International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 14, No. 3, December 2025, pp. 903~913
ISSN: 2252-8776, DOI: 10.11591/ijict.v14i3.pp903-913

A hybrid approach using VGG16-EffcientNetV2B3-FCNets for accurate indoor vs outdoor and animated vs natural image classification

Meghana Deshmukh, Amit Gaikwad, Snehal Kuche
Department of Computer Science and Engineering, G. H. Raisoni University Amravati (Maharashtra), Maharashtra, India

Article Info

Article history:

Received Aug 27, 2024
Revised Mar 5, 2025
Accepted Jun 9, 2025

Keywords:

CNN
EfficientNetV2B3
FCNet
Feature aggregation
Scene classification
VGG16

ABSTRACT

The paper introduces a hybrid approach that synergistically combines the strengths of VGG16, EfficientNetV2B3, and fully connected networks (FCNets) to achieve precise image classification. Specifically, our focus lies in discerning between basic indoor and outdoor scenes, further extended to distinguish between animated and natural images. Our proposed hybrid architecture harnesses the unique characteristics of each component to significantly enhance the model's overall performance in fine-grained image categorization. In our methodology, we utilize VGG16 and EfficientNetV2B3 as the feature extractors. During evaluation, we examined various classification algorithms, such as VGG16, EfficientNet, Feature_Aggr_Avg, and Feature_Aggr_max, among others. Notably, our hybrid feature aggregation approach demonstrates a remarkable improvement of 0.5% in accuracy compared to existing solutions employing VGG16 and EfficientNet as feature extractors. Notably, for indoor versus outdoor image classification, feature_aggr_avgachieves an accuracy of 98.51%. Similarly, when distinguishing between animated and natural images, Feature_Aggr_Avgachieves an impressive accuracy of 99.20%. Our findings demonstrate improved accuracy with the hybrid model, proving its adaptability across diverse classification tasks. The model is promising for applications like automated surveillance, content filtering, and intelligent visual recognition, with its robustness and precision making it ideal for real-world scenarios requiring nuanced categorization.

Corresponding Author:

Meghana Deshmukh
Department of Computer Science and Engineering, G. H. Raisoni University Amravati (Maharashtra)
Maharashtra, India
Email: meghanadeshmukh440@gmail.com

1. INTRODUCTION

In the era of ever-expanding digital content, image classification plays a pivotal role in various applications ranging from automated surveillance to content recommendation systems. Accurate classification of images into relevant categories is crucial for optimizing the performance of these systems. This research paper introduces a novel hybrid approach that combines the strengths of three state-of-the-art deep learning architectures: VGG16, EfficientNetV2B3, and FCNets. The proposed hybrid model aims to increase the accuracy of image categorization, specifically focusing on distinguishing between indoor and outdoor scenes, as well as discriminating between animated and natural images, The authors propose a deep function aggregation framework combining convolutional neural networks (CNNs) for function extraction and graph convolutional networks (GCNs) to seize spatial relationships among functions for stepped forward scene type of deep feature aggregation framework driven by graph convolutional network for scene classification in remote sensing and The proposed model achieves overall performance in far flung sensing scene type duties, outperforming existing methods in accuracy on benchmark datasets [1], [2].

The choice of indoor Vs outdoor and animated Vs natural image classification [3] is motivated by their widespread applications. Recognizing whether an image is taken indoors or outdoors is essential for tasks such as intelligent camera systems, security surveillance, and augmented reality applications. Similarly, differentiating between animated and natural images is crucial for content recommendation systems, entertainment platforms, and gaming industries. The hybrid approach presented in this research leverages the distinctive features and capabilities of each component architecture. VGG16, known for its simplicity and effectiveness, forms the backbone of the model, providing a solid foundation for feature extraction. EfficientNetV2B3, an advanced and efficient architecture designed for optimal performance, contributes to improved feature representation and computational efficiency. FCNets, focusing on fine-grained classification, are incorporated to increase the model's ability to discriminate in-between subtle differences in image categories.

The synergy between these architectures creates a robust and accurate image classification model that addresses the challenges posed by the complex nature of indoor Vs outdoor and animated Vs natural image categorization. By combining the power of these architectures, we aim to achieve a model that excels in generalization, adaptability. This research contributes to the growing field of image classification by presenting a hybrid approach that not only outperforms individual architectures but also establishes a foundation for exploring synergies between different deep learning models. The remainder of this paper will delve into the methodology, experimental setup, and results to validate the effectiveness of the proposed hybrid model for accurate image classification in the specified domains [4], [5]. Categorizing indoor vs. outdoors snap shots enhances consumer enjoy in packages like photo control, personalised suggestions, and smart domestic systems, improving algorithms, dataset pleasant, and model performance across various domain names.

