Grenze International Journal of Engineering and Technology, January Issue

Deep Learning-based Classification of Geographical Land Structure using Satellite Images

M Prabhavathy[1], Valliappan Raman[2] and Putra Sumari[3]
[1-2]Department of AI and DS, Coimbatore Institute of Technology, Coimbatore, India,
Email: prabhavathy.phd@gmail.com, valliappan@cit.edu.in
[3]School of Computer Science, University Sains Malaysia, Penang, Malaysia
Email: putras@usm.my

***Abstract*— Detecting the generic land structure in satellite images is typically done manually, requiring significant time and effort. Thus, the project proposes an approach to reduce human efforts, reducing the cost and time for identifying the land structure. In this project, a deep learning-based approach was developed to automate the process of classifying geographical land structures. The study compared the performance of three architectures, namely CNN, ResNet-50, and Inception-v3. The performance of the proposed model, CNN has achieved an accuracy of 94.8%. The results highlight the potential of deep learning models in scene understanding and their significance in efficiently identifying and categorizing land structures from satellite imagery.**

I. Background study and Aim

Deep learning algorithms have brought a revolution to the computer vision domain [1] by introducing effective solution such as classification of an image. Detecting and classifying geographical land structures from satellite images is an important task in various fields such as environmental monitoring, urban planning, and land management.

Traditionally, this process has relied on manual analysis, which is time-consuming and subject to human error. With the advancements in deep learning and computer vision, there is an opportunity to automate and improve the accuracy of land structure classification. The aim of this project is to develop a deep learning-based model that can accurately classify different geographical land structures in Malaysia using satellite images. By leveraging state-of-the-art architectures such as CNN, ResNet-50, and Inception-v3, the project aims to explore the performance of these models and identify the most effective approach for land structure classification. The ultimate goal is to provide a reliable and efficient solution for automated scene understanding, which can have significant applications in various domains such as land use planning, environmental conservation, and disaster management.

II. Related Work

Deep learning-based approaches have gained widespread popularity in digital image analysis, finding applications in various domain such as human detection [2], vehicle detections [3], crops detections [4], medical image and even in the detection of diseases such as Covid-19 [5]. These techniques have revolutionized the field by leveraging the power of neural networks to automatically learn and extract meaningful features from images, enabling accurate and efficient analysis.

Grenze ID: 01.GIJET.11.1.371

Yang et al. [6] proposed a sparsity model called Dictionary Learning CNN or DL-CNN, where the CNN incorporates dictionary learning layers instead of fully connected layers, as illustrated in Fig. 1. This modification helps reduce the number of parameters in the CNN model while obtaining enhanced sparse features. The DL-CNN model has two nodes in the output layer: the first node represents the discrete probability distribution across different categories, and the second node represents the corresponding discriminant sparse representation. The proposed model achieved a validation accuracy of 96.03% when evaluated on the 15 Scenes dataset.

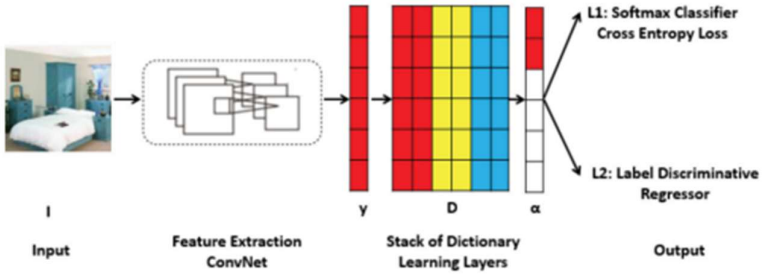

Fig. 1 - Illustration of the DL-CNN architecture [6]

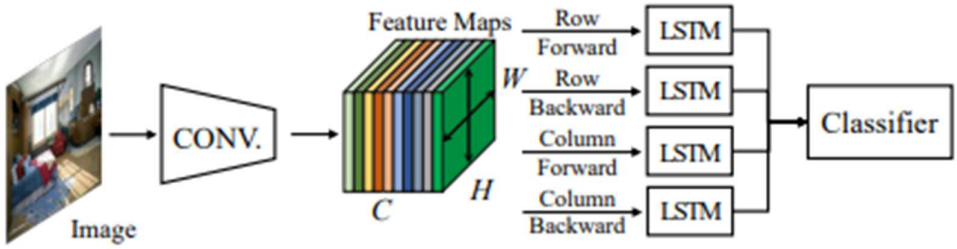

Fig. 2 - Illustration of the CFA architecture [7]

Sun et al. [8] introduced a scene identification model called Contextual Features in Appearance (CFA). In this model, the output of the convolutional layers serves as the input for the Long Short-Term Memory (LSTM) layers, which are organized into four directed sequences in a cyclic manner. The illustration of the CFA architecture has been shown in Fig. 2.

