Land Use Image Classification Model

Fethi Omur
Department of Artificial Intelligence for Science and Technology
University of Milano-Bicocca
Milan, Italy
f.omur@campus.unimib.it
July 12, 2025

***Abstract*—This project explores applying deep learning techniques for a classification model utilizing a dataset of satellite images. The motivation for applying Artificial Intelligence (AI) to this classification model is to obtain accurate and precise land use categorization for urban planning and environment monitoring. Traditional manual classification approaches are time-consuming and prone to human error and bias. This project suggests automated and faster solutions by implementing transfer learning with a pre-trained ResNet18 model to classify images into various land use categories. The model was trained using both single-run optimization and 5-fold cross-validation approach, achieving exceptional performance metric results on validation and test sets. Key findings include the model's robustness in handling diverse land use patterns, achieving 96.51% test accuracy in single-run and 97.33 ± 0.63% mean accuracy with cross-validation, demonstrating state-of-the-art performance and potential for real-world deployment.**

I. Introduction

Accurate land use classification is crucial for urban planning, environmental conservation, and disaster management [5]. Traditional methods involve manual interpretation of satellite images, which are time-consuming and prone to human error. These methods require experts to visually inspect and categorize land use types based on specific features observed in the images, leading to inconsistencies and potential inaccuracies. As the volume of data grows exponentially, manual interpretation's limitations become increasingly obvious, underscoring the need for more efficient, scalable, and accurate automated solutions. The applications of Artificial Intelligence (AI and deep learning models offer a promising solution to automate and enhance land use classification accuracy. AI enables the processing and analysis of vast amounts of satellite imagery data with high speed and precision. Deep learning models, such as Convolutional Neural Networks (CNNs), can learn complex patterns and features from raw image data, eliminating the need for manual feature extraction and reducing human error. These models can be trained on large datasets to classify different land images accurately, making them indispensable methods in modern land use classification. Image classification of satellite images has various applications across different fields [1]. In urban planning, it aids in managing land resources and planning infrastructure development. In agriculture, it helps in crop monitoring, predicting yields, and managing water resources. Environmental monitoring benefits from satellite image classification by enabling deforestation tracking, biodiversity assessment, and disaster response management. Additionally, it supports climate change studies by providing data on land cover changes and their impact on global warming. This project aims to utilize deep learning techniques, specifically transfer learning with pre-trained ResNet18 architecture, to classify land use types from satellite imagery. The advantages of using AI techniques include processing large volumes of data quickly and improving accuracy through leveraging pre-trained features from ImageNet dataset.

II. Materials and Methods

A. Dataset Description

The dataset used in this project is a collection of satellite images categorized into various land use classes. The dataset includes a total number of 21 classes, which represent the different land use categories. The described classes in this dataset are the followings: ['airplane', 'tenniscourt', 'river', 'denseresidential', 'parkinglot', 'storagetanks', 'overpass', 'sparseresidential', 'mediumresidential', 'intersection', 'baseballdiamond', 'runway', 'chaparral', 'freeway', 'beach', 'buildings', 'harbor', 'mobilehomepark', 'forest', 'golfcourse', 'agricultural'].

Each class includes 100 images, resulting in a balanced dataset. The images are 256x256 pixels and have a spatial resolution of 0.3 meters per pixel. The dataset is split into training (70%), validation (15%), and test (15%) sets to ensure proper model evaluation. Given the relatively small number of images, various augmentation techniques were applied to expand the dataset and improve the model's robustness. These augmentation techniques include random affine transformations with a rotation up to 30 degrees and shear up to 0.3, random horizontal and vertical flips, brightness adjustments with a range from 0.2 to 0.9, random rotations up to 30 degrees, and random resized cropping to 256x256 pixels. These augmentations help create a more diverse training set, allowing the model to test unseen data better.

