Deep Learning-Based Multi-Class Image Classification in University Campuses Using EfficientNet

OLA ALKURDI[1], RAHAF HASSONAH[2], RUA'A ALKURDI[3], AND NOOR ALMAZAYDEH[4]

[1] *Department of Artificial intelligence, University of Jordan, Amman, Jordan*

Compiled January 3, 2026

Image classification in complex real-world environments remains a challenging task due to variations in lighting conditions, background clutter, and object appearance. University campuses represent dynamic environments that contain diverse visual elements such as buildings, people, laboratories, trees, and vehicles. This paper presents a deep learning-based approach for automated multi-class image classification in university campus environments using transfer learning. A dataset consisting of 2,370 images collected from a real campus setting is used to train and evaluate the proposed system across five classes. EfficientNetB0, pre-trained on ImageNet, is employed as the backbone model and fine-tuned to adapt to the target classification task. To enhance performance and robustness, an enhanced model configuration is introduced by integrating squeeze-and-excitation attention mechanisms, batch normalization, dropout regularization, and class weighting strategies. The proposed approach is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis. Experimental results demonstrate that the enhanced EfficientNetB0 model achieves a high test accuracy of 98.32%, outperforming the baseline model and ResNet50. The findings confirm the effectiveness of EfficientNet-based transfer learning for accurate and reliable campus image classification and highlight its potential applicability in real-world intelligent monitoring and smart campus systems.

Keywords: Deep Learning; Transfer Learning; Image Classification; Convolutional Neural Networks; EfficientNetB0

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Images captured in university campus environments contain rich visual information related to daily activities, infrastructure, and surrounding natural elements. Modern campuses include a wide variety of visual objects such as trees, people, buildings, laboratories, and cars, all of which play an important role in campus management, safety monitoring, and smart infrastructure planning. With the rapid growth of digital imaging technologies and the increasing availability of large image collections, the automatic classification of campus images has become an important problem in the field of computer vision.

Traditional manual classification of images is a time-consuming and inefficient process that relies heavily on human effort. In large-scale environments such as university campuses, manually organizing images into predefined categories is impractical and prone to errors and inconsistencies. Moreover, variations in lighting conditions, background complexity, and object appearance further increase the difficulty of achieving accurate visual classification. These challenges highlight the need for automated, reliable, and intelligent image classification systems capable of handling real-world visual variability.

Recent advances in machine learning, particularly deep learning, have significantly improved the performance of image classification systems. Convolutional Neural Networks (CNNs) have demonstrated exceptional capability in learning hierarchical and discriminative visual features directly from images. Since their breakthrough success in 2012, numerous CNN architectures have been proposed and successfully applied to a wide range of visual recognition tasks. However, training deep CNN models from scratch typically requires very large labeled datasets and substantial computational resources, which are often unavailable in practical academic or real-world scenarios.

To address these limitations, transfer learning using pre-trained CNN models has emerged as an effective and widely adopted solution. Pre-trained models leverage knowledge learned from large-scale datasets and adapt it to new tasks with limited data. Architectures such as ResNet50 have been commonly used as strong baseline models for image classification. Nevertheless, preliminary experiments conducted in this work indicated that ResNet50 may face limitations in capturing fine-grained visual details within campus scenes. In contrast, EfficientNet introduces a compound scaling strategy that jointly optimizes network depth, width, and input resolution, leading to improved feature representation and computational efficiency. Based on these advantages, EfficientNetB0 was selected as the primary model for this study.

This paper aims to develop an automated image classification system for a university campus environment based on deep learning techniques. The proposed approach utilizes a dataset comprising five campus image classes: trees, people, buildings, laboratories, and cars. EfficientNetB0, pre-trained on a large-scale image dataset, is fine-tuned to address the multi-class classification task. Image preprocessing methods are employed to improve data quality and enhance model performance. The effectiveness of the proposed system is evaluated using standard classification metrics.

The main contributions of this study can be summarized as follows:

- A multi-class image dataset representing five common campus categories is collected from a real university environment.
- Image preprocessing techniques are applied to improve data quality and support effective model training.
- A comparative analysis between ResNet50 and EfficientNetB0 is conducted to evaluate classification performance.
- EfficientNetB0 is adopted to achieve improved feature representation and higher classification accuracy for campus image classification

2. RELATED WORKS

Image classification has been extensively studied in the field of computer vision, particularly with the rapid advancement of deep learning techniques. Earlier image classification approaches relied on handcrafted feature extraction methods combined with traditional machine learning classifiers. However, these approaches were highly dependent on feature engineering and often failed to generalize effectively to complex visual environments, motivating the adoption of deep learning-based solutions [1].