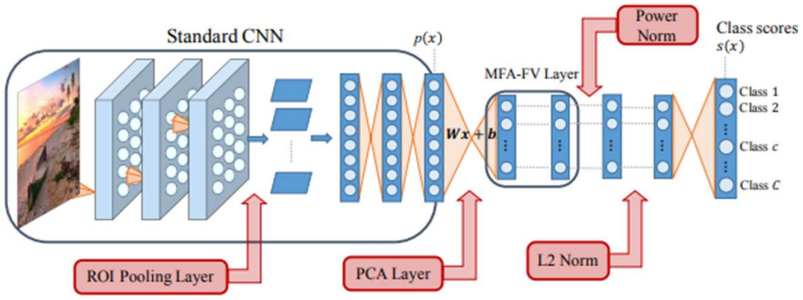

Fig. 3 - Illustration of the MFAFVNet architecture

Li et al. [9] proposed a method for classifying scene images using MFAFVNet, a transfer learning Convolutional Neural Network based on the theory of Mixture of Factor Analyzer (MFA) and Fisher Vector (FV). The MFAFV layers will be added after standard convolutional neural network and then as shown in Fig. 3. The authors added an MFAFV layer to VGG16 and AlexNet to enable comparison. The model was trained on the ImageNet dataset. This approach achieved an accuracy of 80.47% when using AlexNet as the base model and 87.97% when using VGG16.

Damir et al. [10] proposed a method for classifying natural landscape images using transfer learning based on the Inception V3 Convolutional Neural Network

TABLE 1 - OVERVIEW OF RELATED WORKS

Paper	Model	Best Accuracy
[6]	DL-CNN	96.03%
[8]	CFA	78.93%
[9]	AlexNet + MFAFVNet	80.47%
	VGG16 + MFAFVNet	87.97%
[10]	Transfer Learning with Inception V3 CNN	92.9%

Table 1 provides an overview of the related works where researchers have proposed various state-of-the-art methods for scenery detection. Methodology

A. Dataset Description

The team utilized a satellite image dataset obtained from MLRSNet, a multi-label high spatial resolution remote sensing dataset designed for semantic scene understanding [11]. MLRSNet offers diverse perspectives of the world captured by satellites, providing high-resolution imagery.

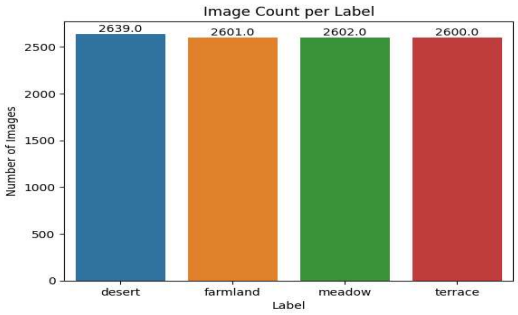

Fig. 4 - Image counts based on label

As previously mentioned, the team utilized a dataset with four distinct labels, each containing approximately 2600 images which the results can be visualized in Fig. 4.

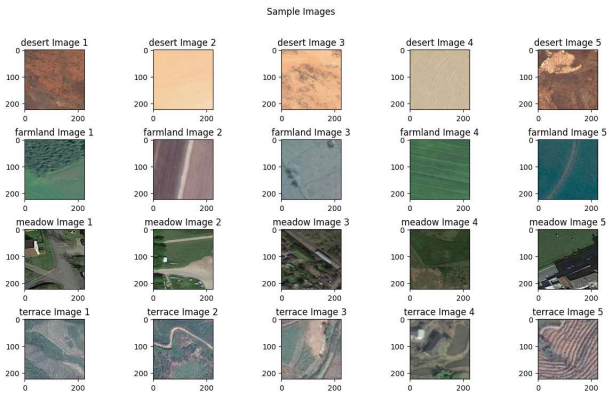

Fig. 5 - Examples of images from the dataset

Fig. 5 shows few examples from the datasets based on desert, farmland, meadow and terrace. This figure also shows the size of the images is at 224x224 pixels. This is because to ensure uniformity and compatibility, all images were resized to a resolution of 224x224 pixels. This resizing process ensures that all images have the same dimensions, facilitating consistent processing and analysis. By standardizing the image size, the team can effectively apply machine learning algorithms and techniques to extract meaningful features and patterns from the dataset.