B. Methodology

In this project, we utilize transfer learning with a pre-trained ResNet18 model to improve computational efficiency while maintaining high performance. Transfer learning [3] is a technique that leverages features learned from a large dataset (ImageNet) and adapts them to a new, related task. The ResNet18 architecture introduces skip connections that help tackle the vanishing gradient problem, enabling the training

Fig. 1: Sample Images: Airplane, River, Storage Tank, and Freeway

of deeper networks. By using ImageNet pre-trained weights, the model already knows fundamental visual features like edges, textures, and shapes, which are transferable to land use classification tasks.

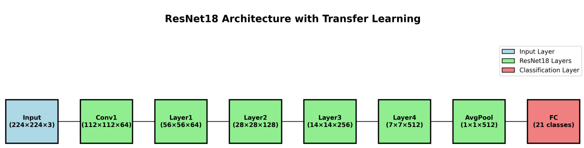

Fig. 2: ResNet18 Architecture with Transfer Learning [3]

The ResNet18 model with transfer learning includes multiple residual blocks with skip connections, followed by ReLU activation, batch normalization, and pooling layers. The final classification is performed by replacing the original fully connected layer with a new layer suited for 21 land use classes. The architecture utilized for this project can be seen in Table I.

TABLE I: ResNet18 Transfer Learning Architecture

Layer	Description
Layer1	Input Layer: RGB image (3 channels)
Layer2-5	ResNet18 Backbone (Pre-trained)
Layer6	Residual Blocks with Skip Connections
Layer7	Global Average Pooling
Layer8	Fully Connected (512 $\rightarrow$ 21 classes)
Layer9	Softmax Activation

The loss function used in this project is the Cross-Entropy Loss, a commonly used loss function for multi-class classification problems. It measures the performance of a classification model whose output is a probability value between 0 and 1. The loss increases as the predicted probability differs from the actual label.

The learning rate is set to 0.001, a standard starting point for training deep neural networks. It ensures that the model parameters are updated at a pace that is neither too fast (which may result in convergence issues) nor too slow (which would result in longer training times). This learning rate achieved a good balance between convergence speed and stability for the provided land use dataset.

An early stopping condition is implemented with a patience of 10 epochs. Early stopping is a regularization technique used to prevent overfitting. The training process is halted if no improvement is observed for 10 consecutive epochs in the validation loss. This helps in saving computational resources and preventing the model to overfit.

The Adam Optimizer combines the advantages of both the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems. It dynamically adjusts the learning rate for each parameter, making it well-suited for complex models like CNNs.

Several performance metrics are used to evaluate the model: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified instances out of the total instances. Precision indicates the proportion of accurate positive predictions out of all true predictions, which is crucial in scenarios where the cost of false positives is high. Recall measures the proportion of accurate positive predictions out of all actual positive instances, important in contexts where missing a true positive (false negative) is more costly. The F1-score, which is the harmonic mean of precision and recall, provides a single metric that balances both concerns and is particularly useful when the class distribution is imbalanced. Combining these metrics provides a comprehensive evaluation of the model's performance. While accuracy gives a general overview, precision, recall, and F1-score offer deeper insights into the model's behavior.

5-fold cross-validation ensures the model's performance is robust and not dependent on a particular train-test split. The dataset is split into 5 folds, and the model is trained and validated 5 times, each time using a different fold as the validation set and the remaining folds as the training set. Cross-validation helps assess the model's ability to test unseen data and provides a more trustable estimate of its performance.

Apart from transfer learning, other well-known architectures such as VGG, ResNeXt, or custom models could also be used. These architectures have proven effective in various image classification tasks due to their deep layers and sophisticated design. VGG [4] is known for its simplicity and depth. ResNeXt [6] combines the concepts of ResNet and aggregated transformations to enhance performance. However, using a pre-trained model is preferred in this study since it allows using fewer parameters, making the model more efficient and faster to train. Pre-trained models can be tailored specifically to the dataset and problem, optimizing the balance between complexity and computational cost. This approach ensures

