

Paper Type: Review Article

Image Classification Using Deep Learning: A Systematic Review

Maher Khalaf Hussein ^{1,*} , Asmaa Alqassab ² , and Lubna Thanoon ALkahla ³ 

¹ Department of Computer Science, Faculty of Education, University of Telafer, Nineveh, Iraq;

maher.k.hussein@uotelafer.edu.iq.

² Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul, Iraq;

asmaa_mow@uomosul.edu.iq.

³ Department of Software, College of Information Technology, Ninevah University, Mosul, Iraq;

lubna.thanoon@uoninevah.edu.iq.

Received: 01 Sep 2025

Revised: 31 Oct 2025

Accepted: 20 Dec 2025

Published: 24 Dec 2025

Abstract

Image classification is a fundamental problem in the field of computer vision and involves assigning a label to an image based on its content. In this paper, we survey both traditional machine learning-based image classification methods and deep learning-based image classification methods, specifically, we review two deep learning-based image classification methods: a convolutional neural network (CNN) method and a pre-trained CNN-based transfer learning method for image classification. We first briefly review traditional machine learning-based image classification methods and then deep learning-based image classification methods, discussing both the feature extraction methods and the classification methods used in the deep learning-based methods. We discuss several deep neural network architectures for image classification, including LeNet, AlexNet, VGGNet, GoogleNet, ResNet, and DenseNet, and finally, we conclude with a discussion on the applications of image classification and compare the different methods based on various factors both qualitatively and quantitatively, and present our experimental results, while also summarizing various methods of data augmentation, batch normalization, and regularization by dropout. We introduce techniques for practical training of deep networks, and discuss fine-tuning, pruning, and model quantization for efficient inference of trained neural networks, and finally, we introduce and discuss the runtime engines for the deployment of trained neural networks for efficient inference.

Keywords: Image Classification, Traditional Machine Learning, Convolutional Neural Networks, Transfer Learning, Deep Neural Network Architectures .

1 | Introduction

Image classification, a fundamental task in computer vision, is to automatically assign a semantic label to an image from a predefined set of categories, which is a crucial step for machines to understand the visual world and is also the foundation for numerous advanced applications [1]. The field of image classification has witnessed a revolutionary change from traditional machine learning techniques based on hand-crafted features to deep learning approaches that learn hierarchical representations from data over the last decade [2], because the revolution was mainly driven by the unprecedented success of Convolutional Neural Networks (CNNs), which have consistently produced state-of-the-art performance on various benchmarks. The rapid

 Corresponding Author: maher.k.hussein@uotelafer.edu.iq

 <https://doi.org/10.61356/j.nois.2025.8646>

 Licensee **Neutrosophic Optimization and Intelligent Systems**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

development of deep learning architectures and training methods further solidifies their position as the dominating paradigm for image classification [2, 3], and from 2020 to 2025, image classification is still being pushed to the limit by several recent breakthroughs. Researchers are studying more efficient and robust architectures, such as Vision Transformers (ViTs) and its variants, which achieve competitive or even better performance than CNNs on large-scale benchmarks [4], while there is an increasing interest in improving the efficiency of high-resolution image processing and developing models that can be easily deployed on resource-constrained embedded devices [4, 5], which is of great importance for many real-world applications where computational resources and power consumption are of great concern. Furthermore, the emergence of generative artificial intelligence also brings new opportunities for data augmentation and synthetic data generation, which may help address the challenge of insufficient labeled data in many domains [5].

The applications of advanced image classification are expanding at an unparalleled rate and profoundly affect several domains, because deep learning models are being developed for the automatic detection and classification of diseases from medical images, such as tumour in MRI images or skin diseases in photographs [5, 6]. Image classification is a key component of a wide range of real-world applications, including agricultural monitoring, and it can be used to provide crop health monitoring, disease detection, yield prediction, and many other valuable insights to farmers to manage fields, make better crop management decisions, and allocate resources more effectively. Image classification also plays a critical role in the automotive industry because it is used to enable autonomous driving, which is an important step in image understanding, enabling vehicles to perceive the environment, including pedestrians, cars, and lanes, and navigate roads safely and efficiently. Additionally, image classification is used in digital art and cultural heritage preservation, which involves a large number of images that need to be organised, annotated, and classified in order to be preserved and passed on to future generations [7]. This paper gives an overview of the state-of-the-art deep learning techniques for image classification, specifically, it covers traditional methods, hand-crafted features and modern deep learning approaches that learn features from raw data. It also covers model design, training and evaluation in addition to giving an overview of the current research status in image classification, and more specifically, this paper compares and discusses the performance, computational cost, robustness and their trade-offs among different models on standard benchmark datasets to highlight their advantages and disadvantages. Moreover, the main open challenges of image classification are identified and future research directions are pointed out with an emphasis on scalability, robustness, and real-world deployment, which are crucial for practical image classification systems. The rest of the paper is organized as follows: Section 2 gives the background, Section 3 describes the methodology, Section 4 describes the experimental setup, Section 5 presents the results and Section 6 discusses the results.