To expand the dataset further, the team employed image augmentation techniques. Image augmentation is a widely used approach to enhance the generalization performance of models by introducing variations to the data [12]. Fig. 6 provides an illustration of various image augmentation transformations, such as rotation, flipping,

shearing, and more which done by one of the “terrace” image. In the context of categorical variables, label encoding, also known as ordinal encoding, assigns a unique integer value to each category within the dataset [13].

Random sampling is a widely used technique for creating a representative subset of a dataset [14]. It is particularly useful for dividing the data into training and testing sets. Another related method is stratified sampling, which is a probability sampling approach that takes into account the distinct groups or categories within the dataset [15]

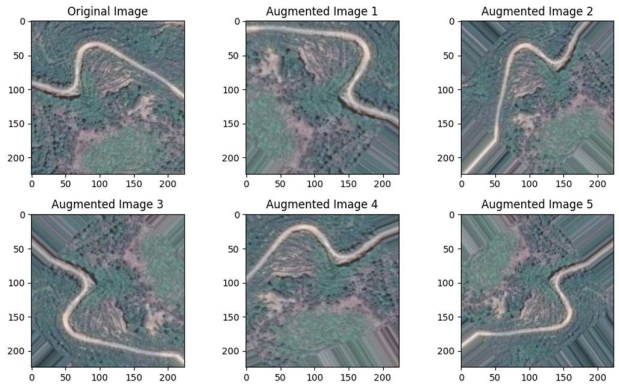

Fig. 6 - Example of image augmentation

B. Convolutional Neural Network (CNN)

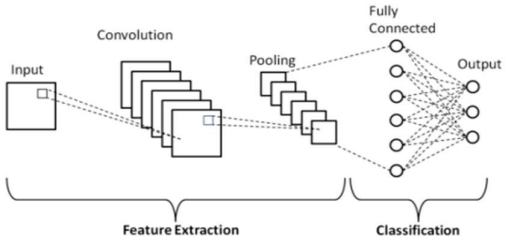

Fig. 7 - Typical architecture of CNN [16]

A Convolutional Neural Network, also known as a CNN, is a type of artificial neural network characterized by its deep feed-forward architecture [17]. CNNs are specifically designed for processing and analyzing visual data, such as images. A CNN typically comprises convolutional layers, which apply sets of learnable filters to the input image, allowing the network to extract relevant features and patterns.

For instance, in an RGB image, the depth would be three, representing the red, green, and blue channels. The convolutional filters in the CNN perform local operations, calculating the dot product between the input and the filter weights. These operations allow the network to capture spatial information and learn meaningful features from the input image [18].

After the convolutional layers, the feature maps are typically down sampled using pooling layers. Pooling layers reduce the spatial dimensionality of the feature maps, which accelerates the training process and helps prevent overfitting by summarizing the most important information in the feature maps [19].

C. Proposed CNN Architecture

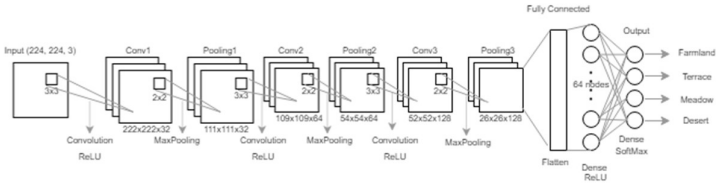

Fig. 8 - Architecture of proposed CNN

The team has proposed a CNN model to classify the geographical land structure, as shown in Fig. 8. The CNN architecture takes input images with dimensions of 224x224x3, representing the height, width, and depth of the images.

Furthermore, the Glorot uniform initialization, also known as the Xavier uniform initialization, is utilized as the kernel initialization method [20]. This initialization method contributes to the overall stability and effectiveness of the learning process in the CNN, enabling the model to learn and extract meaningful features from the input data.

$$f(x) = \max(0, x)$$

Equation 1 - ReLU activation function

The ReLU activation function operates by returning the input value directly if it is positive, and outputting zero if the input is negative as shown in Equation 1.