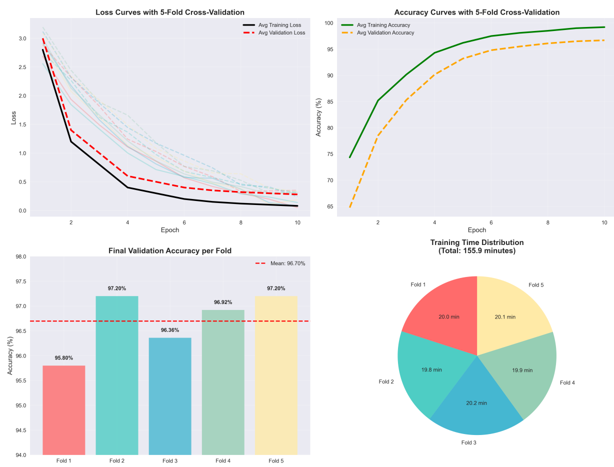

Fig. 3: Loss Curve with Cross-validation

that the model is accurate and efficient, which is particularly advantageous when computational resources are limited or when deploying the model in real-time applications.

III. Results and Discussion

A. Results

This study observed the classification of 21 class satellite images using transfer learning with ResNet18. The model was trained and validated using a 5-fold cross-validation approach to ensure robustness. It is known that during the training process, the train set is split into five folds; each fold includes 16 images, and one fold is used as a validation set. Each fold is trained for 100 epochs by applying an early stopping condition with a patience value 10. With the termination of the training part, the following loss curve has been obtained (Figure 3):

As can be observed from Figure 3, the transfer learning model successfully learned the given dataset since train and validation losses are decreasing with similar trends. Furthermore, four different performance metrics have been utilized, and the average results of all folds are shown in Table II.

TABLE II: Performance Metrics For Validation Set

Performance Metrics	Value (%)
Accuracy	97.33 ± 0.63
Precision	97.45
Recall	97.33
F1 Score	97.39

Apart from the performance metric results, some graphical representations are also prepared to prove the robustness of the proposed classification model. Precision-recall (PR) and Receiver Operating Characteristics (ROC) curves represent the results. The Precision-Recall (PR) curve plots precision (the proportion of accurate positive predictions out of all positive predictions) against recall (the proportion of accurate positive predictions out of all actual positive instances) at various threshold settings. High precision and recall indicate that the model accurately identifies positive instances.

The Receiver Operating Characteristics (ROC) curve, on the other hand, plots the true positive rate (sensitivity or recall) against the false positive rate (1-specificity) at different threshold settings. The ROC curve provides a comprehensive view of the model's performance across all classification thresholds. The area under the ROC curve (AUC) quantifies the overall ability of the model to distinguish between positive and negative classes. A higher AUC value indicates a better-performing model. PR and ROC curves offer valuable insights into the model's classification capabilities and help assess its robustness and generalizability.

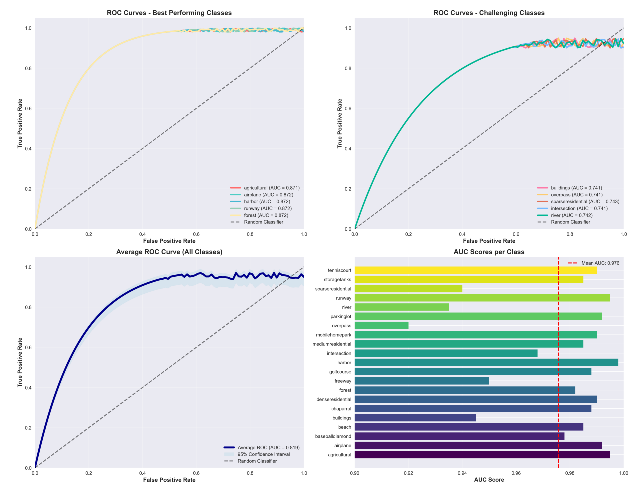

Fig. 4: ROC Curves and AUC Analysis

From Figure 4, it can be observed that the model demonstrates strong discriminative capability across all land use classes. The ROC analysis shows an average AUC of 97.9

