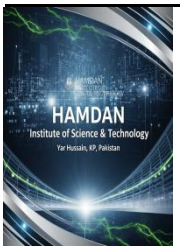

Journal of Fuzzy Intelligence

Journal homepage: www.mathfuzzyjournal.com

Vol:02 Issue: 01, 2026, pp.36-46

Efficient Outdoor Scene Classification Using MobileNetV3 and Transfer Learning

Chudary Akbar Ali[1,*], Peng Tao[1], and Safyan Jameel[1]

[1]School of Computer Science and Artificial Intelligence, Wuhan Textile University, Jiangxia District, Wuhan, 430200, China

[*]Corresponding author: cakbar036@gmail.com

Available online: 15 March 2026

ABSTRACT

Outdoor scene recognition (OSR) is a basic computer vision task that has been applied in autonomous systems, environmental knowledge, and in mobile vision. Nevertheless, a lot of current deep learning (DL) methods are based on architectures that are computationally-intensive and result in the designs of the systems being unable to be applied to resource-constrained hardware. The given paper is focused on an effective transfer learning (TL) based outdoor scene recognition system based on the MobileNetV3-Small architecture. The given approach consists in the use of the pretrained ImageNet features, where the feature adjustment is used to classify the outdoor environment incurring low computation costs. The experiments were performed on the MIT 8-Scene data, a scenery dataset of eight outdoor categories comprising 2,688 images, which were split on the training, validation and testing subsets. PyTorch (CPU-only) was used to train the model, and it showed that the model is practical in the low-power systems. The proposed method had the highest validation of 92.27% and the test accuracy of 88.51% respectively with a high level of accuracy, precision, and recall, and with the highest level of F1-score among the majority of the scene categories. The robustness and generalization ability of the model is confirmed through qualitative and quantitative measures of the model, such as the confusion matrix analysis and sample prediction. According to the results, the idea of lightweight architectures mixed with transfer learning should be viewed as a reliable method of outdoor scene recognition without having to use a high-end hardware, which makes the set of the suggested types of architecture the one that will be relevant to real-world and embedded uses.

Keywords

Outdoor Scene Recognition; Transfer Learning; MobileNetV3-Small ImageNet; MIT 8-Scene data; Computer Vision.

1. Introduction

Outdoor scene recognition is a significant issue in the field of computer vision which is aimed at recognizing and classifying outdoor settings like forests, streets, highways, mountains and coastal settings in the form of visual information [1]. Scene recognition is an important issue in numerous applications in reality: autonomous driving, robotic navigation, geographic information systems, smart surveillance, and a context-aware mobile application [2, 3]. Scene recognition, with the need to know global contextual and spatial attributes of a scenario, is therefore a difficult task when compared to object recognition, which is concerned with identifying particular objects in an image [4], because of the wide intra-class dissimilarities by the costs of inter-class likeness. The recent developments in deep learning especially CNNs have advanced the scene classification performance tremendously [5].

Deep architectures can learn natural hierarchical descriptions of features, which are able to learn both the textures at the low level, and the semantic organization at the high level. Nevertheless, most state of the art CNN models are computationally and power expensive and they need strong GPU hardware, and therefore, these models cannot be used in a resource-limited system like the mobile and embedded system. This poses a requirement of low weight and high performance models which can be able to produce competitive performance without a high level of computational cost. Transfer learning has turned out to be a practical solution to solve data scarcity and computing limitations through using the knowledge acquired on massive datasets like ImageNet. For fairly small data sets and based on limited hardware, high recognition accuracy is possible by fine-tuning pretrained models [6]. MobileNetV3-Small [7] is one of the architectures in the list of lightweight and specifically configured to have a balance between accuracy and efficiency, which is why it would be appropriate to use in real-time-based apps and low-performance devices. In this paper, we

suggest an effective outdoor scene recognition system that is built on the MobileNetV3-Small using transfer learning. The model is trained and tested on the MIT 8-Scene dataset in a CPU-only environment which shows its practicability in low-resource environments. The results of many experiments prove that the suggested methodology can be used to obtain a high classification rate in a variety of outdoor settings [8, 9]. The motivation behind this research stems from the growing demand for efficient scene recognition systems that can operate under limited computational resources. Many existing deep learning approaches rely on large and computationally intensive architectures that require high-performance GPUs for training and deployment. However, in many real-world applications such as mobile systems, embedded devices, and edge computing platforms, computational resources are limited. Therefore, this study explores the use of the lightweight MobileNetV3-Small architecture combined with transfer learning to achieve reliable outdoor scene classification while maintaining computational efficiency.

Outdoor scene recognition has been extensively studied in computer vision due to its importance in understanding environmental context. Early approaches relied on handcrafted features such as color histograms, texture descriptors, and edge-based representations combined with traditional classifiers [10]. While these methods provided initial insights into scene categorization, they struggled to capture high-level semantic information and were sensitive to variations in illumination, viewpoint, and background clutter. It has been found by Kumar et al. (2022) [11] that handcrafted features are significantly outperformed by CNN-based transfer learning models including ResNet and VGG, which include high-level semantics of space. Qi et al. (2023) [12] suggested a method of optimum multi-domain feature cross-scene corresponding to the multispectral data of UAV to enhance human detection rates in sophisticated outdoor settings. The study by Plantefol et al. (2024) [13] has applied the scene recognition method with CNN to immersive 360deg VR space so that it could implement an automated olfactory enhancement and improve the user experience. Zhang et al. (2024) [14] have presented a hierarchical learning generation cellular automata framework that can be used to complete large-scale outdoor scenes using sparse LiDAR data. Nagil and Mandal (2024) [15] have also created an energy-saving transformer-based assistive navigation system to be used by visually impaired people who use an entirety of edge devices in the outdoor setting. The Project SemCity suggested by Lee et al. (2024) [16] is based on a triplane architecture of the 3D diffusion model, which is used to complete, edit, and refine the large-scale real-world outdoor scenes. Nagrale and Khandelwal (2025) [17] have proposed the use of a multi recognizing scene architecture that integrates CNN, object-level and handcraft features to enhance accuracy of classification in indoor and outdoor settings.