With the emergence of deep learning, Convolutional Neural Networks (CNNs) have become the dominant approach for image classification due to their ability to automatically learn hierarchical feature representations from raw image data. CNN-based models have achieved significant performance improvements across several benchmark datasets such as ImageNet, Caltech101, and Caltech256. Despite these successes, training deep CNN architectures from scratch requires large-scale labeled datasets and substantial computational resources, which limits their applicability in many real-world scenarios [2].

To address these limitations, transfer learning has been widely adopted in image classification research. Transfer learning enables the reuse of models pre-trained on large datasets such as ImageNet for new tasks with limited data. Umeaduma. conducted a comprehensive survey of popular pre-trained CNN architectures, including ResNet, EfficientNet, DenseNet, VGG, MobileNet, and ConvNeXt, and evaluated their performance on Caltech101 and Caltech256 datasets. Their study demonstrated that pre-trained models significantly reduce training time while maintaining high classification accuracy, confirming the effectiveness of transfer learning for image classification tasks [2].

Among pre-trained CNN architectures, ResNet introduced residual connections to mitigate the vanishing gradient problem and enabled the training of very deep networks. While ResNet-based models have shown strong performance across various image classification tasks, several studies indicate that they may be less efficient in capturing fine-grained visual details when computational resources or input resolution are constrained. In contrast, EfficientNet proposes a compound scaling strategy that jointly scales network depth, width, and input resolution, resulting in improved parameter efficiency and higher classification accuracy [3].

Recent research has highlighted EfficientNet as a highly effective architecture for transfer learning-based multi-class image classification. Studies applying EfficientNet models to complex classification tasks, such as waste image classification, reported superior performance with fewer parameters and reduced computational cost compared to traditional CNN models [3]. Furthermore, survey-based analyses confirmed that EfficientNet consistently achieves competitive or superior results across multiple datasets, particularly in resource-constrained environments [2].

In the context of campus image classification, existing studies primarily focus on generic object recognition datasets, while limited research addresses multi-class classification within university environments. Therefore, this study builds upon previous work in transfer learning and CNN-based image classification by comparing ResNet50 and EfficientNetB0 for classifying campus images. By leveraging EfficientNet's compound scaling strategy, the proposed approach aims to achieve improved classification accuracy while maintaining reasonable computational efficiency.

EfficientNet introduces a compound scaling strategy that balances depth, width, and input resolution [4]. Transfer learning has shown strong performance in image classification tasks [5]. Transfer learning has shown strong performance in image classification tasks [6]. Class imbalance has a significant impact on CNN classification performance, often biasing models toward majority classes [7]. Transfer learning has been widely adopted in image classification tasks [8]

Several studies have further confirmed the effectiveness of transfer learning, EfficientNet architectures, and data augmentation techniques across diverse application domains, including medical imaging, plant classification, and COVID-19 detection [9–20].

3. METHODS

This paper adopts a structured methodology to develop an automated image classification system for a university campus environment. The workflow begins with data acquisition, where images are collected from various campus locations, followed by exploratory data analysis to examine dataset characteristics and class distribution. The dataset is then partitioned into training, validation, and test subsets to ensure a fair and unbiased evaluation process.

After dataset partitioning, the training data is shuffled to improve model generalization, while caching and prefetching techniques are applied to optimize the data input pipeline and accelerate training. All images are resized to a fixed resolution of 224 × 224 pixels to meet the input requirements of the EfficientNetB0 architecture. Image normalization is then performed using the EfficientNet-specific preprocessing function to enable effective transfer learning from ImageNet-pretrained weights.

To address class imbalance within the dataset, class weights are computed based on the distribution of samples in the training set. These weights are incorporated during the training process to reduce bias toward majority classes and improve classification performance across all categories. Finally, the EfficientNetB0 model is trained using the prepared dataset and evaluated on the validation and test sets to assess its classification performance. The overall workflow of the proposed methodology is illustrated in Fig. 1.

