Land Use and Land Cover Classification of Satellite Images Using Convolutional Neural Network

Jann Rovic Cueto[1,2]

[1] Department of Earth and Space Science, Rizal Technological University, Mandaluyong 1550, Philippines; 2022-104684@rtu.edu.ph

[2] Department of Creative and Product Design, Asia University, Taichung City 41354, Taiwan; 112710342@live.asia.edu.tw

Abstract:

This study investigates the application of deep convolutional neural networks (DCNNs) for precise land-use and land-cover (LULC) classification using high-resolution satellite imagery. Three pre-trained models (Inception, EfficientNetB1, and ResNet50) were examined and optimized for LULC classification. Inception demonstrated the highest success rate at 98%, excelling in vegetation and impermeable surface classification but facing challenges with water recognition. EfficientNetB1 shared water-related challenges with Inception but exhibited balanced performance across classes. Interestingly, ResNet50 resulted in the best for impervious surface classification and best for water classification. However, the study revealed potential issues of overfitting and class-specific complexities, which suggest for further research. Despite high overall accuracy, Inception and EfficientNetB1 faced challenges in water classification, indicating a need for specific solutions. Validation loss curves suggested overfitting in some models, prompting further study on data augmentation and hyperparameter tuning. In conclusion, this study highlights areas for improvement in AI-based LULC classification, suggesting possibilities for further research into advanced transfer learning algorithms, mitigate overfitting, and address class-specific challenges. Future studies can also explore alternative CNN models for classification.

Keyword: Land Use and Land Cover Classification, Machine Learning, Convolutional Neural Network, Transfer Learning.

I. Introduction

Accurate understanding of land cover is crucial for environmental monitoring, resource management, and urban planning. While traditional methods have served us well, their limitations call for advancements in classification techniques (Raoof Naushad et al., 2021). High-resolution satellite imagery became readily available, allowing for the development of efficient image classification algorithms (Piramanayagam et al., 2016) and (Liu et al., 2017). Machine learning, particularly supervised models, demonstrated outstanding performance in land use and land cover (LULC) classification tasks, paving the way for automated and accurate land cover mapping. Despite these advancements, a significant limitation remained. Machine learning models trained for different land cover categories often produced inconsistent results and failed to capture the complex relationships between various types of images (Chen et al., 2022). To overcome this, a novel approach was implemented: we investigate the application of deep learning classification algorithm model utilizing a larger and more diverse training dataset. This approach aimed to enhance classification consistency, improve accuracy, and capture complex relationships, ultimately leading to more robust and reliable results. Research studies supported the effectiveness of this method, demonstrating significantly higher classification accuracy compared to ML approaches (Usman, 2013). Driven by the hypothesis that the deep learning classification k-model with a large and diverse training dataset would yield more robust and reliable results than traditional methods and ML approaches, this study investigated its effectiveness for land cover classification using high-resolution satellite imagery (Ulaby, 2018) and (Zhou et al., 2018). The findings of this research provided valuable insights into the potential and limitations of this novel approach, contributing to the ongoing advancement of efficient and accurate LULC classification techniques.

II. Method

2.1 Data set and data acquisition

High-resolution satellite imagery from the USGS National Map Urban Area Imagery program served as the data source for this research (Yang & Newsam, 2010). These images had already undergone pre-processing, involved slicing and resizing to a standardized dimension of 256 x 256 pixels. The dataset encompassed a diverse range of land use and land cover (LULC) categories. However, for this research study, only 12 specific classes were chosen, carefully consolidated into four macro classes. Vegetation representing all naturally occurring plant life, water encompassing all bodies of water, both natural and artificial, impervious surfaces including roads, buildings, and other man-made structures that impede water infiltration and pervious surfaces encompassing soil, grass, and other surfaces that allow for water infiltration. The classification of different land cover types was strictly based on their inherent natural characteristics, as comprehensively illustrated in Figure 1.

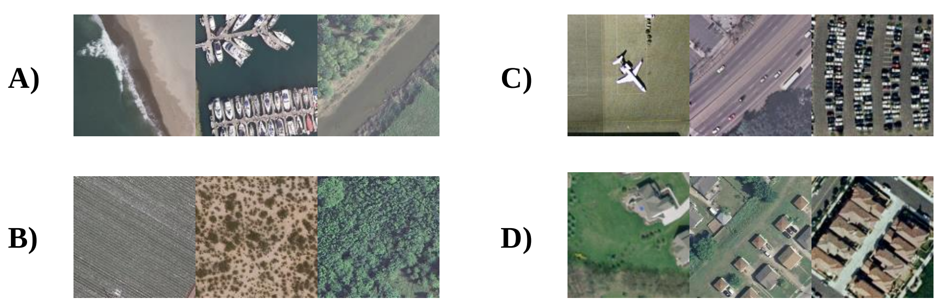

Figure 1. These are the datasets used in this study starting from left image (A) Water, (B) Vegetation, (C) Impervious surfaces and (D) Pervious surfaces.