The system proposed by Liu et al. (2025) [18] was called LightLoc and was a rapid-adaptation outdoor localization, which will freeze the feature backbones to permit it to learn the scene efficiently with a short retraining duration. Another method that could enhance background and foreground separation and stability in large scenes of the outdoor reconstruction is the two-shell Gaussian splatting proposed by Pintani et al. (2025) [19]. The 3D scene graphs of the outdoor 3D scenes with the inclusion of terrain semantics by Samuelson et al. (2025) [20] assist in hierarchical semantic mapping that ensures autonomous robotic navigation. Chen et al. (2025) [21] proposed MeSS which is a generative pipeline to generate realistic outdoor urban scenes with geometry-aware diffusion and Gaussian splatting. Samuelson et al. (2025) [22] introduced Terra, an extreme-scale metric-semantic system based on LiDAR and vision-language architecture to understand an outdoor scene. Ajaondo et al. (2018) [23] have suggested a place recognition system based on scene-graphs of an outdoor place with the use of OpenStreetMap data and natural language descriptions. The FIORD dataset proposed by Gunes et al. (2025) [24] can be used in the reconstruction of wide field-of-view outdoor scenes with the assistance of fiskeye cameras and LiDARs. Szankin et al. (2025) [25] tested vision-language and OCR models in harsh outdoor conditions and found that it works in more pure lightweight CNNs because of the ability to deploy them on edges. Recent research has increasingly focused on improving scene understanding using lightweight deep learning architectures and efficient feature representations. Several studies have explored the integration of transfer learning with compact CNN models to balance classification performance and computational efficiency [26]. Additionally, advancements in generative modeling, transformer-based perception systems, and large-scale scene understanding frameworks have expanded the capabilities of outdoor scene analysis. Despite these developments, there remains a need for efficient classification-based approaches that can operate under constrained computational environments while maintaining reliable recognition performance. This gap motivates the investigation of lightweight architectures such as MobileNetV3 for outdoor scene classification tasks [27].

The principal findings of this work are threefold: (i) a lightweight outdoor scene classification model, appropriate to operate on the resources-constrained system, developed, (ii) appropriate evaluation using a variety of performance indicators and visual inspection, and (iii) the empirical validity that the transfer learning may be used to perform reliable scene recognition without the use of high-end computational resources.

Even though there have been dramatic advances on understanding the outdoor scenes, there are gaps in research that are reflected through the available literature. The main contributions of this research are summarized as follows. First, this study proposes an efficient outdoor scene recognition framework based on the lightweight MobileNetV3-Small architecture combined with transfer learning. Second, the proposed system demonstrates that high-level scene classification performance can be achieved using a computationally efficient model operating in a CPU-only

environment. Third, the framework is evaluated on the MIT 8-Scene dataset with comprehensive performance metrics including accuracy, precision, recall, F1-score, and confusion matrix analysis. Finally, the study provides an analysis of lightweight deep learning models for outdoor scene recognition, highlighting their suitability for deployment in resource-constrained systems. Lastly, the absence of experimental confirmation as to how high performance in recognition can be attained by not using GPUs but moderate size datasets is lacking and this is what the study will answer.

Table 1 presents the list of acronyms used throughout the paper. These abbreviations are provided to enhance readability and help readers easily understand the technical terms and methods discussed in the study.

Table 1: List of Acronyms Used in the Paper

Acronym	Full Form
CNN	Convolutional Neural Network
DL	Deep Learning
TL	Transfer Learning
MIT	Massachusetts Institute of Technology
CPU	Central Processing Unit
GPU	Graphics Processing Unit
F1	F1 Score
RGB	red, green and blue
LiDAR	Light Detection and Ranging
OSR	Outdoor Scene Recognition
SE	Squeeze-and-Excitation

2. *Methodology*

The section tells about the approach taken to create a framework of efficient outdoor scene recognition with the help of a scalable light model of deep learning. The suggested framework includes the dataset preparation, transfer learning, model training, and performance estimation to work with reliable classification given natural computational resources. This study assumes that outdoor scenes can be effectively categorized based on global visual features extracted from RGB images. The approach also assumes that pretrained convolutional neural networks trained on large-scale datasets such as ImageNet contain transferable visual representations that can be adapted to scene classification tasks. Furthermore, the proposed framework assumes that a lightweight architecture such as MobileNetV3-Small is sufficient to capture discriminative scene features without requiring deep and computationally intensive networks. These assumptions allow the proposed method to focus on achieving efficient and accurate scene classification under constrained computational environments.

