	0.738186	0.740476	0.735046	0.977258	0.7275	0.727946
Gradient Boosting	0.685714	0.708677	0.685714	0.688619	0.968370	0.6700	0.670947
SVM	0.850000	0.854606	0.850000	0.848904	0.993463	0.8425	0.842821
XGBoost	0.740476	0.743204	0.740476	0.738354	0.982736	0.7275	0.727868

Figure 4. Comparison of Model performance across Metrics

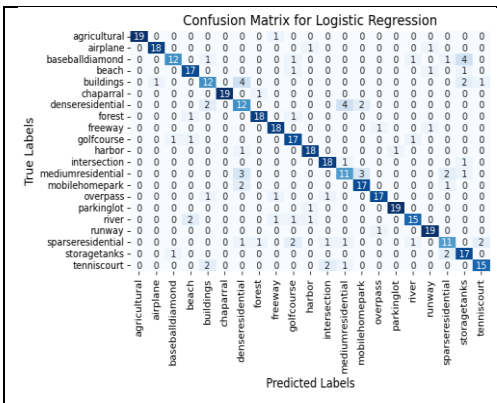

Figure 5. Confusion Matrix for LR

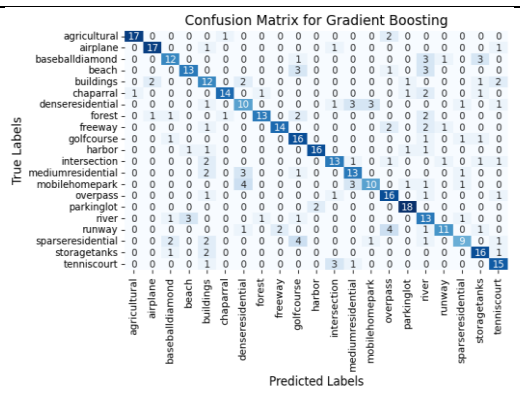

Figure 6. Confusion Matrix for GBst

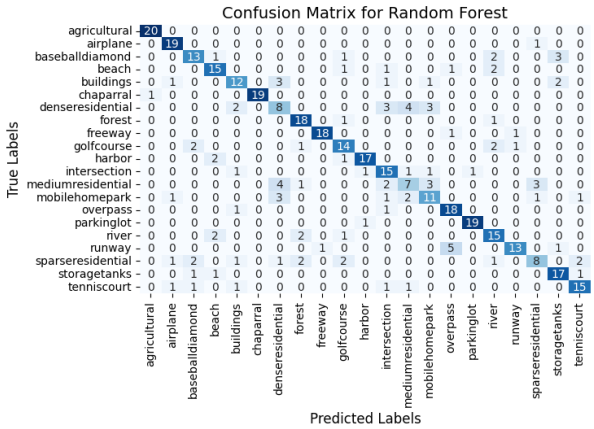

Figure 7. Confusion Matrix for RF

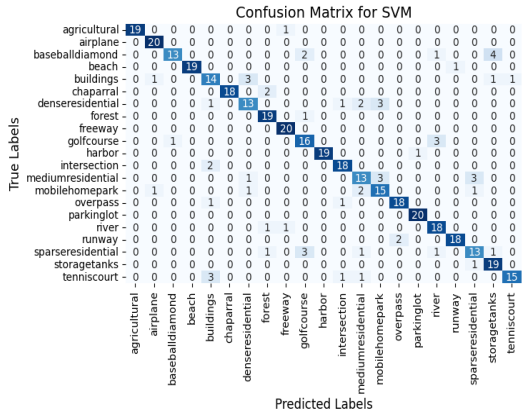

Figure 8. Confusion Matrix for RF

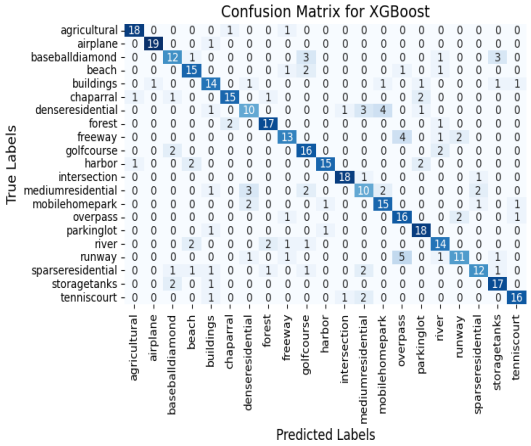

Figure 9. Confusion Matrix for XGBst

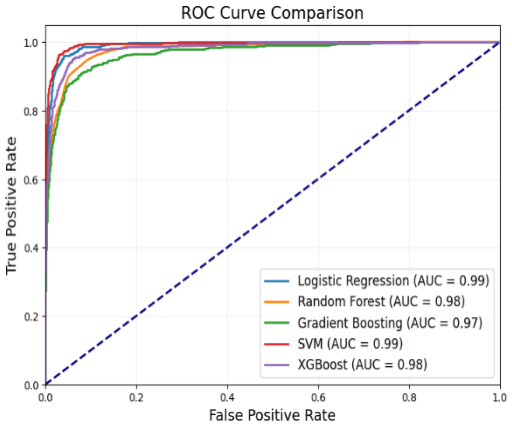

Figure 10. RoC Curve comparison of all the comparative models

B. Class_specific preformance

A close monitoring of class-specific performance metrics exposes some of the model behaviors across the class-wise categories as shown in Fig. 11. In this study, LR seemed to produce very good results in the classes of agriculture, airplane, and chaparral, indicating a 100% precision and very high recalls within these categories. Such results give us an indication that the model is likely to focus in some particular land use types due to clear separability of the classes in the feature space.

In contrast, for RF, high precision and recalls in the agricultural, airplane, and chaparral classes could be seen as indications that the model efficiently identifies these land-use patterns, although problems arise in terms of poor performance in denseresidential and mediumresidential categories where it showed low precision and recall. This reflects general statistics in this model that are lower than those in SVM and LR models.

GBst, on the other hand, shows relatively weaker performance, with good precision 94.4% and 85% recall respectable in agricultural and airplane classes but obviously very poorly with the assemblages of buildings and sparseresidential, where such F1 scores are remarkably lower. It is shown that GBst excels at specific instances due to this limitation of generalization occurring for class imbalance as well as feature heterogeneity.

Figure 11: Class- wise performance metrics of comparative models

VI. DISCUSSION

The study reflects that SVM and LR models are better performing models in classification tasks of land use. Particularly, SVM's remarkable performance can be linked back to its proficiency in dealing with high dimensional data. Despite everything, which is a good way to handle complex class boundary datasets like UCMLU.

Despite weaknesses known to be present in some limited numbers, the RF and XGBst methods, well-proven, appeared to demonstrate weaknesses in modeling complex class distributions. The ability of these algorithms to perform well in the presence of any class may be illustrative of the great difficulty they have submitting intricate patterns found in aerial imagery data.