To facilitate efficient analysis and ensure compatibility with the implemented deep learning algorithms, a vital preprocessing step was undertaken. The original image format, TIFF, was converted to JPEG, as visually depicted in Figure 2. This conversion facilitated seamless integration of the data into the chosen machine learning models, to achieve the success of the research process. Furthermore, the dataset was carefully partitioned, allocating 80% of each class for training and the remaining 20% for validation, as carefully detailed in Table 1. This standardized partitioning enabled strong training of the deep learning model and facilitated rigorous evaluation of its performance on unseen data.

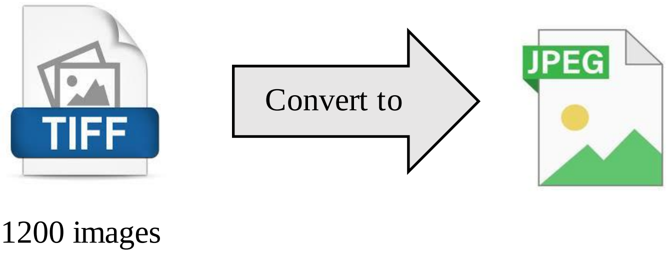

Figure 2. Conversion of image file format.

Table 1. Frequency of the sample training and validation datasets.

Macro Class	Training	Validation
Water	240	60
Vegetation	240	60
Impervious Surfaces	240	60
Pervious Surfaces	240	60
Total Images	960	240

2.2 Methods Procedures

In this study, carefully data preparation procedures were implemented to ensure the datasets were suitable for efficient processing by the chosen machine learning models. Following the consolidation of the dataset into four macro classes (vegetation, water, impervious surfaces, and pervious surfaces), three pre-trained deep learning algorithms were employed for classification: ResNet-50, Inception, and EfficientNetb1, see Figure 3. To optimize performance within the constraints of the available hardware, careful tuning of the training parameters was undertaken (Konar et al., 2020). This involved utilizing a relatively small batch size of 8 and limiting the training epoch to 5. This optimized approach minimized the risk of overfitting while maximizing the efficiency of the learning process. All images were resized to a consistent dimension of 244 x 244 pixels and image augmentation techniques were applied to artificially increase the size and diversity of the training dataset (Mikołajczyk and Grochowski, 2018). These techniques include random rotations with a range of up to 45 degrees, enhancing the model's generalizability by exposing it to a wider range of data variations. To leverage the pre-trained weights and prevent overfitting, the first 15 layers of each pre-trained model were frozen using the learner.freeze(freeze_range=15) function. This strategy ensured the model focused on learning the specific features crucial for the LULC classification task, preventing it from modifying the pre-trained features already optimized for general image recognition. These parameters were carefully evaluated for each pre-trained model, ensuring optimal performance and efficient use of the available computing resources.