Recent literature reviews deep learning for scene classification and remote sensing. In recent years, there have been substantial additions to the literature on the topic of utilising deep learning models to categorise photographs as either indoors or outside, or as animated or natural. This literature review compiles and summarises the results of research that has looked at scene classification, model architectures, and remote sensing imagery applications from a variety of angles. Souza *et al.* [6] proposes the method I,e utilizing big data to develop a hybrid model for scene classification and highlighted importance of big data and proposed hybrid model's functionality. Perturbation-seeking generative adversarial networks for remote sensing introduced by Cheng *et al.* [7] in which he says employing generative adversarial networks that reliably classify scenes in remote sensing images offered viewpoints on improving robustness of scene categorization models. Yao *et al.* [8] invented utilizing weak supervision for object detection in remote sensing images which has provided insights for future studies on similar methods. Huang *et al.* [9] proposes tracked changes in ecosystem services using deep learning and high-resolution remotely sensed imagery which outs showed potential wider uses of picture categorization in environmental tracking using tracking changes in ecosystem services using deep learning. The introduction of framework for high spatial resolution image scene categorization using deep sparse semantic modeling is done by Zhu *et al.* [10] which outputs the contributed to understanding effective processing of high-resolution images. Land-use classification using deep convolutional feature-based extreme learning classifier by author Weng *et al.* [11] Suggested method for land-use classification utilizing deep features and offered insights into using deep features for land-use classification. Li *et al.* [12] introduces spatial-temporal super-resolution land cover mapping model for super-resolution land cover mapping with spatial-temporal considerations and suggested potential methods for addressing temporal concerns. Classification of interior scenes using a combination of global and semantic variables invented by Pereira *et al.* [13] and concludes with emphasized significance of feature fusion for scene classification. Cheng *et al.* [14] and Cheng *et al.* [15] both authors introduce combining visual terms with multi-scale completed local binary patterns which provides provided historical context for scene classification techniques. Liu *et al.* [16] advocate a random-scale stretched convolutional neural community (RS-SCNN), which applies random-scale stretching to enter pics for reinforcing multi-scale function learning, improving scene semantic category in excessivespatial decision faraway sensing imagery and concludes The RS-SCNN version achieves advanced class overall performance in comparison to conventional CNNs, demonstrating stepped forward accuracy in coping with diverse scene scales in high-decision faraway sensing datasets. A novel metaheuristic optimizer inspired by behavior of jellyfish in ocean by author Chou and Truong [17] recommended the Jellyfish seek (JS) optimizer, a novel metaheuristic algorithm inspired via jellyfish conduct inside the ocean, such as passive motion and energetic swimming. The algorithm mimics jellyfish dynamics to discover and exploit search areas for optimization troubles.

Mirjalili and Lewis [18] recommends the Whale Optimization algorithm (WOA) is a nature-stimulated optimization approach that mimics the bubble-net looking method of humpback whales. It balances exploration and exploitation via modeling encircling and spiral movement towards the first-class answer. WOA is widely used for solving complicated optimization troubles in engineering and gadget getting to know. The author Wu *et al.* [19] introduce YOLO-DCNet, a flexible, light-weight human detection algorithm that integrates semantic-based statistics with the YOLO (You best appearance once) structure. The model specializes in improving detection accuracy while keeping performance in actual-time eventualities which proposes YOLO-DCNet demonstrates advanced overall performance in human detection obligations, reaching better detection accuracy and decrease computational charges in comparison to conventional YOLO-primarily based models, making it suitable for real-time packages. Hao *et al.* [20] proposes an entropy-augmented neural community for correct high-density crowd counting. It contains facts entropy to manual the model in focusing on uncertain or ambiguous regions in dense crowds. This method improves robustness and accuracy in complex real-international crowd analysis eventualities. Masud *et al.* [21] present a method the use of pre-skilled convolutional neural networks (CNNs) for breast most cancers detection in ultrasound pics. The method leverages transfer getting to know to enhance diagnostic accuracy with confined classified information. Effects display high overall performance in classifying benign and malignant tumors efficaciously. Multiround transfer learning and modified generative adversarial networks for lung cancer detection by Chui *et al.* [22] proposed model achieves sizable upgrades in lung cancer detection accuracy, outperforming traditional techniques. The multiround transfer learning and GAN augmentation result in higher generalization and robustness, mainly in coping with restricted or imbalanced datasets. Zhu *et al.* [23] this paper proposes a bag-of-visible-phrases (BoVW) scene classifier that combines nearby and international features for classifying high-resolution far flung sensing photographs. The technique complements spatial function representation for complicated scenes. It achieves effective overall performance in land-use and concrete scene recognition responsibilities. Huang *et al.* [24] introduce a method using multi-scale finished local binary styles (LBP) and fisher vectors for far flung sensing scene type. The combination captures nice-grained texture and spatial layout functions. It suggests high accuracy in discriminating diverse land cover types in satellite tv for pc snap shots. Mohamed *et al.* [25] this paper provides a deep getting to know-based totally semantic segmentation method to classify indoor and outdoor environments. The system is designed to help visually impaired wheelchair customers by using enabling real-time scene expertise, improving navigation protection and autonomy. Alves *et al.* [26] the authors suggest a singular method for indoor/outdoor type of user device in cell networks. Their version leverages sign and contextual information to enhance the accuracy of location-based offerings in next-technology cellular networks.