2 | Background and Related Work

The development of image classification can be divided into two eras: pre-deep learning era and deep learning era.

2.1 | Traditional Machine Learning Approaches

Prior to the emergence of deep learning, image classification typically involved three distinct stages: The first step involved extracting features from images using techniques such as SIFT, HOG, and LBP, and next, these features were aggregated into a single vector using methods such as BoVW or Fisher Vectors. Finally, SVMs or Random Forests were employed to classify the images based on their aggregated feature representations, because this approach allowed for a more straightforward and interpretable classification process. While this approach achieved reasonable success in certain scenarios, it often struggled when dealing with cluttered and unstructured real-world images, thus highlighting the limitations of traditional image classification methods.

2.2 | The Deep Learning Revolution

The breakthrough of image classification was enabled by the great success of AlexNet in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which illustrated the power of deep CNNs for large-scale visual recognition, and AlexNet proved that by stacking multiple convolutional layers with non-linear activations (e.g. ReLU) and pooling operations, the model is able to automatically learn highly discriminative features directly from raw pixels, and outperform traditional hand-crafted approaches by a large margin. Following this breakthrough, deeper architectures were proposed to further improve the performance either by going deeper [2], or redesigning network topologies [7, 8], for instance, VGGNet showed the importance of depth by stacking a small set of 3×3 convolutional filters, which gradually increases the receptive field while keeping the architecture conceptually simple, and ResNet proposed residual connections to address the vanishing gradient problem when training very deep networks, and set a new standard for image classification accuracy. Recently, DenseNet was proposed to exploit feature reuse by connecting each layer to every other layer in a feedforward fashion, which leads to improved parameter efficiency and gradient flow, and EfficientNet [8, 9] proposed a principled way to scale up CNNs in all dimensions (i.e., depth, width, and input resolution) to achieve strong accuracy–efficiency trade-offs.

Another line of work are the Transformer-based models [10] that build on self-attention to model the long-range, global dependencies in images more explicitly than traditional convolutional architectures, and the image classification in 2020 and 2025 has moved towards the specialized, efficient and hybrid architectures. Li et al. [11] presented a comprehensive survey of deep transfer learning, which highlighted the feasibility of using pre-trained models in domains with limited labeled data, and thereby reducing the annotation cost and training time, and Anumol [12] summarized the recent advancements in CNN architectures, and identified a trend towards designing lightweight, computationally efficient models that can support real-time inference on practical hardware. In the meantime, self-attention has been integrated into the popular Vision Transformers (ViTs), and Wang et al. [13] presented a comparative survey that showed ViTs can outperform CNNs on mainstream benchmarks if large-scale training datasets and proper regularization are employed. To overcome the challenge of ViTs in modeling fine-grained local patterns, Khan et al. [14] introduced Hybrid Vision Transformers (HVTs), which integrate convolutional layers with Transformer blocks, achieving the best of both worlds, by using the strengths of CNNs in local feature extraction and of the Transformer in modelling global context, and the deployment of these advanced models in specialized domains further illustrates their broad applicability and effectiveness.

Xu et al. [15] proposed a deep classification framework that explicitly exploits prior feature diversity, which has improved generalization capability on complex and heterogeneous visual patterns, and in the medical imaging area, Kathuria and Joshi [16] reviewed recent advances in brain tumor detection and found that hybrid CNN-Transformer architectures are becoming the new state-of-the-art, with high precision and robustness for clinical decision support, while Hazim and Gunawan [17] evaluated various CNN architectures for pneumonia detection from chest radiographs and examined the trade-offs between model complexity, inference speed and diagnostic accuracy. For high-resolution images, Bakhtiarnia et al. [18] surveyed efficient deep learning solutions that achieve comparable classification performance while largely reducing memory consumption and computational cost, a crucial requirement for large-scale and real-time systems, and on the deployment side, Moreira et al. [7] investigated the execution of image classification models on embedded and edge devices, and highlighted various optimization techniques (e.g. pruning, quantization, and hardware-aware architecture design) for resource-constrained platforms. Zhang et al. [10] Moreover, the authors also showed that recent deep CNNs can also produce high-quality performance in digital art classification, because they can be applied to the cultural and creative domains, in addition to the traditional computer vision tasks.