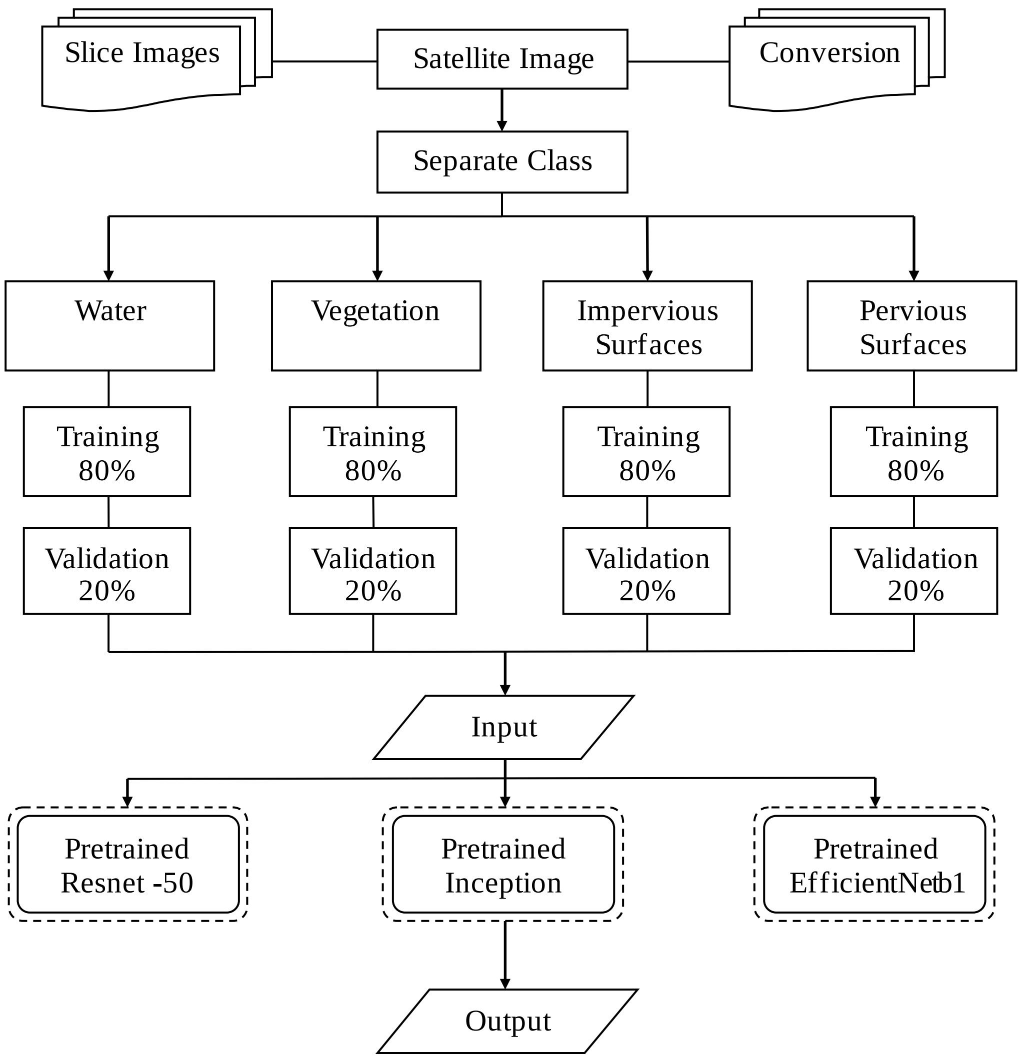

Figure 3. Flow chart of the learning process.

2.3 Machine Learning Models and Performance Metrics

In this section, we first briefly introduce the typical architecture of CNNs and then review the pre-trained CNN models evaluated in our work.

2.3.1 CNN Architecture

A Convolutional Neural Network (CNN) architecture comprises essential components, including various layers such as convolutional layers, pooling layers, and fully connected layers. In each convolutional layer, a fixed number of filters (also referred to as kernels or weights) is employed to generate an equivalent number of feature maps by traversing these filters across the feature maps from the preceding layer. Pooling layers undertake subsampling along the spatial dimensions of the feature maps, reducing their size through methods like max or average pooling. Following the convolutional and pooling layers, fully connected layers are integrated (Zhou et al., 2017). Figure 4 illustrates the standard structure of a CNN model.

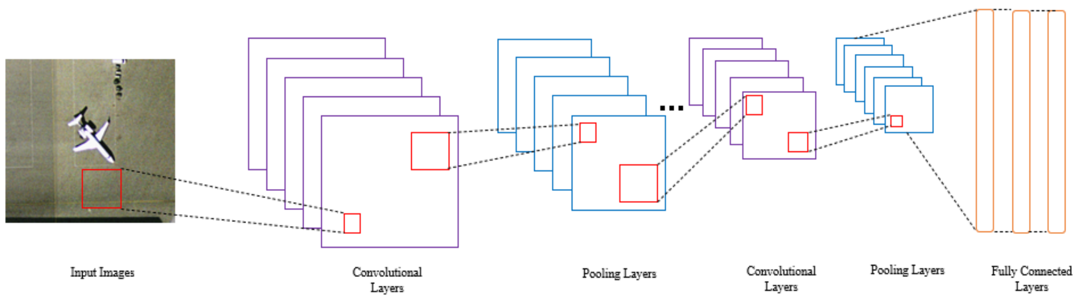

Figure 4. The typical architecture of Convolutional Neural Networks (CNNs.)

2.3.2 Pre-Trained Models

We utilized three pre-trained deep learning models to classify land use and land cover (LULC): ResNet-50, Inception, and EfficientNetb1. These models were chosen for their proven performance in image classification tasks and their ability to transfer learned features to different domains. The categorical cross-entropy loss function was used to measure the model's performance, and the Adam optimizer was used to update the model's parameters. The following metrics were used to evaluate the performance of the trained models (Teemu Kanstrén, 2020).

A CNN algorithm pre-trained ResNet-50 model is used in this study, for LULC classification. This choice was motivated by its proven effectiveness, achieving high accuracy (97.99%) in key papers like (Feng et al., 2021). Additionally, it offered significant efficiency benefits, reducing training time by 90% compared to training from scratch (Vali et al., 2020). Lastly, the pre-trained weights provided valuable transfer learning capabilities, allowing the model to learn specific land cover patterns quickly. Overall, ResNet-50's accuracy, efficiency, and transfer learning capabilities made it the ideal for this research.