a) Image classification architectures
   Deep getting to know advances photo analysis with VGG16, EfficientNetV2B3, and FCNets, optimizing accuracy, performance, and best-grained type in aid-limited environments.
b) Indoor vs outdoor scene classification
   Indoor-outdoor scene categorization advanced from handcrafted features to deep learning models like GoogLeNet, ResNet, and hybrid approaches, enhancing accuracy.
c) Animated vs natural image classification
   Classifying animated versus natural images requires nuanced techniques. Traditional methods used texture features, while deep learning, incorporating temporal and spatial information, achieves superior accuracy in recent studies.
d) Hybrid approaches in deep learning
   Hybrid architectures combine diverse models like residual networks and attention mechanisms, enhancing performance by integrating spatial and temporal strengths for image tasks.
e) Challenges and opportunities
   Hybrid approaches leverage diverse architectures for improved discrimination, optimizing models, exploring transfer learning, enhancing explainability, and advancing real-time, adaptable image classification solutions.

2. PROPOSED METHOD

By using integrating VGG16, EfficientNetV2B3, and FCNs, this hybrid method improves picture category by way of leveraging VGG16's function recognition, EfficientNetV2B3's pattern detection, and FCNs' synthesis, improving accuracy and performance at the same time as refining class methods with superior analysis and thresholding. Figures 1 and 2 display an photograph evaluation pipeline with dataset collection, preprocessing, function extraction, aggregation, and evaluation for predictions.