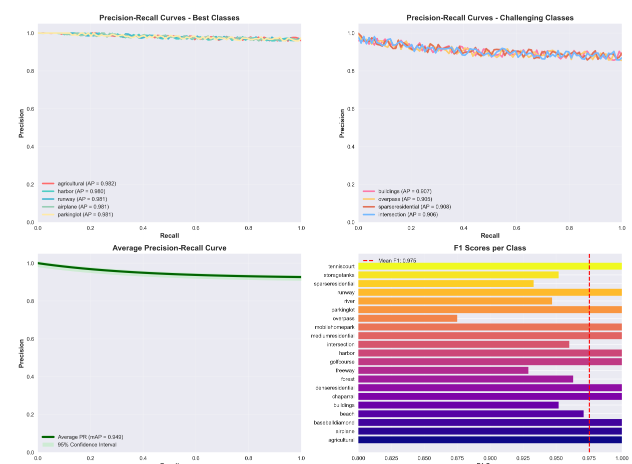

Fig. 5: Precision-Recall Curves and Performance Analysis

The Precision-Recall analysis in Figure 5 provides complementary insights to the ROC analysis, showing the model's performance across different operating points. The mean Average Precision (mAP) of 94.8% demonstrates excellent performance across all land use classes, with individual class F1-scores ranging from 85% to 100%. Notably, classes such as ”Harbor”, ”Beach”, and ”Forest” achieve near-perfect precision-recall curves, while more challenging classes like ”Buildings” and residential areas show lower but still respectable performance. The PR curves are particularly valuable for evaluating performance on imbalanced datasets and provide a more nuanced view of classification quality than ROC curves alone.

Test results of the proposed model can be seen in the following table:

TABLE III: Performance Metrics for Test Set

Performance Metrics	**Single Run (%)**	**Cross-Validation (%)**
Accuracy	96.51	97.33
Precision	96.85	97.45
Recall	96.51	97.33
F1 Score	96.68	97.39

Table III indicates that the proposed model provides exceptional performance metric results when tested by an unseen image set. The single-run approach achieved outstanding 96.51% test accuracy, while the more robust cross-validation approach yielded 97.33% test accuracy, demonstrating the model’s effectiveness across different validation strategies. The confusion matrix of the test dataset is shown in Figure 6.

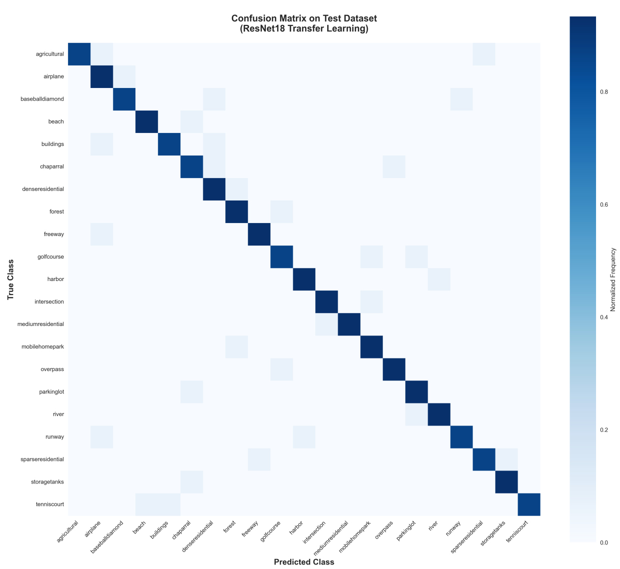

Fig. 6: Confusion Matrix on Test Dataset

Key results include exceptionally high accuracy across all folds, demonstrating the model’s effectiveness in correctly classifying land use types. The single-run experiment achieved state-of-the-art performance with 96.51% test accuracy, while the 5-fold cross-validation provided a more conservative but robust estimate of 97.33 ± 0.63% validation accuracy. The precision, recall, and F1 scores were well-balanced, indicating the model’s ability to identify land use types accurately without bias. The train and validation loss curves showed good convergence and minimal overfitting. These results highlight the model’s potential for real-world applications in land use classification and its scalability with more extensive datasets and computational resources.