Pretrained Inception is designed to maximize the utilization of computing resources within the network, Inception V3 CNN increases depth and width while maintaining computation operations. Inception modules are built as building blocks of this network and are termed by its designers as an optimized network structure with skipped connections (Li et al., 2021). The inception module is repeated spatially by stacking with occasional max-pooling layers to reduce the degree of dimensionality to a manageable computation level (Jahandad et al., 2019).

Building upon recent advancements in deep learning for LULC classification, this study employed the pre-trained EfficientNet B1 model. This selection was informed by compelling research, highlighting EfficientNet B1's exceptional accuracy and efficiency in LULC tasks (Tan and Le, 2019). Furthermore, its ability to extract salient features from diverse land cover types is crucial for differentiating subtle variations within LULC maps. In this study, EfficientNet B1 is utilized to investigate its suitability for achieving accurate and reliable LULC classification results.

2.3.3 Metrics

This section introduces the examination of model performance and the metrics employed in this study. To gauge the efficacy of the CNN models for land-use classification, a range of comprehensive metrics, including precision, recall, F1 score, and accuracy, were utilized to provide a nuanced understanding of their classification capabilities.

Recall is the proportion of actual positive examples that are correctly identified by the model.

$$\frac{TruePositives}{TruePositives + FalseNegatives} = \frac{N.ofCorrectlyPredictedPositiveInstances}{N.ofTotalPositiveInstancesintheDataset} = \frac{N.ofCorrectlyPredictedLULCeachclass}{N.ofLULCclassintheDataset} \tag{1}$$

Accuracy is the overall percentage of predictions that are correct.

$$\frac{TruePositives + TrueNegatives}{TruePositives + TrueNegatives + FalsePositives + FalseNegatives} = \frac{N.ofCorrectPredictions}{N.ofallPredictions} = \frac{N.ofCorrectPredictions}{SizeofDataset} \tag{2}$$

Precision is the proportion of predicted positive examples that are positive.

$$\frac{TruePositives}{TruePositives + FalseNegatives} = \frac{N.ofCorrectlyPredictedPositiveInstances}{N.ofTotalPositiveInstancesintheDataset} = \frac{N.ofCorrectlyPredictedLULCeachclass}{N.ofLULCyoupredictedeachclass} \tag{3}$$

F1-score is the harmonic mean of recall and precision, providing a more balanced measure of model performance.

$$\frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{4}$$

2.3.4 Early stopping

Early stopping is a technique used to prevent deep neural networks from overfitting on training data. It works by monitoring the validation accuracy of the network during training. When the validation accuracy stops improving significantly, training is stopped, and the best weights seen so far are used. This helps to prevent the network from becoming too specialized to the training data, which can lead to poor performance on new data (Naushad et al., 2021).

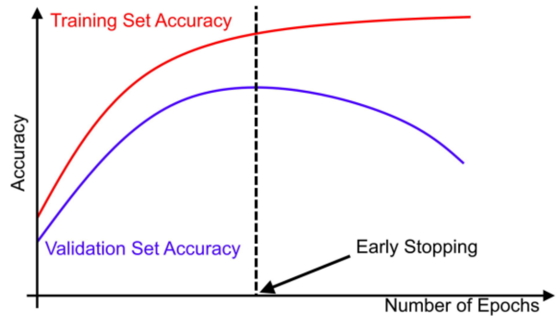

Figure 5. In early stopping, training is stopped after a predefined number of epochs if the validation accuracy hasn't improved by a certain threshold for those epochs (Hatef Dastour and Hassan, 2023).

III. Result

This section presents the primary objective of this study which are to assess the performance of three ML models in the task of Land Use and Land Cover classification using satellite images. The experiment aimed to evaluate the models' capability to accurately classify different land-use classes based on the input imagery. To quantify the performance of the models, key metrics such as recall, precision, F1 score, and accuracy were employed. These metrics were selected to provide a comprehensive understanding of the models' abilities to correctly identify and classify various land-use categories, forming the basis for a detailed analysis of their strengths and limitations.