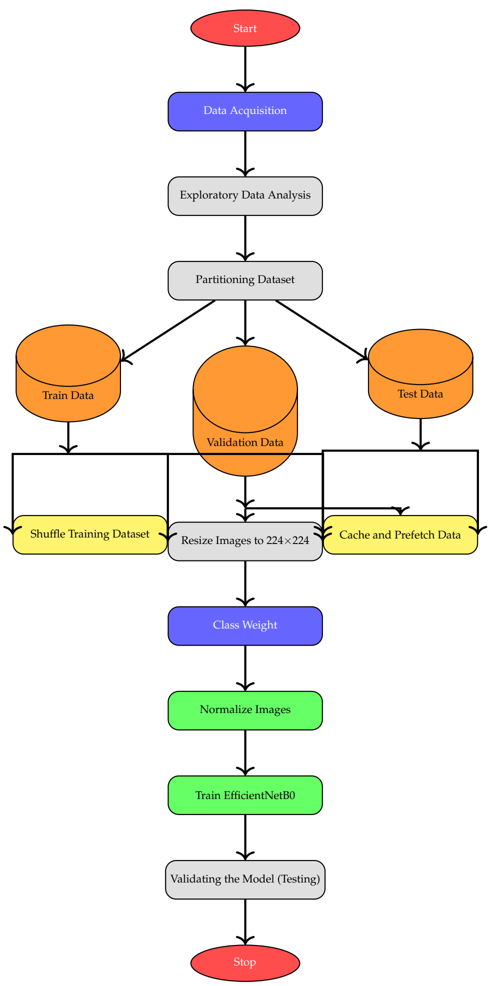

Fig. 1. Overall workflow

A. Data Acquisition

In this paper, the dataset was constructed using images captured manually within a real university campus environment. The images were collected using a digital camera and mobile devices to ensure diversity in image quality and capture conditions. Data acquisition was carried out across multiple locations within the campus, including open spaces, academic buildings, laboratories, and surrounding outdoor areas. This approach was adopted to reflect realistic campus scenes and capture variations in background, lighting, and object appearance.

The collected dataset consists of a total of 2,370 images categorized into five distinct classes: trees, people, buildings, laboratories, and cars. The images were organized into a structured directory format, where each class was stored in a separate folder to facilitate efficient data loading and labeling during model training. To analyze the distribution of images across different categories, a bar chart was generated illustrating the number of images in each class. This visualization helped assess class balance and provided insight into the dataset composition prior to model training.

Fig. 2 illustrates the distribution of images across the five classes before dataset splitting. As shown in the bar chart, the dataset exhibits a relatively balanced class distribution, with minor variations in the number of images per category. This distribution helps reduce class bias during model training and supports stable learning behavior. In addition, Fig. 3 presents one representative sample image from each class in the dataset. These examples highlight the visual diversity within the dataset and demonstrate real-world challenges such as background clutter, varying illumination, and differences in object appearance. Including sample images provides qualitative insight into the dataset characteristics prior to preprocessing and model training.

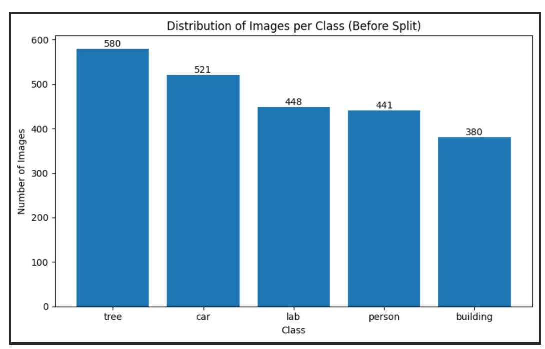

Fig. 2. Distribution of images per class

Fig. 3. Presentation Images of Campus Data

B. Data Preprocessing

To ensure effective model training, several preprocessing steps were applied to the campus image dataset. The dataset was organized into training, validation, and test subsets to enable unbiased evaluation. All images were resized to a fixed resolution of 224 × 224 pixels to match the input requirements of the EfficientNetB0 architecture.

The training dataset was shuffled to enhance generalization, while the validation and test sets were kept unshuffled for consistent evaluation. Dataset performance was optimized using caching and prefetching techniques, which reduce input pipeline overhead and accelerate training.

To address class imbalance, class weights were computed based on the distribution of training samples across the five classes. These weights were later used during training to reduce bias toward majority classes. In addition, image normalization was performed using the EfficientNet-specific preprocessing function to enable effective transfer learning from ImageNet-pretrained weights.

To mitigate class imbalance, class weights were calculated using the following equation:

$$w_c = \frac{N}{C \times n_c} \tag{1}$$

where N represents the total number of training samples, C denotes the number of classes, and n_c is the number of samples belonging to class c.

C. Model Architecture and Training

EfficientNetB0 was initialized with pretrained ImageNet weights and used as the backbone of the proposed model. The model was designed to classify objects commonly observed in the university environment, including cars, people, laboratories, trees, and buildings. EfficientNetB0 employs compound scaling to balance network depth, width, and input resolution, enabling efficient feature extraction while maintaining strong representational capacity.