D. Transfer Learning

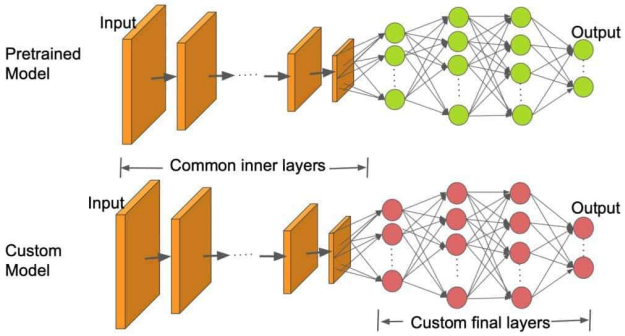

Fig. 9 - Illustration of Transfer Learning

Transfer learning is a technique that involves leveraging the knowledge and insights gained from solving one task and applying it to solve related tasks [21]. It is commonly applied when the domains of the tasks are similar to have a greater positive impact [22]. In the context of artificial neural networks as illustrated in Fig. 9, transfer learning involves utilizing a pretrained model as a starting point. The pretrained model, with its existing layers and weights, serves as a foundation for the custom model.

Residual Network or ResNet, proposed by Kaiming et al. [23], plays a significant role in enhancing the training process of neural networks by formulating the layers as learning residual functions with reference to the previous layer. The concept of residual stems from the difference between the actual observed value and the estimated value [24]. One prominent variant of ResNet is ResNet-50, which is a 50-layer convolutional neural network comprising 48 convolutional layers, one max pooling layer, and one average pooling layer. In the context of classifying scenes, ResNet-50 has been effectively employed. However, for this task, the original output layer of ResNet-50 has been replaced with a dense layer consisting of four nodes, employing the softmax activation function to predict the four categories of the labels.

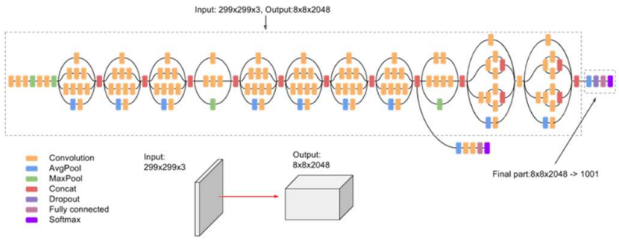

Fig. 10 - Architecture of Inception-v3

In 2014, Szegedy et al. introduced the Inception model, a deep CNN architecture that was developed as part of the GoogleNet during the Large-Scale ImageNet Visual Identification Challenge 2014 [25]. The goal of Inception-v3 was to strike a balance between computational efficiency and low parameter count, making it suitable for practical applications [26]. This CNN architecture, which consists of 48 layers, was trained on the

ImageNet dataset, which comprises over 1 million images. Fig. 10 provides an illustration of the architecture of the Inception-v3 model. In this model, the initial convolutional layer receives input data with dimensions of 299x299x3 and produces an output tensor of size 8x8x2048. Inception-v3 model has also been used to identify the scenes or landscape.

III. Experiment and Result

In order to fine-tune and validate the model, the team has chosen several optimizers and learning rates. These optimizers include the Adam optimizer, Stochastic Gradient Descent (SGD) optimizer, and RMSprop optimizer. The learning rates selected for training are 0.001 and 0.0001. To train the model, each combination of optimizer and learning rate is used. The training process is conducted for a fixed number of epochs, specifically 100 epochs, with a batch size of 64. These models will be trained, tested and validated on same sample of dataset which the training set consists of 8,400 images (60%), the testing set contains 4,200 images (30%), and the validation set comprises 1,400 images (10%).

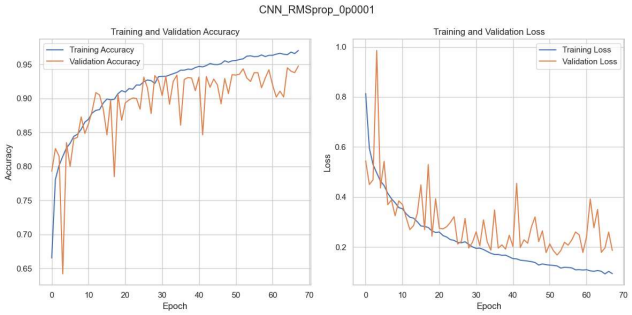

Fig. 11- Training history of CNN with RMSProp and 0.0001 learning rate

Fig. 11 illustrates the training history of a CNN model utilizing the RMSProp optimizer with a learning rate of 0.0001. The plot showcases the progression of training and validation accuracy throughout the epochs. Initially, both training and validation accuracy steadily improve, indicating the model's ability to learn and generalize from the training data. However, as the training progresses, the rate of improvement in training accuracy begins to slow down, reaching a plateau around the 50-epoch mark.