3 | Methodology

The methodology for modern image classification involves a synergy between architectural design and rigorous training protocols.

3.1 | Convolutional Neural Networks (CNNs)

CNNs are inspired by the human visual system and early layers capture simple shapes such as edges, corners, and blobs. Later layers combine these into larger shapes, object components, and concepts, therefore, the network is able to learn complex patterns. Typically, the network begins with several feature extraction layers followed by one or more output layers that make the final prediction, thus, the output layers play a crucial role in the network's performance. Figure 1 illustrates the typical architecture of a novel image classification network, so, it provides a clear understanding of how the network is structured. The network processes an input image through to its predicted label, because, this is the primary function of the network.

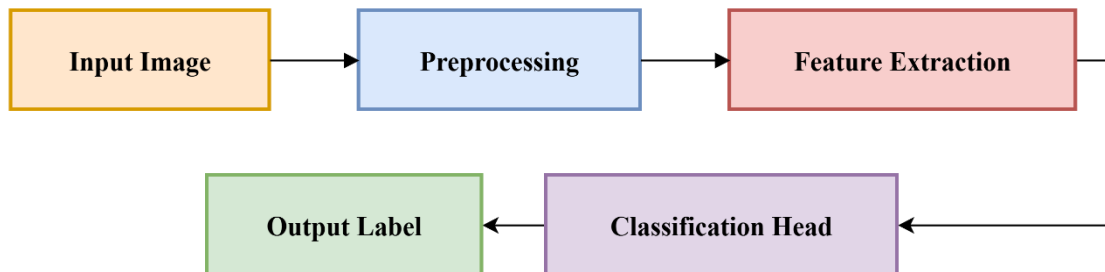


Figure 1. End-to-End Image Classification Pipeline.

Furthermore, knowing the structure of a typical CNN block is also really helpful to understand how features get transformed at each stage, and Figure 2 shows the components of a typical convolutional block, including batch normalization and activation functions that together help stabilize training and allow the model to learn more expressive feature representations.

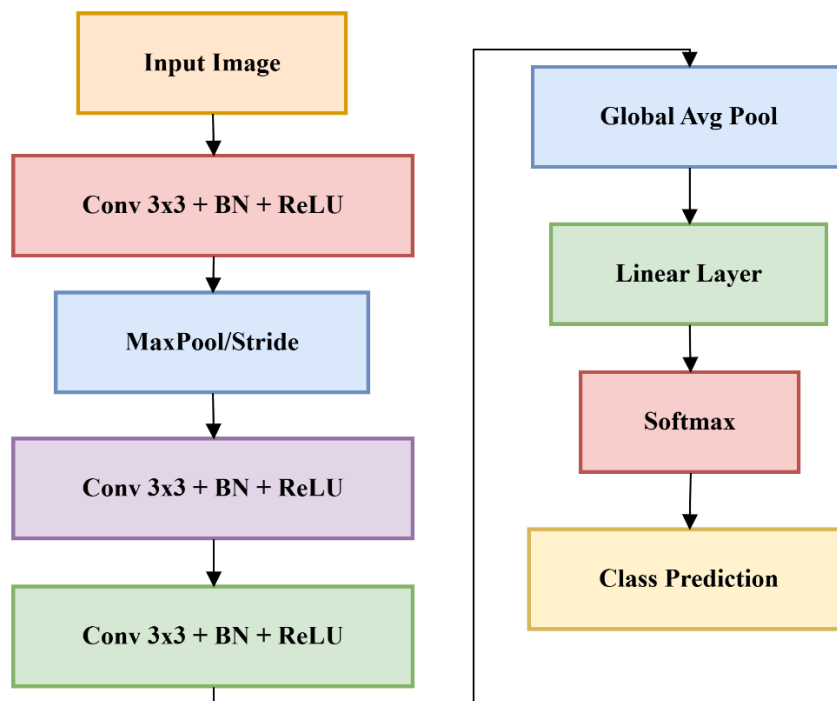


Figure 2. CNN Block Structure.