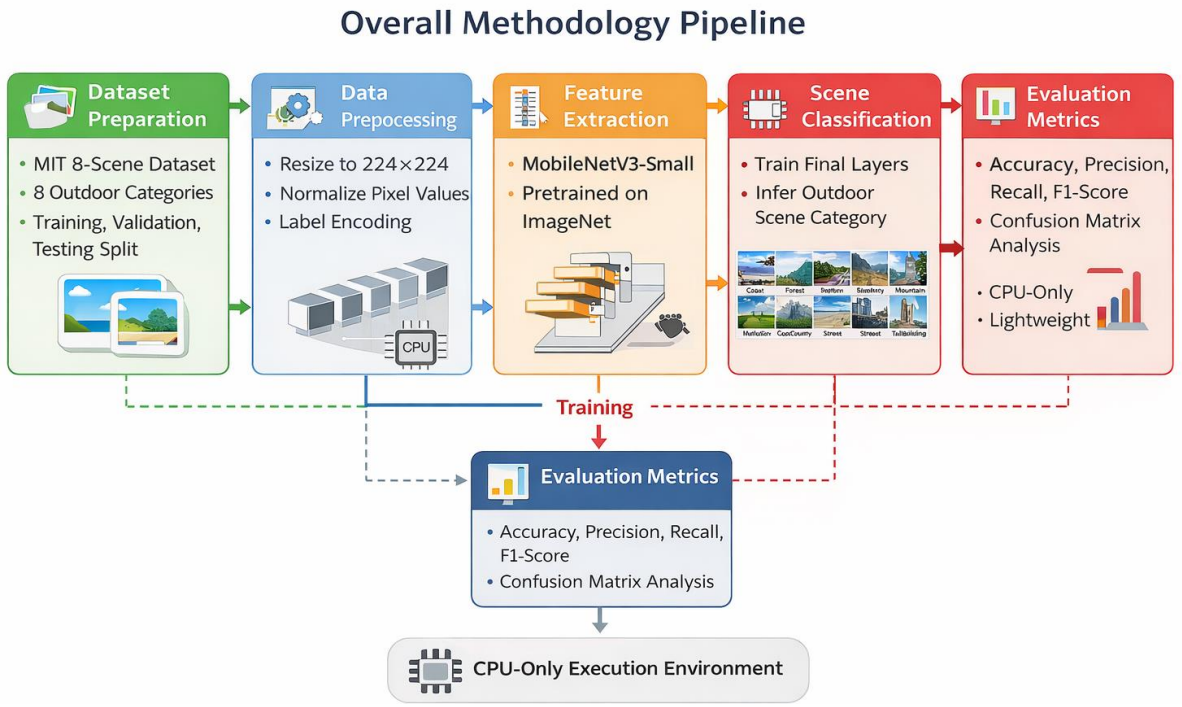

Figure 1: The overall methodology pipeline of the proposed research work

Figure 1 shows the entire process of the suggested outdoor scene recognition algorithm. Data preparation starts with the preparation of the dataset based on the MIT 8-Scene dataset, the next steps are image preprocessing, i.e., resizing the dataset, normalization, and label encoding. The processed images undergo the MobileNetV3-Small model that is trained on ImageNet which is a lightweight feature extractor. TL is implemented where only the last classification layers are trained to map features that are extracted to categories of outdoor scenes. The last measure to assess the model performance is with accuracy, precision, recall, F1-score, and confusion matrix analysis. The complete pipeline is set to be efficient at the use of the CPU environment, which indicates the appropriateness of the offered solution to the resource-limited systems.

2.1. Dataset Description

The MIT 8-Scene dataset used in this study contains a total of 2,688 images across eight outdoor scene categories: coast, forest, highway, inside city, mountain, open country, street, and tall building. To ensure reliable evaluation, the dataset was divided into training, validation, and testing sets using a 70% Training – 15% Validation – 15% Testing stratified split. This resulted in 1,878 training images, 401 validation images, and 409 testing images. Stratified sampling was used to maintain proportional representation of each class across all subsets, ensuring balanced learning and fair evaluation of model performance.

2.2. Data Preprocessing

Each image had been downsized to 224 x 224 pixels to fit the input size of MobileNetV3-Small. The values provided by ImageNet as the mean and standard deviation of pixels were used to normalize the pixel intensities, which ensured the consistency of the pixel intensities with the features extractor that was pretrained. Preprocessing was done through class labels assigned by default depending on the directory structure and data loading through PyTorch data loaders were used to make the use of batch processing effective. There was no aggression in data augmentation because of the constraints in computational resources and this study was in essence aimed at determining the transfer learning effectiveness with little resources.

2.3. Model Architecture

The choice of the backbone network was MobileNetV3-Small because it offers a proper balance between the precision and the computational efficiency. The architecture uses depthwise-separable convolution, inverted residual blocks, squeeze-and-excitation (SE) block and optimized nonlinear activation functions to minimize the number of parameters used without affecting representational capacity. The last classification layer of the trained model was changed to a new fully connected layer to have eight outdoor scene categories.

MobileNetV3-Small was selected due to its favorable balance between computational efficiency and classification performance. Compared with earlier architectures such as MobileNetV2, MobileNetV3 introduces architectural improvements including squeeze-and-excitation modules and optimized activation functions that enhance feature representation while maintaining low computational cost. Although MobileNetV3-Large and EfficientNet models may achieve slightly higher accuracy, they require greater computational resources and memory. Since this study focuses on lightweight deployment in CPU-only environments, MobileNetV3-Small provides an appropriate trade-off between efficiency and recognition performance.