Fig. 4 illustrates the architecture of the EfficientNetB0-based model used in this study. The pretrained EfficientNetB0 backbone was configured without the top classification layers and used as a feature extractor. To enable transfer learning while preserving learned representations, the majority of the backbone layers were frozen, while the last 40 layers were fine-tuned to adapt the model to the target classification task.

After feature extraction, a Squeeze-and-Excitation (SE) attention block was integrated into the architecture. This attention mechanism recalibrates channel-wise feature responses by explicitly modeling interdependencies between feature maps, allowing the model to emphasize informative features relevant to university environmentobjects while suppressing less relevant ones.

Following the attention module, a global average pooling layer was applied to reduce the spatial dimensions of the feature maps. A fully connected dense layer with 256 units and ReLU activation was then used to learn task-specific feature representations, with L2 regularization applied to reduce overfitting. Batch normalization was introduced after the dense layer to stabilize training and improve convergence, followed by an additional ReLU activation and a dropout layer with a dropout rate of 0.3 for further regularization. The final classification layer consisted of a dense softmax layer that outputs class probabilities for the five target categories.

The model was implemented using the TensorFlow and Keras frameworks. The training process employed the Adam optimizer with a learning rate of 1×10^{-4} and sparse categorical cross-entropy as the loss function. The model was trained for up to 45 epochs using class weighting to address class imbalance. Early stopping was applied based on validation loss to prevent overfitting, and a learning rate reduction strategy was used to dynamically adjust the learning rate during training for improved convergence

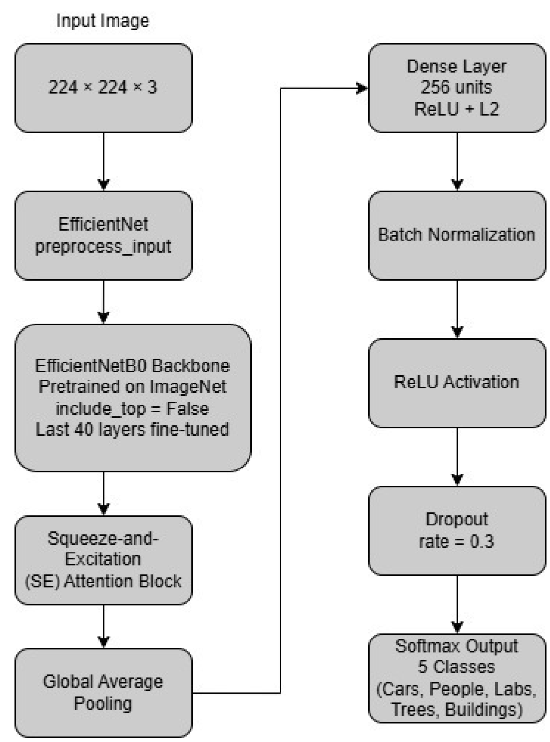

Fig. 4. Model Architecture

D. Model Evaluation

Model evaluation is a crucial stage for assessing the effectiveness of the proposed deep learning models in accurately classifying unseen campus images. In this paper, the evaluation process is conducted using several widely adopted performance metrics, including validation accuracy, classification report, and confusion matrix. These metrics provide both quantitative and qualitative insights into the models' classification capabilities.

Validation accuracy is used to measure the proportion of correctly classified samples in the validation dataset. This metric reflects the model's ability to generalize learned features from the training data to unseen data. A higher validation accuracy indicates better generalization performance. The accuracy metric is computed using Equation (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP represents true positives, TN true negatives, FP false positives, and FN false negatives.

In addition to accuracy, a classification report is employed to provide a more detailed evaluation for each class. The report includes three key metrics: precision, recall, and F1-score. Precision, defined in Equation (2), measures the correctness of the model's positive predictions:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Recall, shown in Equation (3), evaluates the model's ability to correctly identify all true instances of a given class:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

The F1-score, presented in Equation (4), is the harmonic mean of precision and recall. It provides a balanced measure of model performance, particularly in scenarios where class distributions may not be perfectly balanced:

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

Furthermore, the confusion matrix is used as a visual evaluation tool to compare the actual class labels with the predicted labels. This matrix allows for a detailed analysis of misclassifications across different classes and helps identify patterns of confusion between visually similar categories.

To assess the effectiveness of the proposed approach, a comparative evaluation is conducted between EfficientNetB0 and ResNet50. Both models are trained and evaluated under the same experimental settings and dataset splits. The comparison highlights the differences in classification performance and demonstrates the superiority of EfficientNetB0 in terms of feature representation and overall accuracy for campus image classification.