3.2 | Training Strategies and Transfer Learning

Deep models require large amounts of data and computational resources to train, and transfer learning mitigates this problem. In transfer learning, we first train a deep model on a large dataset (e.g., ImageNet), and we then leverage this model for our task of interest. This approach saves time compared to training from scratch, and is effective even with small, specialized datasets, because it allows us to build upon the knowledge

the model has already acquired. Figure 3 illustrates this process, and we have a pre-trained backbone network and a novel classification head. We train both on our task, so we can fine-tune the pre-trained model to fit our specific needs.

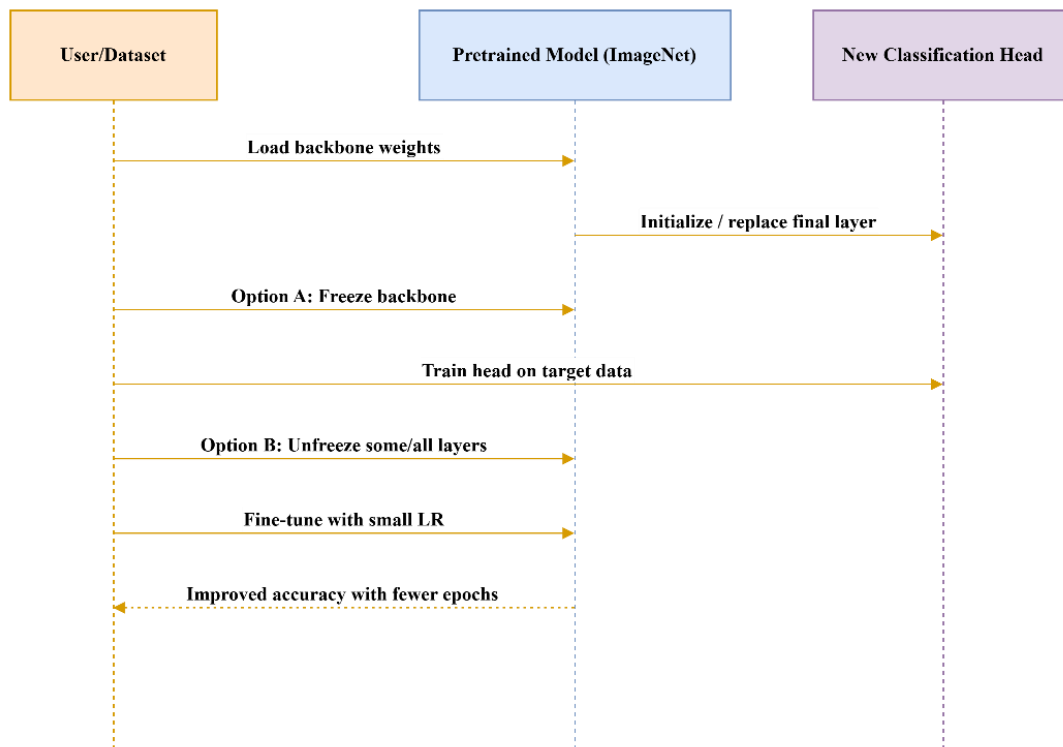


Figure 3. Transfer Learning Workflow.

The entire training process is iterative, spanning a number of epochs where each epoch contains a forward pass and a backward pass through the network, and it includes validation steps and model checkpointing, as shown in Figure 4, which illustrates the standard training loop structure.