GBst's performance was less compared to the other ensembles, possibly due to overfitting and hyperparameters' sensitivity. Further work could therefore be directed to enhance its performance on difficult to classify instances by cross-dwelling upon optimization of such parameters as the learning rate, depth of the trees, etc.

To conclude, though RF, XGBst, and GBst ensemble methods are powerful means with various kinds of classification tasks, straightforward models, such as SVM and LR, might be better suited for some tasks, especially those in which the classes are well distinguished and balanced. The insights provided by the study could prove to be quite important for the future research that is needed in helping to guide the importance of the appropriate model selection for a given land-use classification task. Further studies, for instance, parameter tuning or feature selection, both of which would ideally enhance model performance, could follow to refine our results in aerial image classification.

VII. GRAD-CAM VISUALIZATION AND INTERPRETATION

The Grad-CAM is an important technique for XAI used in understanding the machine learning model decision-making process. It highlights the regions on the image that are most informative with regard to its prediction by the model. For the current study, Grad-CAM was applied to a pre-trained model ResNet50 in order to visualize the focus of a model on specific features in aerial images from the dataset. A custom Grad model was created to simultaneously output the model's final predictions and feature maps from the last convolutional layer (conv5_block3_out).

Sample images in each class were resized to 224x224 pixels pre-processed according standard practices and fed into the Grad model for prediction. The corresponding feature maps were extracted for Grad-CAM using the corresponding features.

Gradients of predicted class score for the feature mapping with respect to the class score were computed using Tenskeleton' s gradient tape. This gradient captures the importance of each feature map element in influencing the prediction. A weighted heatmap was constructed by multiplying the feature map by the gradient values and performing global average pooling. This heatmap highlights the image regions that contributed most significantly to the model's classification decision.

The Grad-CAM heatmap was overlayed on the original image to visually represent the regions of the image the model focused on for its prediction. In land-use classification, heatmaps for buildings were strongly activated the rooftop regions suggesting the model relied on this feature for its classification.

The visual analysis revealed differences in the way the model focused on land-use classes with well-defined features versus more ambiguous classes. For example buildings showed a more localized heatmap while chaparral exhibited wider activations across the image. Fig. 11 emphasizes the use of Grad-CAM for CNN-based feature interpretation in aerial image analysis.