4. RESULT AND DISCUSSION

In this section we will present the experimental results obtained from the image classification in university campus task using two variants of the EfficientNetB0 architecture: a baseline model and an enhanced model incorporating (attention mechanisms, regularization and optimization strategies, and class balancing strategies). We made a detailed analysis to evaluate and compare their performance using multiple evaluation metrics. Our discussion will focus on the models' generalization ability, class-wise behaviour, and the effectiveness of the proposed enhancements for improving classification accuracy in real-world visual recognition scenarios. The evaluation was conducted on a held-out test dataset to assess each model's ability to generalize beyond the training data. Performance indicators including (overall test accuracy, precision, recall, F1-score, and confusion matrices) are analysed to identify strengths and limitations across all five classes: building, car, lab, person, and tree. Furthermore, a comparative analysis is presented to highlight the impact of architectural and training improvements introduced in the enhanced model.

A. Training and Validation Performance

The training behaviour of both the baseline and enhanced EfficientNetB0 models is illustrated through training and validation accuracy and loss curves. These plots provide insights into convergence behaviour and generalization capability.

For the baseline EfficientNetB0 model, the training accuracy increased rapidly and approached nearly 100% after several epochs, while the validation accuracy stabilized between 96% and 98%, as illustrated in Fig. 5. , Fig. 6, Although the validation performance remained high, a small but consistent gap between training and validation accuracy was observed. Similarly, the loss curves show a continuous decrease in training loss, whereas the validation loss converged at a slightly higher level. This behavior indicates the presence of mild overfitting; however, the overall generalization performance of the model remains strong.

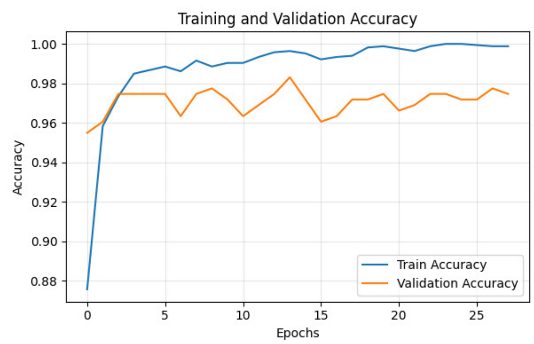

Fig. 5. Training and Validation Accuracy

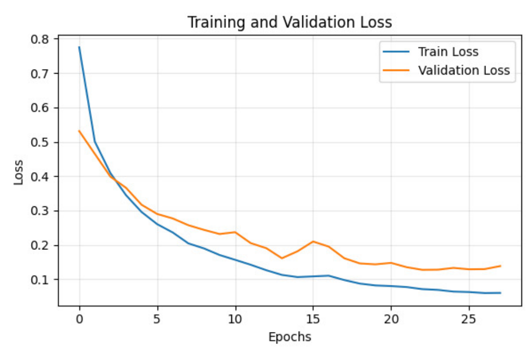

Fig. 6. Training and Validation Loss

In contrast, the enhanced EfficientNetB0 model demonstrated improved training stability and generalization. As illustrated in Fig. 7, Fig. 8, the validation accuracy closely followed the training accuracy throughout the training process, reaching approximately 98–99%. The corresponding loss curves exhibit a rapid and smooth decrease for both training and validation loss, with minimal divergence between them. These results indicate that the introduced enhancements, including the squeeze-and-excitation attention block, batch normalization, dropout regularization, and class weighting, effectively reduced overfitting and improved the convergence behavior. Overall, the enhanced model shows superior generalization compared to the baseline, which is consistent with the observed improvement in test accuracy and classification metrics.