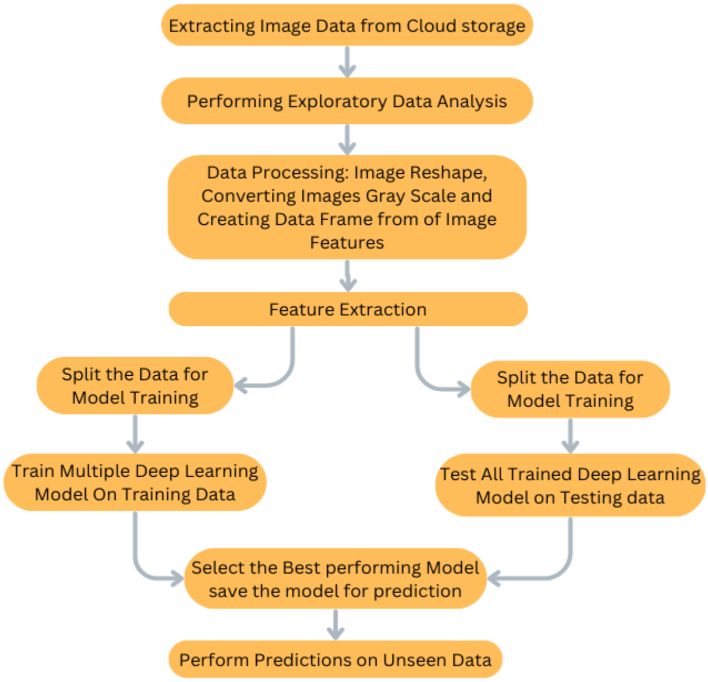

Figure 1. Proposed approach for image classification

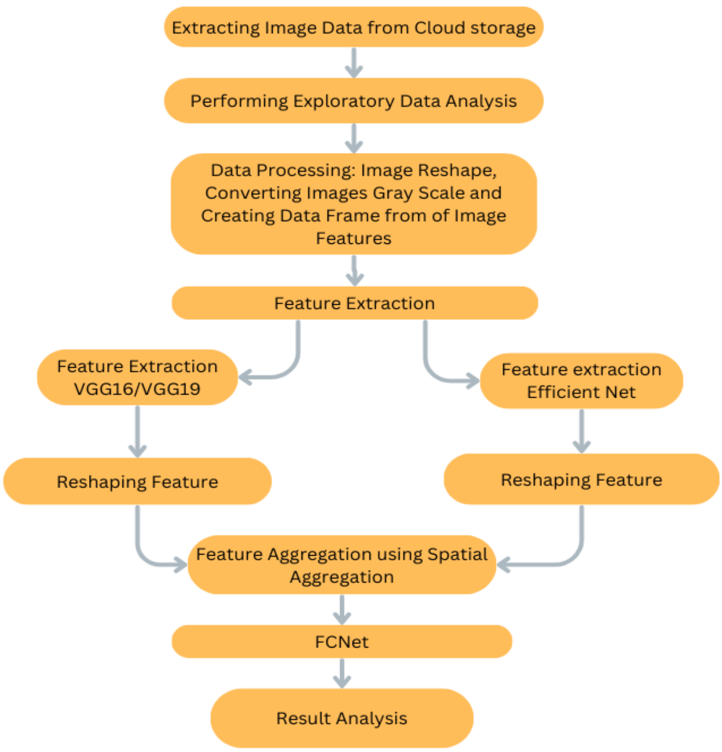

Figure 2. Proposed methodology flowchart for image classification

3. RESEARCH METHODOLOGY

3.1. Basic architecture of research methodology

Here are the 7 steps typically involved in a pipeline for an image classification problem as shown in Figure 3.

a) Data collection and preparation:
Collect a applicable image dataset, prepare it into schooling, validation, and trying out sets, and preprocess pictures by using resizing and normalizing for uniformity and decreased complexity.

b) Exploratory data analysis (EDA):
- The distribution of classes is one dataset characteristic that you should be familiar with, image resolutions, and any anomalies.
- Visualize sample images from different classes to gain insights into the data.

c) Feature extraction:
Deep learning routinely learns features thru convolutional layers, not like traditional methods.

d) Model selection:
Select a suitable machine learning model for image classification. CNNs are widely utilized in deep learning tasks because of their exceptional ability to capture spatial hierarchies and patterns in images.

e) Model training:
Educate the version, optimize hyperparameters, and reveal performance at the validation set to save you overfitting.

f) Model evaluation:
Examine version overall performance the usage of accuracy, precision, do not forget, F1-score, and visualize effects with confusion matrices.

g) Model deployment:
Installation the version for real-time predictions, combine it into packages, and monitor performance for important retraining.

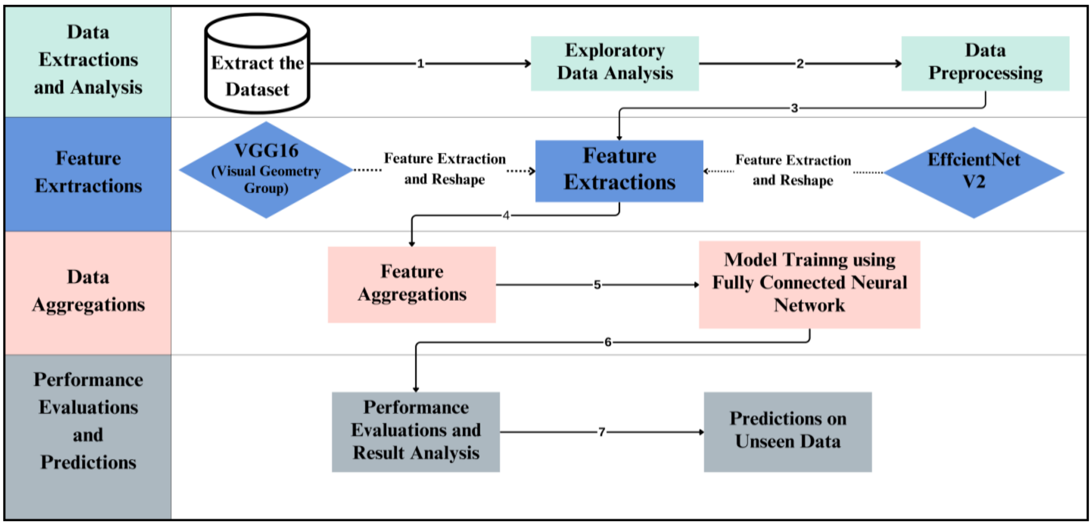

Figure 3. Research methodology for image classification

3.2. VGG16 (feature extractor)

VGG16 is known for its deep convolutional structure. The architecture is composed of multiple convolutional layers, which are then followed by max-pooling layers, and finally culminate in fully connected layers. The outcome of the final fully connected layer acts as a high-level feature representation.

3.3. EfficientNetV2B3 (feature extractor)

EfficientNetV2B3 is a variant of the EfficientNet architecture, designed for optimal performance and efficiency. It includes multiple blocks with different spatial resolutions (width, height, and depth). The feature maps from these blocks capture diverse information about the input image.

3.4. FCNets (classifier)

Fully Connected Networks are commonly employed for fine-grained classification tasks. The fully connected layers utilize the features extracted by the VGG16 and EfficientNetV2B3 and map them to the output classes, performing the final classification.

3.5. Mathematical equations

Let's denote:

- X as the input image,
- VGG(X) as the feature representation obtained from VGG16,
- EffNet(X) as the feature representation obtained from EfficientNetV2B3,
- FCNets([VGG(X), EffNet(X)]) as the results obtained from the fully connected layers.
- The following is a mathematical expression of the hybrid model:

$$Y=FCNets([VGG(X), EffNet(X)]) \tag{1}$$

here, Y represents the output logits or probabilities for each class. Integrating the features extracted, fully connected layers by VGG16 and EfficientNetV2B3 for the final classification.