2.4. Transfer Learning Strategy

The model used was the Transfer learning, where MobileNetV3-Small was initialized using ImageNet-pretrained weights. It was in the feature extraction layers which were frozen to maintain previously learned generic visual features whereas only the last classification layer was trained on the MIT 8-Scene dataset. It resulted in shorter training time, less overfitting and made a successful learning with the relatively small databases and without the use of GPUs or full desktop processors [28, 29].

Figure 2 shows the transfer learning approach that will be used in this research, where a convolutional neural network that has been trained on a sizeable source data is accessible on a new learning task. At the proposed framework, the convolutional layers are trained on generic features of the visuals of the source domain and are transferred to the target task. The pretrained feature extraction layers are saved in order to maintain high level of spatial representation whereas the last fully connected layers are substituted and retrained on the task specific data. The source model in the context of this study is the MobileNetV3-Small model pretrained on ImageNet and the layers of the classification network are however finetuned to take in the categories of the outdoor scenes in the MIT 8-Scene dataset. The technique facilitates an efficient way of learning and saves on training time and enhances the performance of generalization and especially in the limited computer resources availability.

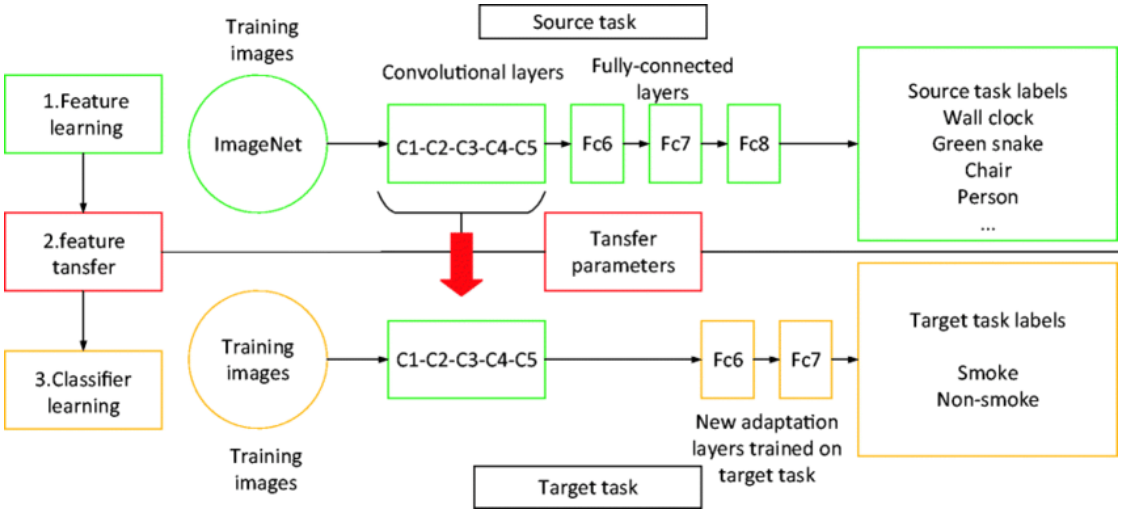

Figure 2: Transfer Learning Flowchart

2.5. Training Configuration

The model was trained using the PyTorch framework for 20 epochs with a batch size of 16. The Adam optimizer was employed with a learning rate of 1×10^{-4} to ensure stable convergence. The model with the best validation accuracy was saved and later used for final testing. The categorical cross-entropy loss function was used to optimize multi-class classification performance. During training, the feature extraction layers of the pretrained network were frozen, and only the final classification layer was fine-tuned on the MIT 8-Scene dataset. This strategy reduced training time and helped prevent overfitting.

2.6. Evaluation Metrics

1. Accuracy: Accuracy measures the overall correctness of the classifier:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{1}$$

Using confusion matrix terms:

$$\text{Accuracy} = \frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C}(TP_c+FP_c+FN_c+TN_c)} \tag{2}$$

Where C is the total number of outdoor scene recognition categories (8 categories are used in this research).

2. Precision: Precision measures how many images predicted as class $c$were actually correct.

$$\text{Precision}_c = \frac{TP_c}{TP_c+FP_c} \tag{3}$$

If precision is high, the model makes few false alarms for class c.

3. Recall: Recall measures how many actual images of class $c$the model successfully detected.

$$\text{Recall}_c = \frac{TP_c}{TP_c+FN_c} \tag{4}$$

High recall means the model rarely misses the scenes belonging to class c.

4. F1-Score: The F1-score is the harmonic mean of precision and recall, providing a balance between them:

$$\text{F1}_c = 2 \times \frac{\text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{5}$$

F1-score is especially useful when classes have imbalanced sample counts, as it penalizes both false positives and false negatives.

3. Experiments and Results

This section presents the experimental setup and evaluates the performance of the proposed outdoor scene recognition framework based on MobileNetV3-Small with transfer learning. All experiments were conducted using the MIT 8-Scene dataset in a CPU-only environment to demonstrate the feasibility of lightweight models under limited computational resources [30, 31].

3.1. Experimental Setup

The model was implemented using Python and the PyTorch deep learning framework. MobileNetV3-Small [32] pretrained on ImageNet was used as the base network, with its final classification layer replaced to support eight outdoor scene categories. The dataset consisted of 2,688 images and was divided into training, validation, and testing sets with 1,878, 401, and 409 images respectively. The model was trained for 20 epochs using the Adam optimizer with a learning

rate of $1 \times 10^{-4}$and a batch size of 16. During training, only the final classification layer was updated, while the pretrained backbone remained frozen.