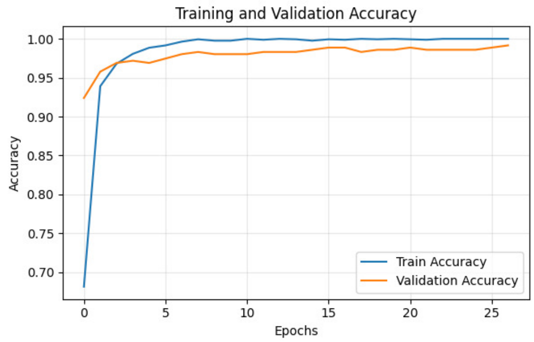

Fig. 7. Training and Validation Accuracy

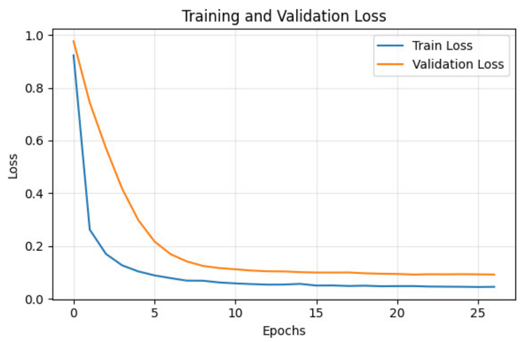

Fig. 8. Training and Validation Loss

B. Classification Metrics

B.1. Baseline Model Performance

The baseline EfficientNetB0 model achieved an overall test accuracy of 96.09% on the test dataset. The detailed class-wise precision, recall, and F1-score are summarized in Table 1, which presents the corresponding performance table.

Table 1. Classification performance of the baseline EfficientNetB0 model

Class	Precision	Recall	F1-score	Support
Building	0.88	0.98	0.93	57
Car	1.00	0.94	0.97	79
Lab	0.96	0.97	0.96	68
Person	0.97	0.94	0.95	67
Tree	0.99	0.98	0.98	87
Accuracy			**0.96**	358
Macro Avg	0.96	0.96	0.96	358
Weighted Avg	0.96	0.96	0.96	358

As shown in Table 1, the classification report demonstrates strong performance across all five classes, with F1-scores ranging from 0.93 to 0.98. The *tree* and *car* classes achieved particularly high precision and recall values, reflecting their visually distinctive characteristics. In contrast, the *building* class exhibited a comparatively lower precision of 0.88, indicating occasional misclassification with structurally similar categories. Furthermore, the macro-averaged and weighted F1-scores of 0.96 confirm the overall robustness of the model despite minor class-wise performance variations.

B.2. Enhanced Model Performance

The enhanced EfficientNetB0 model achieved a higher test accuracy of 98.32%, representing a clear improvement over the baseline model. The detailed class-wise precision, recall, and F1-score are reported in Table 2, which summarizes the classification performance of the enhanced model.

Table 2. Classification performance of the enhanced EfficientNetB0 model

Class	Precision	Recall	F1-score	Support
Building	0.96	0.96	0.96	57
Car	0.99	0.99	0.99	79
Lab	0.99	0.99	0.99	68
Person	0.99	1.00	0.99	67
Tree	0.99	0.98	0.98	87
Accuracy			**0.98**	358
Macro Avg	0.98	0.98	0.98	358
Weighted Avg	0.98	0.98	0.98	358

As shown in Table 2, all classes achieved F1-scores of 0.96 or higher, with the *person* class reaching an F1-score of 0.99. Improvements in both precision and recall were observed across all categories, particularly for the *building* class, where precision increased from 0.88 to 0.96, indicating a substantial reduction in misclassification. These results demonstrate that the integration of attention mechanisms, class weighting, and regularization techniques significantly enhanced the model's ability to distinguish between visually similar classes, while maintaining high overall accuracy and reducing overfitting.

C. Confusion Matrix Analysis

The confusion matrices provide a detailed insight into the class-wise prediction behavior of both the baseline and enhanced EfficientNetB0 models. For the baseline model, the confusion matrix shown in Fig. 9 indicates that the majority of samples were correctly classified, with only a limited number of misclassifications observed across several classes. Notable confusion occurred between the *building* and *tree* classes, as well as between the *person* and *lab* classes. For instance, several *tree* samples were incorrectly predicted as *building*, while a small number of *person* images were misclassified as *lab*. These errors can be attributed to shared structural features and complex background elements commonly present in campus environments.

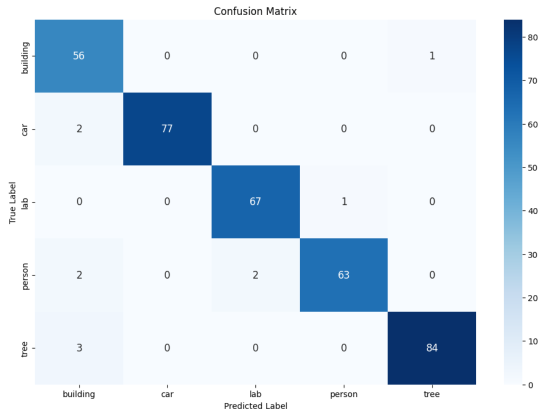

Fig. 9. Confusion Matrix – Baseline Model

In contrast, the enhanced EfficientNetB0 model demonstrates improved class separation, as illustrated in Fig. 10. The number of misclassifications is further reduced, and several classes exhibit near-perfect prediction performance. Notably, the *person* class achieved complete separation with no misclassified samples. The remaining errors are primarily confined to visually similar categories, such as *building* and *tree*, which often share overlapping textures and spatial structures in real-world campus environments.