3.1 Pretrained ResNet – 50

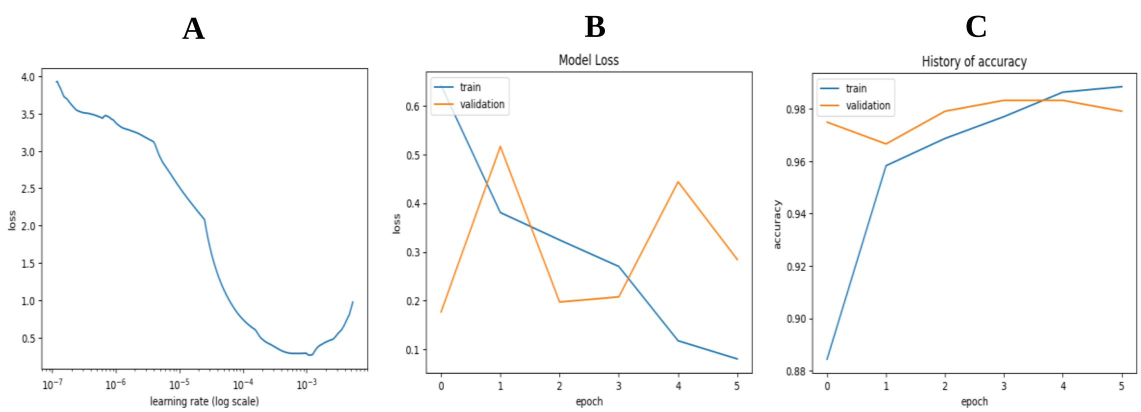

Figure 6. Plot presents the learning rate (A), loss curves (B), and accuracy curves (C) of the pre-trained ResNet-50 model, offering insights into its performance for LULC classification.

Table 2. Metric performance of Pretrained ResNet – 50 learning result.

Model Performance Comparison	Pretrained ResNet -50			
	Vegetation	Water	Impervious Surfaces	Pervious Surfaces
Precision	94%	100%	98%	98%
Recall	100%	93%	98%	98%
F1- Score	97%	97%	98%	98%
Accuracy	97%			

The plot shows the relationship between the training and validation loss and accuracy for a K-train pre-trained ResNet-50 model trained on a LULC classification task. The loss measures how well the model's predictions match the ground truth labels. The accuracy measures the proportion of correct predictions made by the model. The graph suggests that the K-train pre-trained ResNet-50 model can learn the training data effectively. However, the model may be overfitting the training data, as evidenced by the plateauing validation loss curve. This could lead to poor performance on new, unseen data. The table 2 presents the performance metrics of a pre-trained ResNet-50 model employed for the task of LULC classification. The results of the pre-trained ResNet-50 model performed exceptionally well in classifying land cover types.

3.2 Pretrained Inception

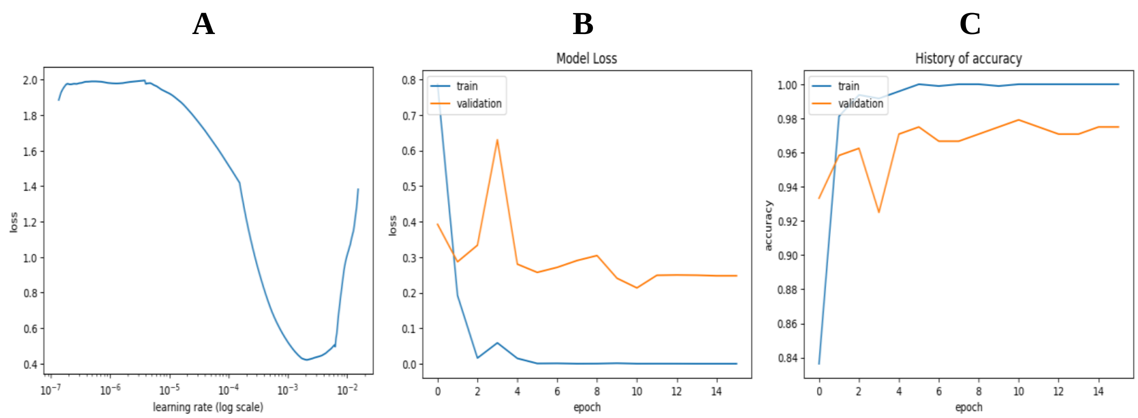

Figure 7. Plot presents the learning rate (A), loss curves (B), and accuracy curves (C) of the Pretrained Inception model.

Table 3. Metric performance of Pretrained Inception learning result.

Model Performance Comparison	Pretrained Inception			
	Vegetation	Water	Impervious Surfaces	Pervious Surfaces
Precision	98%	97%	97%	100%
Recall	98%	98%	98%	97%
F1- Score	98%	98%	98%	98%
Accuracy	98%			

The results suggest in the graphs that it becomes evident that the model effectively learns the training data. However, a plateau in the validation loss curve suggests potential overfitting, which could negatively impact performance on unseen data**.** While the results appeared impressively high, potentially meeting the minimum success rate for the classification, the validation loss suggested overfitting during training. This raises concerns about the model's ability to generalize to unseen data, even with its seemingly strong performance.