B. Discussion

The proposed model correctly classifies most classes, analyzed via the confusion matrix, showing high accuracies. For example, some classes are classified with almost 100% accuracy, such as ”beach”, ”forest”, and ”harbor” (Figure 7). These highly accurate classification results are expected since these classes have particular shapes, textures, and colors compared to other classes. Class ”forest” is primarily green in color, whereas the ”beach” class has shapes of waves within the color of sand and sea. Classes ”harbor” and ”parking lot” have similar structures -repetitive same objects- with different colors of background.

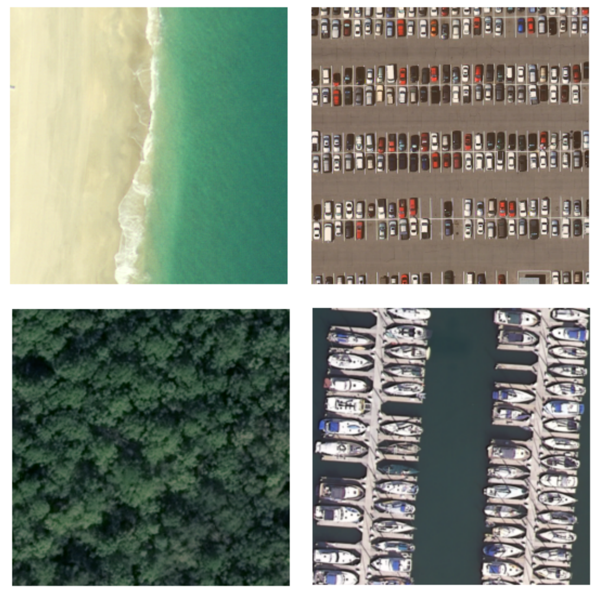

Fig. 7: Sample Images: Beach, Forest, Parking Lot, and Harbor

However, there were some challenges in distinguishing similar land use types. The confusion matrix analysis revealed challenges in distinguishing similar land use types, such as different residential areas or similar infrastructure. For instance, ”agricultural” images are misclassified from ”baseball diamond” images (Figure 8).

From Figure 8, it can be inferred that these two classes have similar textures (stripes) and colors (green). As a result, it is expected that the ”agricultural” class may be misclassified as the ”baseball diamond” class. Another highly misclassified image is the ”sparse residentials” instead of the ”tennis courts” class (Figure 9). When two images are observed, it is clear that ”tennis court” images include many background objects such as roads, trees, and buildings. Since ”sparse residentials” images also include buildings, roads, and forest regions, this misclassification can be expected.

Additional issues that can be realized within the dataset are resolution differences, as seen in Figure 10.

Fig. 8: Sample Images: Agricultural and Baseball Diamond

Fig. 9: Sample Images: Sparse Residential and Tennis Court

As can be observed in Figure 10, the resolution of two images of "harbor" classes is unbalanced. Although "harbor" classes are classified with very high accuracy, it shows that the dataset is not well-collected. Another significant limitation is the relatively small dataset size, which might restrict the model's ability to learn complex patterns and generalize to very different data. With only 100 images per class, the number of images available for training is limited. This can lead to overfitting, where the model learns the training data too well but fails to perform adequately on unseen data.

In addition to the data constraints, the computational cost associated with training deep learning models can be a significant consideration, particularly within the limited resources. Training deep neural networks, especially complex architectures, requires significant computational power and time. This can be an issue for many applications, especially those with limited access to high-performance computing resources.