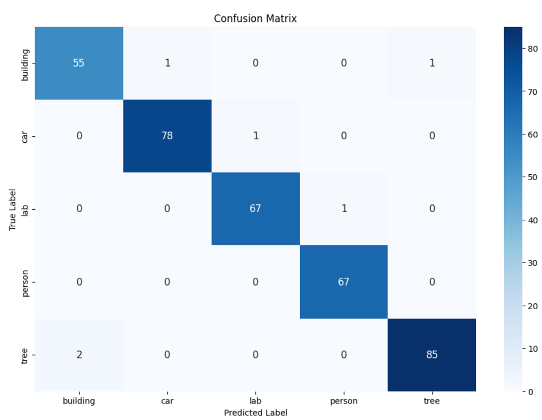

Fig. 10. Confusion Matrix – Enhanced Model

Overall, the enhanced model's confusion matrix confirms that the integration of attention mechanisms, class weighting, and additional regularization techniques effectively reduces inter-class confusion and improves robustness, particularly for classes that were more prone to misclassification in the baseline configuration.

D. Comparative Analysis of Baseline and Enhanced Models

A direct comparison between the two models shows that the enhanced EfficientNetB0 outperforms the baseline across all evaluated metrics. In particular, the overall test accuracy increased by 2.3 percentage points, from 96.0% to 98.3%.

The observed performance gains can be attributed to the following enhancements:

- **The squeeze-and-excitation (SE) attention block**, which improves feature discrimination by modeling inter-channel dependencies.
- **Batch normalization and dropout**, which enhance training stability and effectively reduce overfitting.
- **Class weighting**, which mitigates bias toward majority classes and improves class-wise balance.
- **A lower learning rate**, enabling finer weight updates during optimization and smoother convergence.

E. Misclassified Sample Analysis

We further investigated the classification errors of the enhanced EfficientNetB0 model, conducting qualitative analysis of the misclassified test samples. After evaluating the enhanced model on the held-out test dataset, only six images were misclassified out of 358 test samples, confirming the model's strong generalization capability and high robustness.

Figure 11 presents a visualization of the misclassified test images along with their corresponding ground-truth and predicted labels. The limited number of errors indicates that the proposed enhancements were effective in minimizing incorrect predictions.

A closer inspection of the misclassified samples reveals that most errors occurred between visually or contextually similar categories, particularly *building* and *tree*, *car* and *building*, *car* and *lab*, and *lab* and *person*. Unlike the *building–tree* cases, these misclassifications are not caused by the presence of multiple object categories within the same image, but rather by similarities in structural patterns, textures, and contextual cues learned by the model.

For example, laboratory environments may contain metallic surfaces, rectangular structures, or equipment layouts that resemble vehicle components under certain viewpoints or lighting conditions, leading to confusion between the *car* and *lab* classes. Similarly, confusion between the *lab* and *person* classes can be attributed to the presence of human-associated visual cues, such as laboratory coats or indoor backgrounds, which may bias the model toward the *person* class even when a clearly visible individual is absent.

Despite these challenges, the very small number of misclassified images (only six samples) demonstrates that the enhanced model maintains a high level of discriminative capability. This qualitative error analysis complements the quantitative evaluation metrics presented earlier and further confirms that the enhanced EfficientNetB0 model effectively reduces inter-class confusion while maintaining strong performance across all categories.

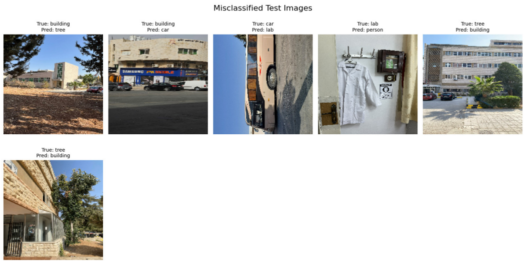

Fig. 11. Visualization of Misclassified Test Images – Enhanced EfficientNetB0 Model

F. Discussion and Limitations

The enhanced EfficientNetB0 model has had high classification accuracy and robustness this demonstrates its potential suitability for real-world image recognition applications. These applications include smart surveillance systems, autonomous navigation, and intelligent urban monitoring platforms, where accurate and reliable object classification is critical. Furthermore, EfficientNetB0's lightweight architecture and computational efficiency make it particularly suitable for deployment on resource-constrained devices, such as edge and embedded systems.