Table II - Training Results for CNN

MODEL	OPTIMIZER	LEARNING RATE	ACCURACY	LOSS
CNN	ADAM	0.001	92.1%	0.2331
		0.0001	93.7%	0.2079
	SGD	0.001	61.3%	1.0444
		0.0001	74.9%	0.6328
	RMSProp	0.001	92.1%	0.2530
		0.0001	**94.8%**	**0.1727**

Table II presents the training results of the CNN model using different combinations of optimizers and learning rates. The findings indicate that CNN performs better with a lower learning rate of 0.0001, regardless of the optimizer used. Among the models trained with various parameters, the CNN model trained with RMSProp optimizer stands out with a test accuracy of 94.8% and a test loss of 0.1727. On the other hand, the CNN model trained with SGD optimizer did not perform well in this case, exhibiting the lowest results with a test accuracy of 61.3% and a test loss of 1.0444 when the learning rate was set to 0.001. These observations suggest that a lower learning rate facilitates better performance for the CNN model in this scenario, and the choice of optimizer also plays a significant role in achieving optimal results.

Next, Table III shows the performance of ResNet-50 trained with different parameters. In contrast to the CNN model, the impact of a low learning rate on the ResNet-50 model's performance is not as significant. However, it

is worth noting that using a low learning rate in combination with the RMSProp optimizer may result in poorer performance.

TABLE III - TRAINING RESULTS FOR RESNET-50

MODEL	OPTIMIZER	LEARNING RATE	ACCURACY	LOSS
RESNET-50	ADAM	0.001	76.5%	0.6137
	SGD	0.0001	**77.3%**	**0.6219**
		0.001	45.3%	1.1683
		0.0001	50.3%	1.3311
	RMSProp	0.001	75.0%	0.6712
		0.0001	65.2%	0.8243

TABLE II - TRAINING RESULTS FOR INCEPTION-V3

MODEL	OPTIMIZER	LEARNING RATE	ACCURACY	LOSS
INCEPTION-V3	ADAM	0.001	93.5%	0.1958
	SGD	0.0001	93.7%	0.1867
		0.001	91.9%	0.2350
		0.0001	85.7%	0.4225
	RMSProp	0.001	93.2%	0.2048
		0.0001	**93.8%**	**0.1849**

Table IV shows the training performance of Inception-v3 under different parameter configurations. Interestingly, the performance of the three optimizers is quite similar in this case.

The Inception-v3 model trained with the RMSProp optimizer and a learning rate of 0.0001 achieved the best results, with a test accuracy of 93.8% and a test loss of 0.1849. Close behind, the model trained with the Adam optimizer and a learning rate of 0.0001 achieved a test accuracy of 93.7% and a test loss of 0.1867. Although slightly lower, it is a competitive performance. Furthermore, the performance gap between the SGD optimizer and the other optimizers has narrowed. The SGD model trained with a learning rate of 0.0001 achieved an 85.7% test accuracy and a test loss of 0.4225. While still not performing as well as the other optimizers, this demonstrates a relatively improved performance for the SGD optimizer with Inception-v3.

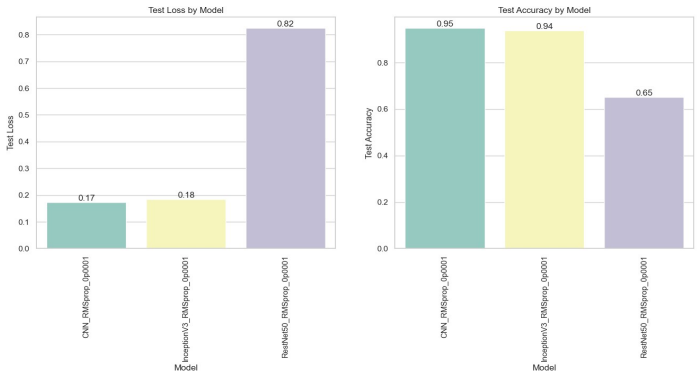

Fig. 12 - Comparison between different model

To summarize the previous results, Fig. 12 presents a bar chart comparing the performance of the best model from each of the different algorithms. The chart clearly demonstrates that the CNN model utilizing the RMSProp optimizer and a learning rate of 0.0001 perform better the ResNet-50 and Inception-v3 models.