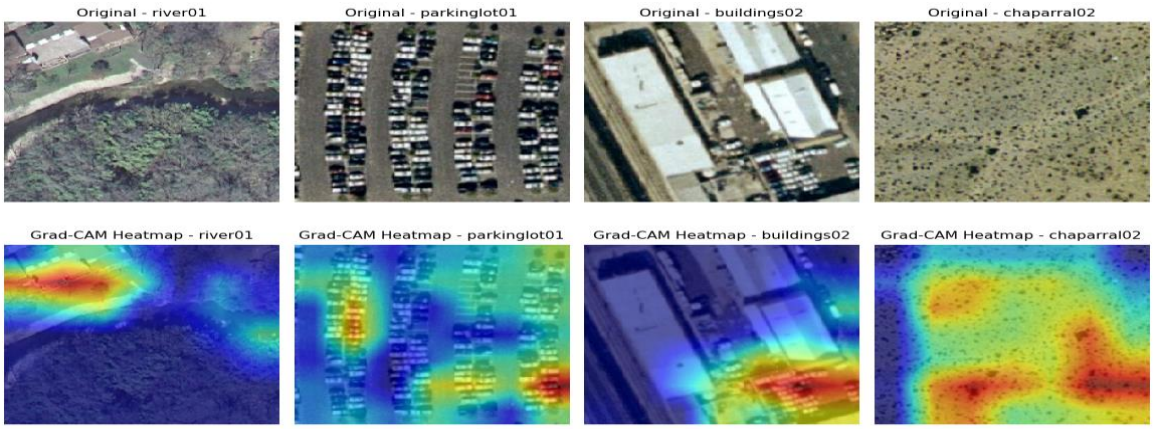

Figure 11: Comparison of Original Aerial Images and their Corresponding Grad-CAM Heatmaps for Different Terrain Types

VIII. CONCLUSION

This study presented a complete analysis of various machine learning classifiers applied to the dataset for aerial image classification using a database of 2100 images. Through an optimized data preprocessing pipeline incorporating advanced data augmentation techniques the models were trained and evaluated to assess their performance across several metrics including accuracy, precision recall F1-score AUC-ROC Cohen's Kappa and MCC. Our results highlighted the superior performance of the SVM model which achieved the highest accuracy (85%) and AUC-ROC score (0.99), demonstrating its robustness and ability to generalize well across the dataset. LR, RF and GBst followed with notable performance in terms of precision and recall but slightly lower overall classification accuracy than SVM.

Furthermore the incorporation of Grad-CAM enabled us to visually interpret the decision making processes in the models. By overlaying class-specific heatmaps on the images, the key regions focused by the models focused upon such as rooftops, agricultural areas and roads. These visualizations provided insight into the spatial reasoning of the models, demonstrating their ability to focus on semantically meaningful areas and in particular in the case of deep learning models.

IX. RESEARCH DIRECTIONS

While this study discerned model performance and interpretability, there are several scope for further exploration that could better enhance both the accuracy of the models in aerial image classification tasks:

- Exploring ensemble methods such as stacking voting bagging and stacked can leverage individual classifiers' strengths to potentially improve overall classification performance.

- Investigating the performance of deep learning models including CNNs and Vision Transformers, on this dataset can unlock the potential for more sophisticated feature extraction and representation
- Fine-tuning prerained models like ResNet and optimizing hyperparameters can significantly enhance model performance and generalization.
- Beyond Grad-CAM exploring explainability methods like SHAP and LIME can provide deeper insights into the model decision-making, building trust and ensuring the reliability of AI systems in critical applications.
- Leveraging transfer learning on models prepre-trained on similar datasets can improve performance especially with limited data.
- Combining aerial imagery with other data sources, such as GIS data and environmental sensor readings can enhance classification accuracy and enable more thorough analyses.
- Incorporating temporal features of aerial imagery such as seasonal changes and urban growth can provide valuable insights into land use dynamics and support applications like land management and urban development planning.

By pursuing these research directions, further research can be explored to advance the development of precision and interpretable AI systems for remote sensing applications.

REFRENCES

[1] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," Remote Sens., vol. 10, no. 7, p. 1119, 2018.

[2] Mather, P. M., & Koch, M. (2011). Computer Processing of Remotely-Sensed Images: An Introduction. Wiley, Online ISBN:9780470666517, DOI:10.1002/9780470666517

[3] Jensen, J.R. (2005) Introductory Digital Image Processing: A Remote Sensing Perspective. 3rd Edition, Prentice Hall, Upper Saddle River, 505-512.