The performance improvements observed in the enhanced model suggest that the incorporation of (attention mechanisms, regularization and optimization strategies, Batch Normalization and class balancing strategies) contributes to improve generalization and stability, which are essential characteristics for real-world deployment. The reduction in misclassification errors, as shown in the confusion matrix analysis and other evaluation metrics , further supports the model's robustness when handling visually similar object categories.

Despite these promising results, several limitations should be acknowledged. First, although the dataset is sufficient for experimental evaluation, it may not fully represent the diversity of real-world conditions, such as severe occlusions, extreme lighting variations, or rare object appearances. Second, all experiments were conducted in a controlled offline environment, and the model's performance under real-time operational constraints was not assessed. Additionally, only the EfficientNetB0 variant was evaluated, limiting insights into the scalability of the approach to larger model variants.

Future work will focus on expanding the dataset to include more diverse and challenging samples, evaluating higher-capacity EfficientNet variants, and validating the model's performance in real-time and real-world deployment scenarios.

5. CONCLUSION

This paper presented a deep learning-based multi-class image classification framework for university campus environments using EfficientNetB0 and transfer learning techniques. A real-world campus dataset containing five object categories(buildings, cars, laboratories, people, and trees) was constructed and used to evaluate the proposed approach. The study investigated both a baseline EfficientNetB0 model and an enhanced version incorporating attention mechanisms, regularization techniques, and class balancing strategies.

Experimental results demonstrated that the enhanced EfficientNetB0 model significantly outperformed the baseline configuration and ResNet50 across all evaluation metrics. The enhanced model achieved a test accuracy of 98.32%, with consistently high precision, recall, and F1-scores across all classes. Confusion matrix analysis further confirmed a substantial reduction in misclassification errors, particularly among visually similar categories such as buildings and trees. The integration of squeeze-and-excitation attention, batch normalization, dropout, and class weighting contributed to improved feature discrimination, training stability, and generalization performance.

Overall, the results indicate that EfficientNetB0 provides an effective balance between classification accuracy and computational efficiency, making it suitable for real-world applications in smart campus systems, surveillance, and intelligent urban monitoring. Despite the strong performance, this study has some limitations, including the use of a single campus dataset and offline evaluation conditions. Future work will focus on expanding the dataset to include more diverse environments, exploring larger EfficientNet variants, and validating the proposed system under real-time deployment scenarios.

REFERENCES

1. A. S. Muhammad Hallaj Nashrulloh, J. Intell. Syst. Technol. Informatics (2025).
2. L. I. Umeaduma, World J. Adv. Res. Rev. (2024).
3. M. M. R. Umme Ayman *et al.*, (2024).
4. M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, (2020).
5. J. W. D. R. Abdullah Caliskan, Hasan Badem, Electronics (2025).
6. S. I. Anas Mohammad Ishfaqul Muktadir, Taimur Rahman, Int. Conf. on Learn. Represent. (ICLR), Tiny Pap. (2024).
7. F. J. Pulgar, A. J. Rivera, F. Charte, and M. J. del Jesus, "On the impact of imbalanced data in convolutional neural networks performance," in *Proceedings of the International Conference on Hybrid Artificial Intelligence Systems (HAIS)*, (2017), Lecture Notes in Artificial Intelligence (LNAI).
8. M. Hussain, J. J. Bird, and D. R. Faria, "A study on cnn transfer learning for image classification," in *Proceedings of the UK Workshop on Computational Intelligence (UKCI)*, (2018).
9. S. J. Pan and Q. Yang, IEEE Trans. on Knowl. Data Eng. (2010).
10. J. Deng *et al.*, IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR) (2009).
11. K. He, X. Zhang, S. Ren, and J. Sun, IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR) (2016).
12. J. Yosinski *et al.*, Adv. Neural Inf. Process. Syst. (NeurIPS) (2014).
13. W. Rawat and Z. Wang, Neural Comput. (2017).
14. M. Buda, A. Maki, and M. A. Mazurowski, Neural Networks (2018).
15. E. C. Too *et al.*, Comput. Electron. Agric. (2019).
16. X. Li and W. Zhang, IEEE Access (2020).
17. L. Alzubaidi *et al.*, J. Big Data (2021).
18. M. A. Talukder, M. A. Layek, M. Kazi, *et al.*, Comput. Biol. Med. (2024).
19. T. Ozturk and M. Talo, Comput. Biol. Med. (2020).
20. C. Shorten and T. M. Khoshgoftaar, J. Big Data (2019).