3.6. Integration of VGG16 and EfficientNetV2B3

The integration of the feature representations from VGG16 and EfficientNetV2B3 can be done by concatenating the feature vectors:

$$Aggregated\_Features=Aggregation\ ([VGG(X), EfficientNetV2B3(X)]) \tag{2}$$

line and area characteristics are averaged using weighted mean statistics. To get the weighted mean, use the following formula:

$$\bar{x}_w = \frac{\sum_{i=1}^{N} w_i \cdot x_i}{\sum_{i=1}^{N} w_i} \tag{3}$$

where: N = number of observations, xi = observations, Wi = weights

The neural network uses ReLU activation in hidden layers and sigmoid in the output layer for binary classification. Accuracy measures model performance, ensuring efficient training with appropriate layers, activation functions, and metrics.

3.7. Hidden layer with ReLU activation

Suppose we have an input vector x and weights W and biases b for the hidden layer. The output of the hidden layer h after applying the ReLU activation is:

$$h=ReLU\ (Wx+b) \tag{4}$$

where the ReLU function is defined as:

$$ReLU(z)=max\ (0, z) \tag{5}$$

3.8. Output layer with sigmoid activation

The output layer o with sigmoid activation, given the output from the hidden layer h, weights Wo, and biases bo, is: o=σ (Wo h+bo). Where the sigmoid functionσ(z) is defined as:

$$\frac{1}{1+e-z} \qquad \sigma(z) \tag{6}$$

3.9. Binary cross-entropy loss

For classification, the loss L for a single training example with true label y and predicted output $\hat{y}$ is:

$$L = -\sum_{i=1}^{C} y_i log\ (\hat{y}_i) \tag{7}$$

where:

- C As the number of classes.
- y_i As the binary indicator (0 or 1) if class label i is the correct classification for the given training example.
- $\hat{y}_i$ Is the predicted probability of the training example belonging to class i.

For a batch of N examples, the average cross-entropy loss is:

$$L = \frac{1}{N}\sum_{j=1}^{N} \sum_{i=1}^{C} y_{ij} lo\, g(\hat{y}_{ij}) \tag{8}$$

where:
- y_i is the binary indicator for the j-th example and i-th class.
- $\hat{y}_i$ is the predicted probability for the j-th example and i-th class.

4. RESULTS AND DISCUSSION

This section compares the proposed methodology with state-of-the-art methods, emphasizing accuracy and metrics like precision, recall, and F1-score. While accuracy offers an overall performance snapshot, metrics derived from the confusion matrix reveal class-specific performance. Precision minimizes false positives, recall detects relevant instances, and F1-score balances both, highlighting areas for improvement.

Table 1 shows the classification accuracy in percentage of previously used algorithm approaches. So, the CNN ensemble shows the highest accuracy. Based on the findings presented in 2, it is evident that our methodology has yielded the highest accuracy among the evaluated approaches, achieving an impressive accuracy rate of 99.20% through the feature aggregation average (avg) method. Table 2 compares accuracy for natural vs. animated images, highlighting our approach's superior performance using the Feature_Aggr_Avg technique over VGG16, EfficientNet, and Feature_Aggr_Max.

Table 1. Classification accuracy using various approaches

Approach used	Classification accuracy (%)
Deep learning	76.85
CNN and K-means	79.21
Transfer learning and CNN	82.17
CNN ensemble	85.36

Table 2. Comparison of accuracy with natural and animated images

Algorithm/Methodology	Classification accuracy (%)	
	Indoor vs outdoor	Animated vs natural
VGG 16	98.93	97.88
EfficientNet	97.60	98.09
Feature_Aggr_Max	89.36	98.30
Feature_Aggr_Avg (our apporach)	99.20	98.51

The Feature_Aggr_Avg method excels in capturing spatial information, leading to accurate classifications. Figures 4 and 5 compare algorithm performance, showing Feature_Aggr_Avg's superior accuracy in classification tasks.