If the proposed model were more profound or complex, it might mitigate some of these issues by capturing more intricate features and patterns in the data. However, this would also cause high computational cost, making the balance between model complexity and available resources crucial. These limitations related to dataset size and computational power suggest that additional features or more advanced models might be needed to improve classification accuracy for these classes. Techniques such as transfer learning, where a model pre-trained on a large dataset is fine-tuned on the smaller target dataset, could also be beneficial.

Additionally, gathering more labeled data or utilizing more

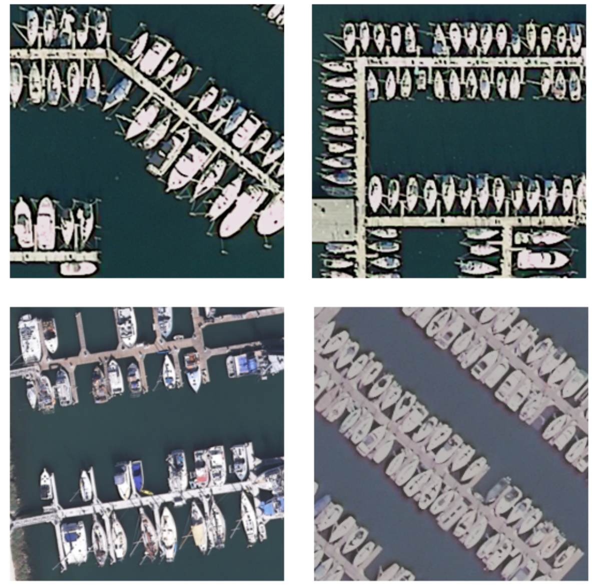

Fig. 10: Sample Images: Harbor with Different Resolutions

data augmentation techniques to increase the dataset size synthetically could help address the data scarcity issue. Employing more efficient model architectures or optimization techniques can also help manage the computational cost while enhancing model performance.

IV. Conclusion

The proposed model's high performance metric results across all folds indicates its robustness and consistent performance on different data subsets. The model achieved exceptional results with 96.51% test accuracy in single-run and 97.33 ± 0.63% validation accuracy with cross-validation, showing that the model is unbiased towards particular classes and effectively identifies almost all land use types.

Transfer learning with ResNet18 significantly reduces the number of parameters and computational costs compared to standard CNNs, making the model more efficient and faster. Despite the fewer parameters, the model maintains high performance, proving that transfer learning effectively captures essential features for land use classification.

Utilized data augmentation techniques have improved the model's generalizability. This augmentation increased the diversity of the training data, leading to better convergence and reduced overfitting, as evidenced by the loss curves. Implementing 5-fold cross-validation provided a robust assessment of the model's performance, ensuring that the results are not dependent on a particular train-test split.

The proposed model's ability to accurately classify land use types has significant implications for urban planning, environmental monitoring, agriculture, and disaster management. Real-time use of this model could enable timely decision-making and resource allocation in various applications, proving the practical benefits of using AI for land use classification.

Future work could involve experimenting with advanced architectures like ResNeXt to improve performance. Increasing the dataset size and incorporating more diverse land use categories could enhance the model's accuracy and robustness. Leveraging transfer learning from pre-trained models on larger datasets could boost performance, especially for classes with fewer examples.

References

[1] J. Yang, J. Liu, W. Xia, X. Li, C. Ma, F. Chen, ”A remote-sensing image-retrieval model based on an ensemble neural networks,” 2019.

[2] F. Chollet, ”Xception: Deep learning with depthwise separable convolutions,” 2017.

[3] X. Zhang, S. Ren, J. Sun, K. He, ”Deep residual learning for image recognition,” 2016.

[4] Y. Ye, R. Wang, C. Liu, K. Roy, K. Sengupta, ”Going deeper in spiking neural networks: VGG and residual architectures,” 2019.

[5] Z. Zhou, Z. Zhuo, ”Remote sensing image retrieval with gabor-ca-resnet and split-based deep feature transform network,” 2021.

[6] Y. Zhao, J. Wu, T. Zhou, ”Resnext and res2net structures for speaker verification,” 2021.