3.3 Pretrained EfficientNetb1

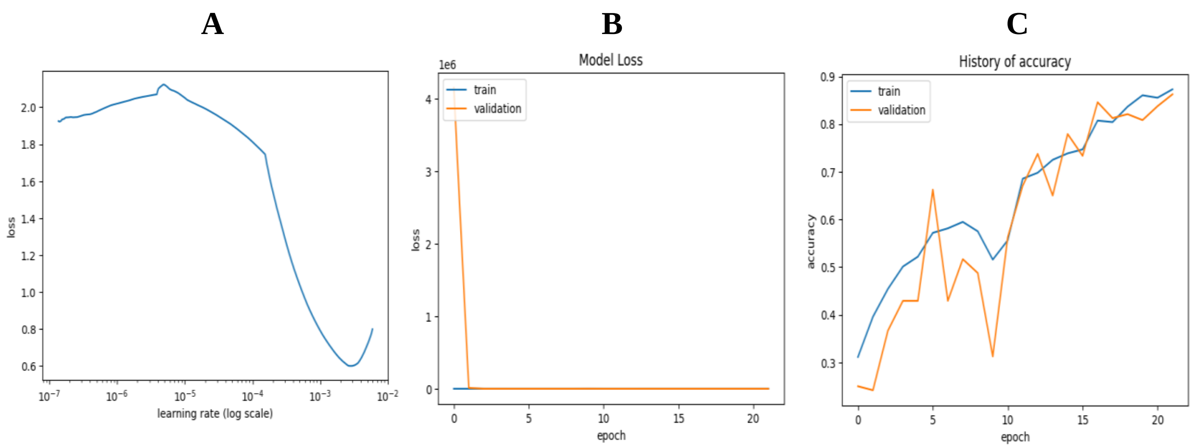

Figure 8. Plot presents the learning rate (A), loss curves (B), and accuracy curves (C) of the Pretrained EfficientNetb1 model.

Table 4. Metric performance of Pretrained EfficientNetb1 learning result.

Model Performance Comparison	Pretrained EfficientNetb1			
	Vegetation	**Water**	**Impervious Surfaces**	**Pervious Surfaces**
Precision	98%	82%	82%	77%
Recall	98%	90%	67%	83%
F1- Score	98%	86%	73%	80%
Accuracy	85%			

The EfficientNetB1 model showed a good result for LULC classification, but further investigation into its performance on individual classes and exploration of different training methods or hyperparameters could potentially improve its accuracy and generalizability. The model achieved an accuracy of 85% on this LULC classification task. While its performance on vegetation was strong, its accuracy on other classes, particularly impervious and pervious surfaces, was lower. This suggests that the model may not be generalizable to all LULC classes and may require further training or adjustments to improve its performance.

3.4 Confusion Matrix

In the evaluation of three pre-trained models—Inception, EfficientNetB1, and

ResNet50—for land-use and land-cover classification, the results suggest that distinct strengths and weaknesses were observed across various land-use classes. Inception demonstrated exceptional accuracy in classifying vegetation and impervious surfaces, achieving near-perfect results. However, it faced challenges in accurately identifying water. EfficientNetB1 exhibited a balanced performance, excelling in both vegetation and pervious surfaces but encountering similar challenges in water classification. Moreover, ResNet50 emerged as a good model in water classification, achieving outstanding accuracy in this class and demonstrating impressive dominance in impervious surface classification.

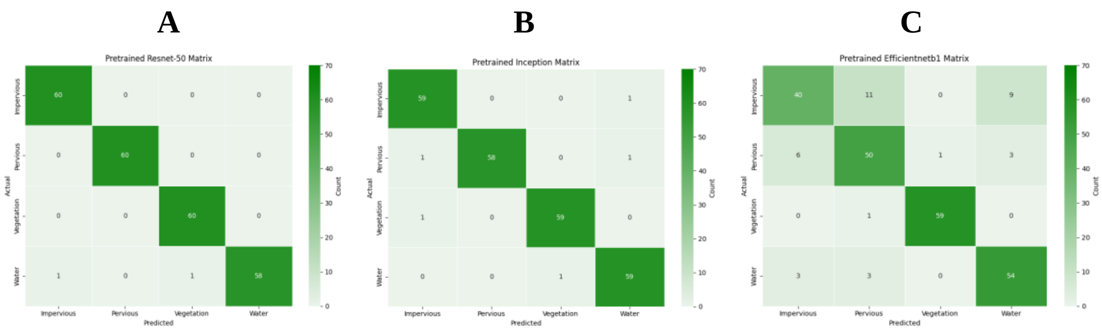

Figure 9. Confusion matrix of all pretrained models from left plot (A) ResNet-50, (B) Inception, and (C) EfficientNetb1.