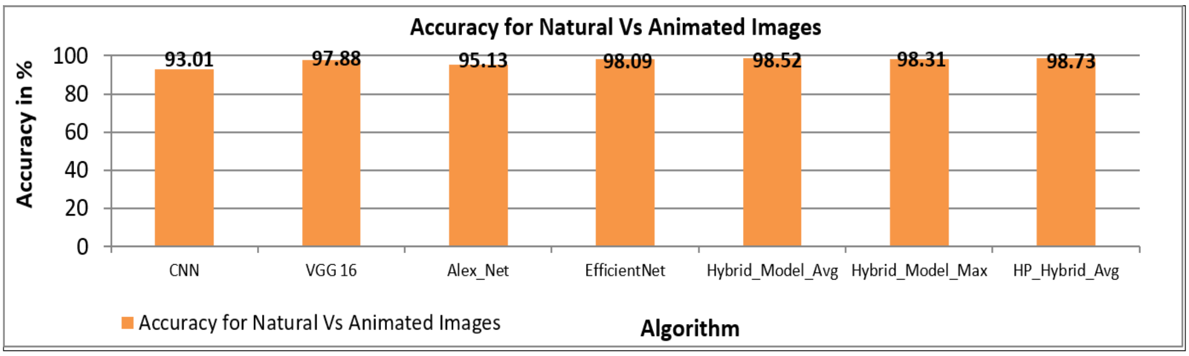

Figure 4. Accuracy of natural vs animated images

Feature_Aggr_Avg outperforms VGG16 and EfficientNet, improving accuracy by 0.5% across various classification tasks. Figure 5 shows for the task of indoor versus outdoor image classification, Feature_Aggr_Avg achieves an impressive accuracy of 98.51%. This signifies its effectiveness in accurately distinguishing between indoor and outdoor scenes. Similarly, in the classification of animated versus natural images, Feature_Aggr_Avg exhibits exceptional performance, achieving an accuracy rate of 99.20%.

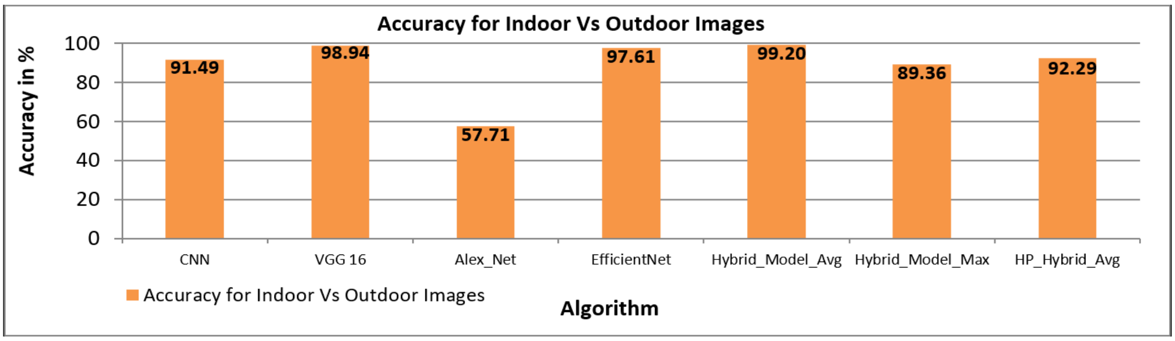

Figure 5. Accuracy of indoor vs outdoor images

The bar plots serve as a visual confirmation of the superior accuracy achieved by Feature_Aggr_Avg compared to other algorithms. Its results demonstrate how the hybrid feature aggregation method improves classification accuracy, outperforming popular feature extractors like VGG16 and EfficientNet.

Exploratory data analysis results: Figures 6 and 7 show conversions for natural vs animated and indoor vs outdoor. Channel-wise distributions (red, green, blue) reduce dimensionality when converted to grayscale.

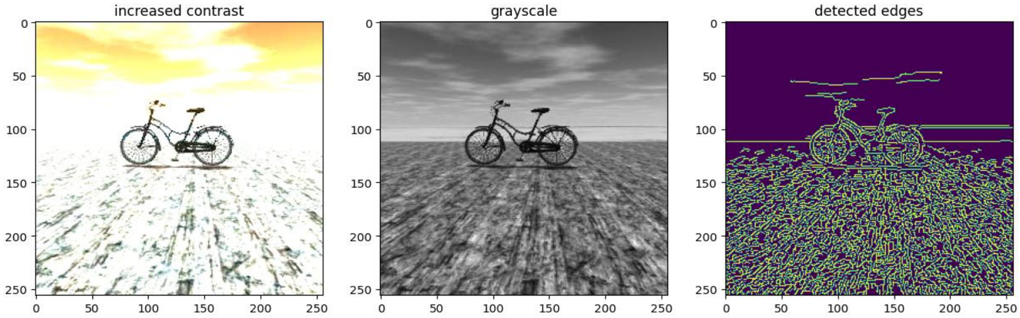

Figure 6. Conversion for natural vs animated images

Figure 7. Channel wise distribution natural vs animated images

Tables 3 and 4 evaluate CNN, VGG16, AlexNet, EfficientNet, and hybrid fashions on indoor-outside and animated-herbal snap shots, comparing accuracy, precision, remember, F1 rating, and greater. Table 4 is model results for animated vs natural.