After identifying the best model, which is the CNN trained with RMSProp optimizer and a learning rate of 0.0001, further analysis has been conducted. Fig. 13 presents the classification report of the best model. The

reports shows that the model trained performed very well, achieving high precision, recall and F1-scores across all classes. The precision scores for all classes are notably high, ranging from 0.92 to 0.98.

	precision	recall	f1-score	support
desert	0.98	0.98	0.98	1050
farmland	0.92	0.94	0.93	1050
meadow	0.95	0.92	0.93	1050
terrace	0.95	0.95	0.95	1050
accuracy			0.95	4200
macro avg	0.95	0.95	0.95	4200
weighted avg	0.95	0.95	0.95	4200

Fig. 13 - Classification report for the best model

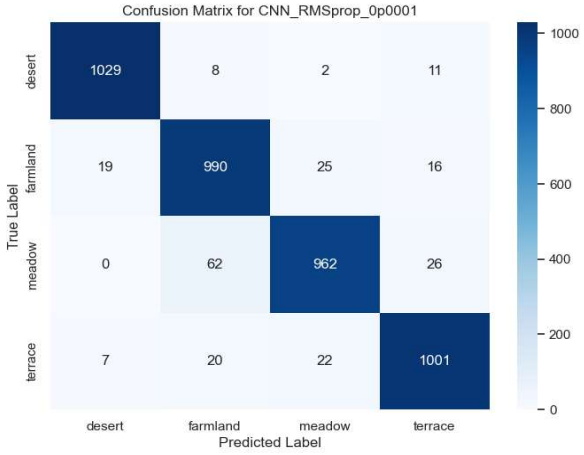

Fig. 14 - Confusion Matrix for the best model

Fig. 14 displays the confusion matrix, providing a visual representation of the model's predictions. The confusion matrix reveals the model's performance in correctly predicting the classes or labels of the images. Upon examining the matrix, it is evident that the model performs exceptionally well, with a high number of images correctly predicted across all classes.

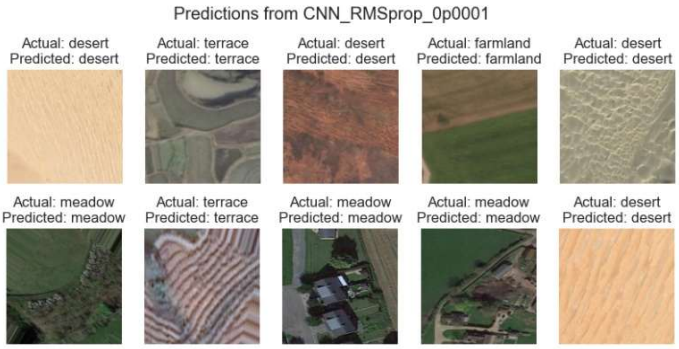

Fig. 15 - Predictions done by the best model

Fig. 15 presents the predictions made by the CNN model on a selection of samples from the testing set. For this particular analysis, 10 samples have been chosen to evaluate the model's performance. Remarkably, the model excels by correctly predicting all the classes or labels assigned to the images. This outcome demonstrates the high accuracy and effectiveness of the proposed CNN model in accurately detecting and classifying the scenes or geographical land.

IV. Conclusion

In conclusion, this project aimed to develop a deep learning model for the classification of geographical land structures with three different architectures, namely CNN, ResNet-50, and Inception-v3 to identify the most suitable model for this task. The experimental results revealed that the CNN model trained with RMSProp optimizer and a learning rate of 0.0001 achieved the best performance with the lowest test loss (0.17) and the highest test accuracy (95%) among all the tested models. Furthermore, in-depth analysis of the CNN model showcased its excellent precision, recall, and F1-scores across all classes, indicating its robustness and reliability in accurately identifying different land structures. Consequently, it can be concluded that the proposed CNN model outperformed ResNet-50 and Inception-v3 for the specific task of scene detection. This project demonstrates the efficacy of deep learning models in scene understanding and showcases the potential for their application in various real-world scenarios related to land structure identification and classification.