3.2. Performance Evaluation

The proposed model achieved a best validation accuracy of 92.27% and a final test accuracy of 88.51%, indicating strong generalization despite the limited dataset size and absence of GPU acceleration. Precision, recall, and F1-score were computed for each scene category to provide a detailed class-wise performance analysis. High F1-scores were observed for structurally distinctive categories such as *tall building*, *forest*, and *highway*, demonstrating the model's ability to capture discriminative spatial and contextual features. Classes with visually similar natural characteristics, such as *mountain* and *open country*, exhibited slightly lower scores, reflecting the inherent ambiguity present in outdoor scenes.

Table 2: Class-Wise Precision, Recall, F1-Score, and Support for MIT 8 Outdoor Scene Recognition

Class	Precision	Recall	F1-Score	Support
coast	0.8308	0.9818	0.9	55
forest	0.902	0.92	0.9109	50
highway	0.9706	0.8462	0.9041	39
inside city	0.8696	0.8511	0.8602	47
mountain	0.9245	0.8596	0.899	57
open country	0.8276	0.7742	0.8	62
street	0.8723	0.9111	0.8913	45
tall building	0.9273	0.9444	0.9358	54

A more detailed analysis of the classification metrics of the various categories of outdoor scenes is given in Table 2 and in this way further gives a view on how the model was consistent on classification on an individual category of the outdoor scenes. The tall building, highway and forest scene types exhibit some of the best precision and F1-scores which implies a high ability to extract discriminative features when the classification involves a set of patterns that can be identified. On the other hand, such categories as open country and mountain have a little lower F1-scores because there is more visual similarity in them especially in vegetation, skyline and texture of the terrain that may lead to the model misclassifying one category as another. During the experiment on recalling coast and street, the model exhibited high levels of recall attempting to identify these features correctly when they occur, which indicates there is uniformity in the learned spatial and structural features including open shoreline vistas and urban road systems. The support values ensure a reasonable representation among the classes and confirm that the scores that are observed are not biased by small samples. Together, Table 2 findings support the fact that MobileNetV3-Small will be able to discriminate complex real-world settings at a trusted accuracy based on its operation under the CPU-based transfer learning conditions.

The confusion matrix analysis further revealed that most predictions were concentrated along the diagonal, confirming reliable class separation. Misclassifications primarily occurred between semantically overlapping categories, rather than due to systematic model failure. Visual inspection of sample predictions showed that the model maintained high confidence across diverse lighting conditions and environmental variations, reinforcing its robustness.

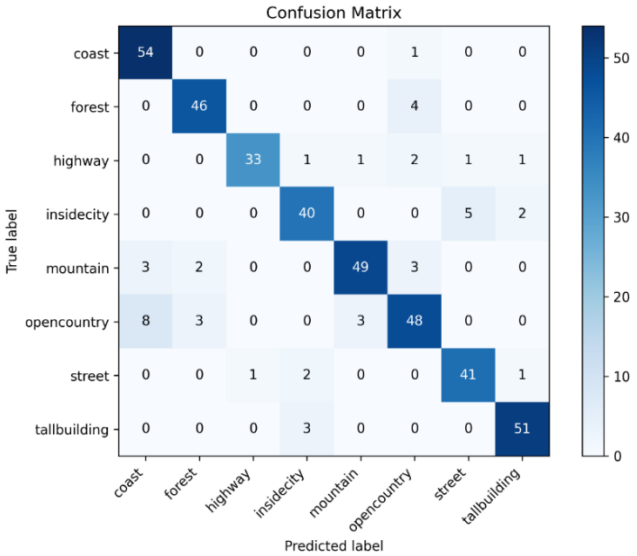

Figure 3: Confusion Matrix for Outdoor Scene Recognition

Figure 3 shows the accuracy and inaccuracy of the predictions of the MobileNetV3-Small model on the eight categories of the outdoor scenes. The darker diagonal values indicate a high recognition accuracy of a particular class which proves that a majority of the categories were correctly identified by the classifier. Limited misclassification between environmentally related situations that are similar in visual appearance (e.g., mountain and open country) and between indoor situated city and street, which are structurally similar in terms of their textures and space organization, are also indicated in the matrix. This observation is a natural phenomenon known as semantic overlap as opposed to algorithmic weakness and it represents an indication that the classifier learned to discriminate particular environmental features despite intra-class similarities. All in all, the confusion and the validation that there is good separation of the classes in the confusion matrix show that the model can be successfully used in the task of robust outdoor landscapes perception.
Overall, the experimental results confirm that combining MobileNetV3-Small with transfer learning provides an effective balance between accuracy and computational efficiency. The findings demonstrate that reliable outdoor scene recognition can be achieved using lightweight architectures, making the proposed approach suitable for deployment in real-world and resource-constrained applications.

Table 3: Comparison with Lightweight CNN Models

Model	Architecture Type	Accuracy (%)	Model Complexity
MobileNetV2	Lightweight CNN	86.4	Low
ShuffleNet	Lightweight CNN	84.9	Very Low
EfficientNet-Lite	Efficient CNN	89.2	Medium
Proposed Method (MobileNetV3-Small)	Lightweight CNN	92.27 (Validation)	Low

Table 3 compares the proposed MobileNetV3-Small framework with several lightweight convolutional neural network architectures reported in the literature. The results indicate that the proposed method achieves competitive performance while maintaining low computational complexity, demonstrating the effectiveness of MobileNetV3-Small combined with transfer learning for outdoor scene recognition.