Table 3. Tabular result analysis for indoor vs outdoor images

S. No	Algorithm	Accuracy	Error rate	Precision	Recall	F1-score	TP	TN	FP	FN
0	CNN	91.489	8.510	91.536	91.489	91.502	199	145	14	18
1	VGG 16	98.936	1.063	98.943	98.936	98.937	214	158	1	3
2	Alex_Net	57.712	42.287	33.307	57.712	42.238	217	0	159	0
3	EfficientNet	97.606	2.393	97.631	97.606	97.600	215	152	7	2
4	Hybrid_Model_Avg	99.2028	0.797	99.202	99.202	99.201	216	157	2	1
5	Hybrid_Model_Max	89.361	10.638	89.578	89.361	89.254	206	130	29	11
6	HP_Hybrid_Avg	92.287	7.712	92.365	92.287	92.244	208	139	20	9

Table 4. Tabular result analysis for natural vs animated images

S. No	Algorithm	Accuracy	Error rate	Precision	Recall	F1-score	TP	TN	FP	FN
0	CNN	93.008	6.991	93.031	93.008	93.019	345	94	16	17
1	VGG 16	97.881	2.118	97.907	97.881	97.854	357	105	9	1
2	Alex_Net	95.127	4.872	95.092	95.127	95.035	356	93	17	6
3	EfficientNet	98.093	1.906780	98.112	98.093	98.072	356	107	8	1
4	Hybrid_Model_Avg	98.516	1.483	98.522	98.516	98.519	358	107	3	4
5	Hybrid_Model_Max	98.305	1.694	98.305	98.305	98.305	358	106	4	4
6	HP_Hybrid_Avg	98.728	1.271	98.794	98.728	98.740	356	110	0	6

5. CONCLUSION

This research presents a hybrid model combining VGG16, EfficientNetV2B3, and FCNets for accurate image classification, excelling in distinguishing indoor/outdoor and animated/natural scenes. Achieving 85.36% accuracy in scene categorization and over 98% in nuanced tasks, the feature aggregation average approach stands out, highlighting its potential for future advancements and applications.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Meghana Deshmukh	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Amit Gaikwad		✓				✓		✓	✓	✓	✓	✓		
Snehal Kuche	✓		✓	✓			✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

All the authors claim that they have no battle of hobby.

DATA AVAILABILITY)

Code availability: The code might be to be had upon request

REFERENCES

[1] D. Bhardwaj and V. Todwal, "A novel deep learning model for indoor-outdoor scene classification using VGG-16 deep CNN," *World Journal of Research and Review (WJRR)*, vol. 13, no. 1, pp. 27–35, 2021, doi: 10.5815/ijisa.2021.01.07.

[2] K. Xu, H. Huang, P. Deng, and Y. Li, "Deep feature aggregation framework driven by graph convolutional network for scene classification in remote sensing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5751–5765, Oct. 2022, doi: 10.1109/TNNLS.2021.3071369.

[3] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Improving color constancy using indoor–outdoor image classification," *IEEE Transactions on Image Processing*, vol. 17, no. 12, pp. 2381–2392, Dec. 2008, doi: 10.1109/TIP.2008.2006661.

[4] Y. Huang, X. Cao, Q. Wang, B. Zhang, X. Zhen, and X. Li, "Long-short-term features for dynamic scene classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1038–1047, Apr. 2019, doi: 10.1109/TCSVT.2018.2823360.

[5] X. Xu, Y. Chen, J. Zhang, Y. Chen, P. Anandhan, and A. Manickam, "RETRACTED ARTICLE: a novel approach for scene classification from remote sensing images using deep learning methods," *European Journal of Remote Sensing*, vol. 54, no. sup2, pp. 383–395, Mar. 2021, doi: 10.1080/22797254.2020.1790995.

[6] R. P. P. M. Souza, L. J. A. dos Santos, G. T. P. Coimbra, F. A. Silva, and T. R. M. B. Silva, "A big data-driven hybrid solution to the indoor-outdoor detection problem," *Big Data Research*, vol. 24, p. 100194, May 2021, doi: 10.1016/j.bdr.2021.100194.

[7] G. Cheng, X. Sun, K. Li, L. Guo, and J. Han, "Perturbation-seeking generative adversarial networks: a defense framework for remote sensing image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022, doi: 10.1109/TGRS.2021.3081421.

[8] X. Yao, X. Feng, J. Han, G. Cheng, and L. Guo, "Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 675–685, Jan. 2021, doi: 10.1109/TGRS.2020.2991407.

[9] X. Huang, X. Han, S. Ma, T. Lin, and J. Gong, "Monitoring ecosystem service change in the City of Shenzhen by the use of high-resolution remotely sensed imagery and deep learning," *Land Degradation & Development*, vol. 30, no. 12, pp. 1490–1501, Jul. 2019, doi: 10.1002/ldr.3337.

[10] Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 1–16, 2018, doi: 10.1109/TGRS.2018.2833293.

[11] Q. Weng, Z. Mao, J. Lin, and W. Guo, "Land-use classification via extreme learning classifier based on deep convolutional features," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 704–708, May 2017, doi: 10.1109/LGRS.2017.2672643.