IV. Discussion

The 3 model introduce a good result but it’s still the model loss and validation loss suggest that the ResNet- 50 and inception may get to complex when apply to unseen data as shown in the graph plot. In this study the EfficientNetB1 model achieved an impressive 80% training accuracy, but the plateauing validation loss curve suggests potential overfitting. This raises concerns about its generalizability to unseen data, especially considering the impact of image resolution on CNN performance (Laith Alzubaidi et al., 2021) and (Ioffe and Szegedy, 2015).

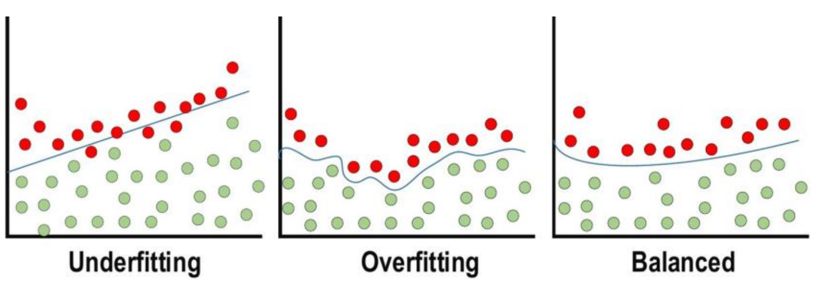

Figure 10. Overfitting and Underfitting issues.

Studies have shown that low-resolution images can lead to information loss, making it harder for the model to capture crucial features and relationships between different LULC classes (Alzubaidi et al., 2021). This could explain the model's struggle with generalizability despite its seemingly high training accuracy. Additionally, incorporating data augmentation techniques specifically designed for LULC tasks, like random rotations and flips, could further improve the model's robustness to variations in image resolution and lighting conditions. Other challenges that may the model interacts is the amount of data that is given to the model, it suggests that the big amount of data helps the model to learn more about the characteristics of the image classification (Alzubaidi et al., 2021). Studies suggest that the DL performance works may not be good as the amount of data may not balance. In the case of this study augmentation is adapted to increase the data and flip it once no other types applied which may suggest being the model hard to learn and overfit the result see Fig. 10 (Shorten and Khoshgoftaar, 2019). Still, there is a limitation in this study due to many obstructions from both water, vegetation, and pervious surfaces (Naushad et al., 2021). Some of the water parts may contain boats and built ups near harbor, which could hinder the learning process of the models (Konar et al., 2020) and (Vali et al., 2020). The future study may establish guidelines for each feature in the image, arrange them based on specific region of water and pervious surfaces before training, and separate the individual classes accordingly. Further study might also include segmentation of each image features in each macro classes which can improve the validation of prediction (Abu et al., 2020) and (Henry et al., 2019). In this research, the performance of Pretrained Inception and Pretrained EfficientNetb1 with validation datasets were intensively compared. The accuracy of Pretrained Inception on the USGS National Map Urban Area Imagery dataset was found to be better than Pretrained EffiecinetNetb1 by at least 13% of the overall accuracy set. As mentioned in Table 3, the best performing model of Pretrained Inception achieved 98%, whereas it was 85% for Pretrained EfficientNetb1. Thus, it was understood that Pretrained Inception performed better that Pretrained EffiencentNetb1. As shown in the results, the accuracy of 98% achieved using Pretrained Inception with RGB bands was higher than the highest accuracy of 97% achieved using the Pretrained ResNet-50 model with RGB bands.

V. Conclusion

In conclusion, this study investigated the effectiveness of Inception, EfficientNetB1, and ResNet50 models for LULC classification utilizing pre-trained Deep Convolutional Neural Networks (DCNNs). Results showed that all models had potential capabilities, with Inception obtaining the best overall accuracy. Individual advantages and disadvantages were, nevertheless, revealed by class-specific analysis, emphasizing the need to select models appropriate for LULC classification tasks. The study also found that certain models may have been overfitted, which suggests more research into data augmentation and hyperparameter tuning strategies to increase generalizability. This study provides a basis for more advanced transfer learning in LULC classification by providing a structure for future research exploring further complex models and methods. Researchers can more effectively analyze complex earth observational data by utilizing artificial intelligence, which will eventually assist in fulfilling sustainable development goals.

References

Abu, N., Sarker, A., Paul, O., Ali, A., Amin, M. A., Rahman, A. M., 2020. LULC Segmentation of RGB Satellite Image Using FCN-8. *ArXiv.org*. doi.org/10.48550/arXiv.2008.10736.