References

[1] M. Hassaballah and A. I. Awad, Deep Learning in Computer Vision: Principles and Applications, CRC Press Taylor & Francis Group, 2020.

[2] W. Rahmaniar and A. Hernawan, "Real-Time Human Detection Using Deep Learning on Embedded Platforms: A Review," Journal of Robotics and Control (JRC), vol. 2, 2021.

[3] Y. Wang, W. Deng, Z. Liu and J. Wang, "Deep Learning-based Vehicle Detection with Synthetic Image Data," IET Intelligent Transport Systems, vol. 13, 2019.

[4] C. Ke, N. T. Weng, Y. Yang, Z. M. Yang, P. Sumari, L. Abualigah, S. Kamel, M. Ahmadi, M. A. Al-Qaness, A. Forestiero and A. R. Alsoud, "Mango Varieties Classification-Based Optimization with Transfer Learning and Deep Learning Approaches," in Classification Applications with Deep Learning and Machine Learning Technologies, Cham, Springer International Publishing, 2023, pp. 45-65.

[5] P. Sumari, S. J. Syed and L. Abualigah, "Turkish Journal of Computer and Mathematics Education (TURCOMAT)," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 6, pp. 2001-2011, 2021.

[6] Y. Liu, Q. Chen, W. Chen and I. Wassell, "Dictionary Learning Inspired Deep Network for Scene Recognition," The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), pp. 7178-7185, 2018.

[7] D. Zeng, M. Liao, M. Tavakolian, Y. Guo, B. Zhou, D. Hu, M. Pietik{\"a}inen and L. Liu, "Deep learning for scene classification: A survey," arXiv preprint arXiv:2101.10531, 2021.

[8] N. Sun, W. Li, J. Liu, G. Han and C. Wu, "Fusing Object Semantics and Deep Appearance Features for Scene Recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 99, 2018.

[9] Y. Li, M. Dixit and N. Vasconcelos, "Deep Scene Image Classification with the MFAFVNet," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

[10] D. Krstinic, M. Braovic and D. Božić-Štulić, "Convolutional Neural Networks and Transfer Learning Based Classification of Natural Landscape Images," Journal of Universal Computer Science, vol. 26, pp. 244-267, 2020.

[11] X. Qi, P. Zhu, Y. Wang, L. Zhang, J. Peng, M. Wu, J. Chen, X. Zhao, N. Zang and P. Mathiopoulos, "MLRSNet: A Multi-label High Spatial Resolution Remote Sensing Dataset for Semantic Scene Understanding," 2021.

[12] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 60, 2019.

[13] K. Potdar, T. Pardawala and C. Pai, "A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers," International Journal of Computer Applications, vol. 175, pp. 7-9, 2017.

[14] R. Gemulla, "Sampling Algorithms for Evolving Datasets," 2008.

[15] V. Parsons, "Stratified Sampling," in Wiley StatsRef: Statistics Reference Online, Hyattsville, MD, USA, 2017.

[16] S. Palakodati, V. Ch, Y. Dasari and S. Bulla, "Fresh and Rotten Fruits Classification Using CNN and Transfer Learning," Revue d'Intelligence Artificielle, vol. 34, pp. 617-622, 2020.

[17] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti and D. De, "Fundamental Concepts of Convolutional Neural Network," in Recent Trends and Advances in Artificial Intelligence and Internet of Things, 2020, pp. 519-567.

[18] M. Kabir, "Convolutional Neural Network," 2021.

[19] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan, "Review of deep learning: concepts, {CNN} architectures, challenges, applications, future directions," Journal of Big Data, vol. 8, no. 1, p. 53, 2021.

[20] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010.

[21] C. Sharma, "Transfer Learning and its application in Computer Vision: A Review," in Transfer Learning and its application in Computer Vision, Waterloo, Canada, 2022.

[22] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, "A comprehensive survey on transfer learning," Proceedings of the IEEE, vol. 1, pp. 43-76, 2020.

[23] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[24] J. Liang, "Image classification based on RESNET," Journal of Physics: Conference Series, vol. 1634, 2020.
[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
[26] J. Cao, M. Yan, Y. Jia, X. Tian and Z. Zhang, "Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals," EURASIP Journal on Advances in Signal Processing, vol. 2021, 2021.