[4] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[5] Azzari, G., & Lobell, D. (2017). Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. Remote Sensing of Environment, 202, 64-74.

[6] M. A. Shafaey, M. A. Salem, H. M. Ebied, M. N. Al-Berry, and M. F. Tolba, "Deep learning for satellite image classification," in Proc. Int. Conf. Adv. Intell. Syst. Inform., vol. 845, 2019, pp. 383–391.

[7] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 4, pp. 549–553, Apr. 2017.

[8] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-D deep learning approach for remote sensing image classification," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 4420–4434, Aug. 2018.

[9] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active transfer learning network: A unified deep joint spectral–spatial feature learning model for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 3, pp. 1741–1754, Mar. 2019.

[10] Abdi, H. and Williams, L.J. (2010) Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433-459. http://dx.doi.org/10.1002/wics.101

[11] Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical Automated Data Augmentation with a Reduced Search Space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 702–703.

[12] Norman R. Draper, Harry Smith, Applied Regression Analysis, 9 April 1998, Online ISBN:9781118625590 |DOI:10.1002/9781118625590, Wiley Series in Probability and Statistics.

[13] Breiman, L. (2001) Random Forests. Machine Learning, 45, 5-32. http://dx.doi.org/10.1023/A:1010933404324

[14] Jerome H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis, Volume 38, Issue 4, 2002, Pages 367-378, ISSN 0167-9473, https://doi.org/10.1016/S0167-9473(01)00065-2.

[15] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. Machine Learning, 20, 273-297. http://dx.doi.org/10.1007/BF00994018

[16] Friedman, Jerome. (2000). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. 29. 10.1214/aos/1013203451.
[17] Pal, M. (2005). Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217–222. https://doi.org/10.1080/01431160412331269698
[18] Giles M. Foody, Status of land cover classification accuracy assessment, Remote Sensing of Environment, Volume 80, Issue 1, 2002, Pages 185-201, ISSN 0034-4257, https://doi.org/10.1016/S0034-4257(01)00295-4.
[19] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.
[20] Jayanthi, S., Geraldine, J.M. (2019). Development and Investigation of Data Stream Classifier Using Lagrangian Interpolation Method. In: Saini, H., Sayal, R., Govardhan, A., Buyya, R. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 74. Springer, Singapore. https://doi.org/10.1007/978-981-13-7082-3_38
[21] Dr. S Jayanthi, "Investigation of Land - Use and Land - Cover Classification with synergization of Resnet50, PCA, and Machine Learning Classifiers" Letters in High Energy Physics, ISSN: 2632-2714, Volume 2024
[22] Alem, A., & Kumar, S. (2022). Deep Learning Models Performance Evaluations for Remote Sensed Image Classification. IEEE Access, 10, DOI:10.1109/ACCESS.2022.3215264.
[23] Hamza et al., "An Integrated Parallel Inner Deep Learning Models Information Fusion With Bayesian Optimization for Land Scene Classification in Satellite Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 9888-9903, 2023, doi: 10.1109/JSTARS.2023.3324494.
[24] Papoutsis, N. I. Bountos, A. Zavras, D. Michail, and C. Tryfonopoulos, "Benchmarking and scaling of deep learning models for land cover image classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 195, pp. 250-268, 2023. https://doi.org/10.1016/j.isprsjprs.2022.11.012
[25] Z. Ma, M. Xia, H. Lin, M. Qian, and Y. Zhang, "FENet: Feature enhancement network for land cover classification," International Journal of Remote Sensing, vol. 44, pp. 1702-1725, 2023, https://doi.org/10.1080/01431161.2023.2190471
[26] Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, Volume 202, 2017, Pages 18-27, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2017.06.031.
[27] M. Aljebreen, H. A. Mengash, M. Alamgeer, S. S. Alotaibi, A. S. Salama and M. A. Hamza, "Land Use and Land Cover Classification Using River Formation Dynamics Algorithm With Deep Learning on Remote Sensing Images," in IEEE Access, vol. 12, pp. 11147-11156, 2024, doi: 10.1109/ACCESS.2023.3349285
[28] Huang, Q., Xia, H., & Zhang, Z. (2023). Clustering Analysis of Integrated Rural Land for Three Industries Using Deep Learning and Artificial Intelligence. IEEE Access, 11, 110530-110543, DOI: 10.1109/ACCESS.2023.3321894
[29] Dataset: http://weegee.vision.ucmerced.edu/datasets/landuse.html