Chen, S., Lei, F., Dong, S., Zang, Z., Zhang, M., 2022. Land use/land cover mapping using deep neural network and sentinel image dataset based on google earth engine in a heavily urbanized area, China. *Geocarto International*, 37(27), 16951–16972. doi.org/10.1080/10106049.2022.2120551.

Feng, X., Gao, X., Luo, L., 2021. A ResNet50-Based Method for Classifying Surface Defects in Hot-Rolled Strip Steel. *Mathematics*, 9(19), 2359–2359. doi.org/10.3390/math9192359.

Hatef Dastour, Hassan, Q. K., 2023. A Comparison of Deep Transfer Learning Methods for Land Use and Land Cover Classification. *Sustainability*, 15(10), 7854–7854. doi.org/10.3390/su15107854.

Henry, Christopher J., Christopher D. Storie, Muthu Palaniappan, Victor Alhassan, Mallikarjun Swamy, Damilola Aleshinloye, Andrew Curtis, and Daeyoun Kim, 2019. Automated LULC map production using deep neural networks. *International Journal of Remote Sensing*, 40(11), 4416-4440. doi.org/10.1080/01431161.2018.1563840.

Jahandad, Suriani Mohd Sam, Kamilia Kamardin, Nur, N., Mohamed, N., 2019. Offline Signature Verification using Deep Learning Convolutional Neural Network (CNN) Architectures GoogLeNet Inception-v1 and Inception-v3. *Procedia Computer Science*, 161, 475–483. doi.org/10.1016/j.procs.2019.11.147.

Kanstrén, T., 2020. A look at precision, recall, and F1-score. *Towards Data Science*, 12(10).

Konar, J., Khandelwal, P., Tripathi, R., 2020. Comparison of various learning rate scheduling techniques on convolutional neural network. *In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (pp. 1-5). IEEE. doi.org/10.1109/SCEECS48394.2020.94.

Laith Alzubaidi, Zhang, J., Humaidi, A. J., Al-Dujaili, A. Q., Duan, Y., Omran Al-Shamma, José Santamaría, Fadhel, M. A., Muthana Al-Amidie, Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1). doi.org/10.1186/s40537-021-00444-8.

Li, Y., Yang, C., Huang, J., Liu, S., Zhuo, Y., Lu, X., 2021. Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. *Future Generation Computer Systems*, 114, 358–367. doi.org/10.1016/j.future.2020.08.015.

Liu, P., Zhang, H., Eom, K., 2017. Active Deep Learning for Classification of Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2), 712–724. doi.org/10.1109/jstars.2016.2598859.

Mikołajczyk, A., Grochowski, M., 2018. Data augmentation for improving deep learning in image classification problem. *In 2018 international interdisciplinary PhD workshop (IIPhDW)* (pp. 117-122). IEEE.doi.org/10.1109/IIPHDW.2018.8388338.

Piramanayagam, S., Schwartzkopf, W., Koehler, F. W., Saber, E., 2016. Classification of remote sensed images using random forests and deep learning framework. *Proceedings of SPIE*. doi.org/10.1117/12.2243169.

Raoof Naushad, Kaur, T., Ebrahim Ghaderpour, 2021. Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study. *Sensors*, 21(23), 8083–8083. doi.org/10.3390/s21238083.

Rashid, M., Hegde, C., Rajashekarappa, K., 2020. Transfer Learning Models for Land Cover and Land Use Classification in Remote Sensing Image. Journal of the Indian Society of Remote Sensing, 48(6), 1101-1111. doi.org/10.1007/s12524-020-01120-w.

Shorten, C., & Khoshgoftaar, T. M., 2019. A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), 1-48.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *In International conference on machine learning* (pp. 6105-6114). PMLR.

Ulaby, F. T., 2018. Introduction to Satellite Remote Sensing: Atmosphere, ocean, land, and cryosphere applications [Book Review]. *IEEE Geoscience and Remote Sensing Magazine*, 6(4), 109–110. doi.org/10.1109/mgrs.2018.2873040.

Usman, B., 2013. Satellite imagery land cover classification using k-means clustering algorithm computer vision for environmental information extraction. *Elixir International Journal of Computer Science and Engineering*, 63, 18671-18675.

Vali, A., Comai, S., Matteucci, M. (2020). Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sensing*, 12(15), 2495–2495. doi.org/10.3390/rs12152495.

Yang, Y., & Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. *Research Gate*, 270–279. doi.org/10.1145/1869790.1869829.

Zhou, W., Newsam, S., Li, C., Shao, Z., 2018. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *Isprs Journal of Photogrammetry and Remote Sensing*, 145, 197–209. doi.org/10.1016/j.isprsjprs.2018.01.004.