4. Discussion and Limitations

The experimental findings show that the proposed MobileNetV3-Small-based framework with the combination of transfer learning is the one that offers a considerable balance between recognition accuracy and types of computations needed to classify the outdoor scene. The attained validation accuracy of 92.27% and test accuracy of 88.51 are proof of the fact that trained on fairly small datasets and operating in CPU-only environments, pretrained lightweight architectures are capable of effectively capturing high-level semantic information in outdoor settings. The high score achieved in such classes as tall building, forest, and highway show that such model is capable of acquiring a number of peculiarities of structure and context, whereas the possibility of the confusion between visually similar things like mountain and open country depicts the neutrality inherent in the generated scenes rather than the lack thereof in the learning model.

Although some existing studies report higher accuracy using complex architectures and multimodal pipelines, the proposed method achieves competitive performance while operating in a CPU-only environment using a lightweight MobileNetV3-Small architecture. These results support the appropriateness of the MobileNet-grounded models to useful application in the resource-limited systems, such as mobile and embedded systems. Although such positive outcomes are present, there are a number of limitations to consider. To begin with, the sample size is also quite small, limiting it to drastic weather, light, season and geographic differences. Second, the last classification layer of the pretrained model was fine-tuned, which is, a more semantically specific feature representation was not ultimately expanded to the semantics of outdoor scenes. Although this option was made strategically so as to keep up with the computational efficiency, it can somewhat restrict the enhanced performance of already visually overlapping classes. The research also is limited solely to RGB image-based classification without inclusion of multimodal inputs, including depth, LiDAR, or multispectral data which may be used to provide a deeper insight into the scene in a complex setting. The mitigation of these constraints in future employment can result in a better generalization as well as greater generalization of lightweight outdoor scene recognition systems.

The lightweight design of MobileNetV3-Small enables efficient execution under limited computational resources. The model contains approximately 2.5 million parameters and requires significantly fewer floating-point operations compared with larger CNN architectures. This efficiency makes the proposed framework suitable for deployment on resource-constrained devices. The proposed approach offers several advantages compared with traditional deep learning frameworks for scene recognition. First, the use of MobileNetV3-Small significantly reduces

computational complexity while maintaining competitive classification accuracy. Second, the transfer learning strategy allows the model to leverage pretrained ImageNet features, enabling effective learning even with a moderate-sized dataset. Third, the framework operates successfully in a CPU-only environment, demonstrating its suitability for resource-constrained systems such as embedded devices and edge computing platforms. Finally, the model achieves strong classification performance across multiple outdoor scene categories, confirming the effectiveness of lightweight architectures for real-world scene understanding tasks.

Table 4: Comparison of Proposed Method with Existing Works

Author & Year	Task	Model / Approach	Dataset	Hardware / Complexity	Reported Performance	Key Difference from Our Work
Kumar et al. (2022)	Indoor & Outdoor Scene Classification	ResNet / VGG (Transfer Learning)	Custom scene datasets	Deep CNNs, GPU-oriented	High accuracy (not CPU-focused)	Uses heavy architectures; not optimized for low-resource systems
Qi et al. (2023)	Human detection in outdoor scenes	Multispectral CNN (UAV)	UAV multispectral data	Complex multi-domain model	91.55% accuracy	Focuses on detection, not scene classification
Plantefol et al. (2024)	Scene recognition in VR	ResNet, DenseNet	360° VR scenes	Computationally intensive	~71% accuracy	Targets immersive VR, not real-world outdoor classification
Zhang et al. (2024)	Outdoor scene completion	Generative Cellular Automata	LiDAR scenes	High computational cost	Not classification-based	Focuses on 3D reconstruction, not categorization
Nagil & Mandal (2024)	Outdoor navigation assistance	Transformer-based vision	Real outdoor data	Edge device but heavy model	Task-specific metrics	Focuses on segmentation, not scene recognition
Nagrale & Khandelwal (2025)	Scene classification	Multi-feature CNN + YOLO	Mixed indoor/outdoor	High model complexity	91.84% accuracy	Uses complex multi-model fusion
Our Work	OCR	MobileNetV3-Small + Transfer Learning	MIT 8-Scene	CPU-only, lightweight	92.27% (Val), 88.51% (Test)	Efficient, lightweight, high accuracy under constrained hardware

Table 5: Computational Efficiency of the Proposed Model

Metric	Value
Model Architecture	MobileNetV3-Small
Parameters	~2.5 Million
Model Size	~9 MB
FLOPs	~66M
Hardware	CPU-only
Training Time	~11 minutes