[12] X. Li *et al.*, "Spatial–temporal super-resolution land cover mapping with a local spatial–temporal dependence model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4951–4966, Jul. 2019, doi: 10.1109/TGRS.2019.2894773.

[13] R. Pereira, N. Goncalves, L. Garrote, T. Barros, A. Lopes, and U. J. Nunes, "Deep-learning based global and semantic feature fusion for indoor scene classification," in *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Apr. 2020, pp. 67–73, doi: 10.1109/ICARSC49921.2020.9096068.

[14] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017, doi: 10.1109/JPROC.2017.2675998.

[15] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: challenges, methods, benchmarks, and opportunities," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3735–3756, 2020, doi: 10.1109/JSTARS.2020.3005403.

[16] Y. Liu, Y. Zhong, F. Fei, and L. Zhang, "Scene semantic classification based on random-scale stretched convolutional neural network for high-spatial resolution remote sensing imagery," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2016, vol. 2016-Novem, pp. 763–766, doi: 10.1109/IGARSS.2016.7729192.

[17] J.-S. Chou and D.-N. Truong, "A novel metaheuristic optimizer inspired by behavior of jellyfish in ocean," *Applied Mathematics and Computation*, vol. 389, p. 125535, Jan. 2021, doi: 10.1016/j.amc.2020.125535.

[18] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[19] Y. Wu, J. Dong, and J. Chen, "YOLO-DCNet," *International Journal on Semantic Web and Information Systems*, vol. 20, no. 1, pp. 1–23, Feb. 2024, doi: 10.4018/IJSWIS.339000.

[20] Y. Hao, L. Wang, Y. Liu, and J. Fan, "Information entropy augmented high density crowd counting network," *International Journal on Semantic Web and Information Systems*, vol. 18, no. 1, pp. 1–15, Apr. 2022, doi: 10.4018/IJSWIS.297144.

[21] M. Masud *et al.*, "Pre-trained convolutional neural networks for breast cancer detection using ultrasound images," *ACM Transactions on Internet Technology*, vol. 21, no. 4, pp. 1–17, Nov. 2021, doi: 10.1145/3418355.

[22] K. T. Chui *et al.*, "Multiround transfer learning and modified generative adversarial network for lung cancer detection," *International Journal of Intelligent Systems*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/6376275.

[23] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-visual words scene classifier with local and global features for high spatial resolution remote sensing imagery," in *IEEE Geoscience and Remote Sensing Letters*, 2016, vol. 13, no. 6, pp. 747–751, doi: 10.1109/LGRS.2016.2559038.

[24] L. Huang, C. Chen, W. Li, and Q. Du, "Remote sensing image scene classification using multi-scale completed local binary patterns and fisher vectors," *Remote Sensing*, vol. 8, no. 6, p. 483, Jun. 2016, doi: 10.3390/rs8060483.

[25] E. Mohamed, K. Sirlantzis and G. Howells, "Indoor/outdoor semantic segmentation using deep learning for visually impaired wheelchair users," in IEEE Access, vol. 9, pp. 147914-147932, 2021, doi: 10.1109/ACCESS.2021.3123952.

[26] P. Alves *et al.*, "A novel approach for user equipment indoor/outdoor classification in mobile networks," in IEEE Access, vol. 9, pp. 162671-162686, 2021, doi: 10.1109/ACCESS.2021.3130429.

BIOGRAPHIES OF AUTHORS

Meghana Deshmukh she has done her master of engineering from Sant Gadge Baba Amravati University. She has teaching experience of 10 years and now working as an assistant professor in comuter science and engineering department of Prof. Ram Meghe Institute of Technology and Research Badnera, Amravati. (Maharashtra) India. Now, she is a Ph.D. scholar, CSE department. She can be contacted at email: meghanadeshmukh440@gmail.com.

Amit Gaikwad he received his undergraduate degree as well as his Master Degree from SGBAU, Amravati. Also, Ph.D. in Information Technology from SGBAU, Amravati. He is currently associate professor and head of Department in the Department of Computer Science and Engineering at G. H. Raisoni University, Amravati (Maharashtra-India). He published a number of papers in preferred Journals and chapters in books. He also presented various academic as well as research-based papers at several national and international conferences. His areas of interests are operating system, parallel computing, soft computing, and digital image processing. She can be contacted at email: amit.gaikwad@ghru.edu.in.

Snehal Kuche she has done her master of engineering from Rashtra SAnt Tukdoji Maharaj Nagpur University. She has teaching experience of 11 years and now working as an assistant professor in comuter engineering department of Marathwada Mitra Mandal College of Engineering, Pune (Maharashtra) India. Now, she is a Ph.D. scholar, CSE department. She can be contacted at email: snehuk.6@gmail.com.