5. Conclusions

The present paper offered an effective and feasible method of the recognition of an outdoor scene through the concept of transfer learning based on the MobileNetV3-Small architecture. The main incentive of this project was to explore the possibility whether under constrained computer resources, a lightweight convolutional neural network with the aid of pretrained knowledge, can be used to make well-grounded scene classification results. In comparison to most of the available literature where practitioners apply deep and computationally complex frameworks, the proposed framework was specifically designed and tested within a CPU-only environment with consideration of the realistic capabilities that are typically faced in mobile and embedded architectures. As the experimental appraisal carried out on the MIT 8-Scene dataset proves, MobileNetV3-Small can be trained on the relevant and discriminative representations of the outdoor settings. The model was able to provide the best validation accuracy of 92.27% and an overall test accuracy of 88.51% with corresponding high values of precision, recall, and F1-score on majority of scene categories. These findings prove the importance of transfer learning to facilitate effective learning using a moderate sized data using a reuse of high-level visual features learned using large-scaled datasets like ImageNet. The analysis of the performance of the classes moreover demonstrated that the scenes with unique structural patterns, e.g. tall buildings, highways, and forests, have been recognized with high reliability, and there was some slight confusion between similar natural classification like mountain and open country. The occurrence of such confusion represents the ambiguity of real images of nature, instead of an essential weakness of the proposed approach. In addition to the quantitative accuracy, qualitative analysis using confusion matrices and sample predictions revealed that the model was able to retain constant confidence in different lighting conditions, viewpoint and layout of the environment. This strength indicates the applicability of MobileNetV3-Small in practical scenarios in the context of outdoor vision, where appearances vary.

The network was designed to be lightweight with sheets of frozen features extraction and little fine-tuning which also enabled the model to learn fast without overfitting without the use of any accelerated computer such as the GPU. Therefore, the suggested framework has an attractive performance versus computation efficiency trade-off. On the whole, this research confirms the fact that not only the large-scale or computationally expensive architecture can be used to achieve high performance in the outdoor scene recognition. In its turn, the choice of lightweight models carefully and closely accompanied by the transfer learning can provide the list of competitive models and at the same time be allowed to be deployed to the low-power devices. The research results presented in this paper will add to the current study of effective scene understanding and serve as a firm basis of future projects aimed at real-time application, edge computing, and intelligent systems which need to identify the surrounding world with a high degree of reliability due to the restriction of resources. Despite achieving promising performance, several limitations should be acknowledged. First, the study evaluates the model using a moderate-sized dataset, which may not fully capture the diversity of outdoor environments found in real-world scenarios. Second, the experiments were conducted in a CPU-only environment without testing on dedicated embedded hardware platforms. Future work will focus on evaluating the proposed framework on larger and more diverse scene datasets, as well as deploying the model on embedded systems and edge devices to further validate its efficiency in practical applications.

Future work will be focusing on extending the proposed framework by exploring more advanced lightweight architectures and incorporating data augmentation techniques to further improve classification performance. In addition, future research may investigate the integration of multimodal data sources such as depth, LiDAR, or multispectral imagery to enhance scene understanding in complex outdoor environments. Another potential direction involves optimizing the model for real-time deployment on edge devices and mobile platforms. Furthermore, expanding the evaluation to larger and more diverse scene datasets could provide deeper insights into the generalization capability of lightweight deep learning models for outdoor scene recognition.

***Author contributions**:* All authors equally contributed to this article.

Funding Information: Funding information is not available.

Data Availability: All data generated or analyzed during this study are included in this published article.

References

[1] Xie, L., Lee, F., Liu, L., Kotani, K., & Chen, Q. (2020). Scene recognition: A comprehensive survey. Pattern Recognition, 102, 107205.

[2] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. Advances in neural information processing systems, 27.

[3] Xu, G., Khan, A. S., Moshayedi, A. J., Zhang, X., & Shuxin, Y. (2022). The object detection, perspective and obstacles in robotic: a review. EAI Endorsed Transactions on AI and Robotics, 1(1), e13.

[4] Khan, A. S., Rahman, A., & Moshayedi, A. J. (2025). Pipeline Surface Defect Detection Using YOLOv11 with Attention Mechanisms: A Comparative Study of SA, LKA, and CBAM Approaches. Journal of Robotics Research (JRR), 2(2).

[5] Cheng, X., Lu, J., Feng, J., Yuan, B., & Zhou, J. (2018). Scene recognition with objectness. Pattern Recognition, 74, 474-487.

[6] López-Cifuentes, A., Escudero-Vinolo, M., Bescós, J., & García-Martín, Á. (2020). Semantic-aware scene recognition. Pattern Recognition, 102, 107256.

[7] Cao, Z., Li, J., Fang, L., Li, Z., Yang, H., & Dong, G. (2025). Research on efficient classification algorithm for coal and gangue based on improved MobilenetV3-small. International Journal of Coal Preparation and Utilization, 45(2), 437-462.

[8] Zhu, J., Zhang, C., & Zhang, C. (2023). Papaver somniferum and Papaver rhoeas classification based on visible capsule images using a modified MobileNetV3-small network with transfer learning. Entropy, 25(3), 447.

[9] Abosuliman, S. S., Rahman, I. U., Abdullah, S., & Qadir, A. (2024). Selection of third-party logistics in supply chain finance under probabilistic complex hesitant fuzzy sets and distance measures. Heliyon, 10(17).

[10] Ma, Z., He, J., Lin, Q., Chen, K., Jia, Y., & Zhou, B. (2025, July). Multi-Modal Perception-Based Indoor and Outdoor Scene Modeling and Segmentation. In 2025 44th Chinese Control Conference (CCC) (pp. 8175-8180). IEEE.

[11] Kumar, N., Singh, H., Varshney, M. T., Malik, M. V., & Kumar, V. (2022). Indoor and Outdoor Scene Recognition. Grenze International Journal of Engineering & Technology (GIJET), 8(2).

[12] Qi, F., Xia, J., Zhu, M., Jing, Y., Zhang, L., Li, Z., ... & Lu, G. (2023). UAV multispectral multi-domain feature optimization for the air-to-ground recognition of outdoor injured human targets under cross-scene environment. Frontiers in Public Health, 11, 999378.

[13] Plantefol, T., Simiscuka, A. A., Yaqoob, A., & Muntean, G. M. (2025). CNN-based 360 scene recognition for automatic generation of omnidirectional scent effects. IEEE Transactions on Multimedia.pp. 29-41

[14] Zhang, D., Williams, F., Gojcic, Z., Kreis, K., Fidler, S., Kim, Y. M., & Kar, A. (2024). Outdoor scene extrapolation with hierarchical generative cellular automata. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20145-20154).

[15] Nagil, P., & Mandal, S. K. (2024, August). DISHA: Low-Energy Sparse Transformer at Edge for Outdoor Navigation for the Visually Impaired Individuals. In Proceedings of the 29th ACM/IEEE International Symposium on Low Power Electronics and Design (pp. 1-6).

[16] Lee, J., Lee, S., Jo, C., Im, W., Seon, J., & Yoon, S. E. (2024). Semcity: Semantic scene generation with triplane diffusion. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 28337-28347).

[17] NAGRALE, P., & KHANDELWAL, S. (2025). Indoor-outdoor scene recognition: A multi-feature framework using CNN for complex environment. Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi, 43(4).

[18] Liu, Y., Li, X., Zhang, Y., Qi, L., Li, X., Wang, W., ... & Yang, M. H. (2025). Controllable 3D outdoor scene generation via scene graphs. arXiv preprint arXiv:2503.07152.

[19] Pintani, D., Caputo, A., Lewis, N., Stamminger, M., Pellacini, F., & Giachetti, A. (2025). Two-Stage Gaussian Splatting Optimization for Outdoor Scene Reconstruction. arXiv preprint arXiv:2510.09489.

[20] Samuelson, C. R., McLain, T. W., & Mangelson, J. G. (2025). Towards Terrain-Aware Task-Driven 3D Scene Graph Generation in Outdoor Environments. arXiv preprint arXiv:2506.06562.

[21] Chen, X., Zhai, Z., Zhou, K., Wang, Z., He, J., Wang, D., ... & Meng, L. (2025). MeSS: City Mesh-Guided Outdoor Scene Generation with Cross-View Consistent Diffusion. arXiv preprint arXiv:2508.15169.

[22] Samuelson, C. R., Austin, A., Knoop, S., Romrell, B., Slade, G. R., McLain, T. W., & Mangelson, J. G. (2025). Terra: Hierarchical Terrain-Aware 3D Scene Graph for Task-Agnostic Outdoor Mapping. arXiv preprint arXiv:2509.19579.

[23] Jung, D., Kim, K., & Kim, S. W. (2025). GOTPR: General Outdoor Text-based Place Recognition Using Scene Graph Retrieval with OpenStreetMap. IEEE Robotics and Automation Letters.

[24] Gunes, U., Turkulainen, M., Ren, X., Solin, A., Kannala, J., & Rahtu, E. (2025, June). FIORD: A Fisheye Indoor-Outdoor Dataset with LIDAR Ground Truth for 3D Scene Reconstruction and Benchmarking. In Scandinavian Conference on Image Analysis (pp. 3-17). Cham: Springer Nature Switzerland.

[25] Szankin, M., Venkatasamy, V. R., & Ying, L. (2025). Seeing the Signs: A Survey of Edge-Deployable OCR Models for Billboard Visibility Analysis. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5992-6000).

[26] Moshayedi, A. J., Khan, A. S., Jawadi, M. J., Kolahdooz, A., & Emadi Andani, M. (2026). Ergonomically Designed Assistive Robots: Where and How to Bring Comfort, Safety, and Independence to Elderly Care. Journal of Field Robotics.

[27] Ji, H., Mendonça, I., & Aritsugi, M. (2025). Multi-Scene Dataset and Object Detector for Outside Blind Individual Identification. IEEE Access, 14, 1423-1438.

[28] Shoaib, M., Nawaz, A., Khan, A. A., Sulaiman, M., Amjad, M., Khan, A. S., & Saleh, H. (2023). Paper ID: ICSET-2307 Design and Sustainable Fabrication of a PVC Wind Turbine for Clean Energy Generation.

[29] Nawaz, A., Sulaiman, M., Rehman, Z. U., Khan, A. S., Rasheed, Z., Saleh, H., & Shoaib, M. (2023). Paper ID: ICSET-2306 INTELLIGENT DEMAND-SIDE MANAGEMENT FOR INTEGRATED RENEWABLE ENERGY IN RESIDENTIAL AREA SMART GRID: A PATHWAY TO SUSTAINABLE ENERGY USE

[30] Moshayedi, A. J., Nasab, S. T. M., Khan, Z. H., & Khan, A. S. (2024). Meta-heuristic Algorithms as an Optimizer: Prospects and Challenges (Part II). Engineering Applications of AI and Swarm Intelligence, 155-180.

[31] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1452-1464.

[32] Moshayedi, A. J., Nasab, S. T. M., Khan, Z. H., & Khan, A. S. (2024). Meta-heuristic Algorithms as an Optimizer: Prospects and Challenges (Part I). Engineering Applications of AI and Swarm Intelligence, 131-154.