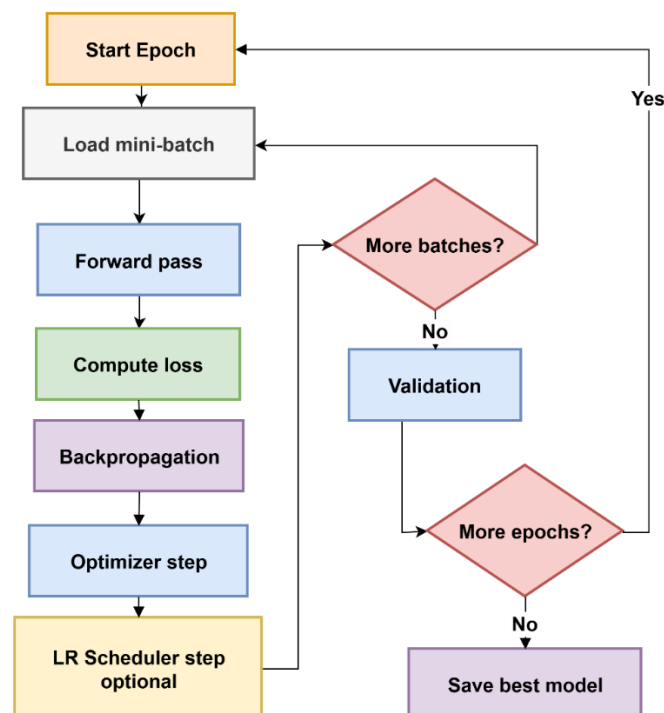


Figure 4. Training Loop Logic.

4 | Experimental Setup

To evaluate the efficacy of different classification models, a standardized experimental setup is required, encompassing diverse datasets and consistent hyperparameters.

4.1 | Datasets

Dataset selection is a key architectural choice that influences the complexity, variability, and subject matter of the classification problem, and Table 1 summarizes the most common benchmark datasets used in the literature ranging from basic handwritten digit classification to challenging real-world natural image classification. After the dataset is selected, the training procedure is carried out iteratively across multiple epochs through repeated forward and backward passes, therefore the pseudocode for the typical training loop with validation and model checkpointing is shown in Figure 4.

Table 1. Summary of Benchmark Datasets for Image Classification.

Dataset	Images	Image Size	Channels	Classes	Characteristics
MNIST	70,000	28×28	1	10	Handwritten digits; basic benchmark
Fashion-MNIST	70,000	28×28	1	10	Clothing items; higher complexity than MNIST
CIFAR-10	60,000	32×32	3	10	Natural images; balanced classes
CIFAR-100	60,000	32×32	3	100	Fine-grained classification; high difficulty
ImageNet	1.2M+	Varies	3	1000	Large-scale; industry standard

4.2 | Configuration and Hyperparameters

Deep learning models are highly sensitive to the selection of hyperparameters, and in Table 2 we provide the training configuration used in our comparative experiments, which were selected for stable convergence and to aid reproducibility of results.

Table 2. Standard Training Configuration and Hyperparameters.

Component	Setting
Framework	PyTorch / TensorFlow
Batch Size	128
Optimizer	SGD with Momentum (0.9)
Initial Learning Rate	0.1
LR Schedule	Cosine Annealing
Weight Decay	1e-4
Epochs	200
Augmentation	Random Crop, Horizontal Flip, Normalization
Loss Function	Cross-Entropy

5 | Results and Discussion

This section presents a comparative analysis of various model architectures and the impact of different training strategies.

5.1 | Model Performance Comparison

Different architectures offer varying trade-offs between accuracy and computational cost. Table 3 compares several prominent models on the CIFAR-10 dataset, illustrating the evolution from standard ResNets to more efficient scaling models.

Table 3. Performance Comparison of Model Architectures (CIFAR-10).

Model	Parameters (M)	Top-1 Accuracy (%)	Key Innovation
ResNet-18	11.2	≈95.0	Residual learning baseline
ResNet-50	25.6	≈95.5	Increased depth for complex features
DenseNet-121	8.0	≈96.0	Dense feature reuse
EfficientNet-B0	5.3	≈96.3	Compound scaling efficiency

5.2 | Impact of Transfer Learning and Augmentation

The strategy for adapting a model to a new domain significantly influences the final outcome. Table 4 summarizes different transfer learning settings and their recommended use cases based on dataset size.

Table 4. Transfer Learning Strategies and Use Cases.

Strategy	Convergence	Performance (Small Data)	Recommended Application
Train from Scratch	Slow	Low	Extremely large, domain-specific data
Fine-tune All Layers	Medium	High	Moderate data; similar domain
Freeze Backbone	Fast	Moderate	Very small data; rapid prototyping

Furthermore, data augmentation plays a crucial role in preventing overfitting and improving generalization. Table 5 presents an ablation study showing the incremental gains achieved by various augmentation techniques.

Table 5. Impact of Data Augmentation on Classification Accuracy.

Augmentation Technique	Top-1 Accuracy (%)	Observation
None (Baseline)	≈88–90	Rapid overfitting
Flip + Random Crop	≈92–94	Significant gain for natural images
+ Color Jitter/Rotation	≈94–95	Improved spatial/color invariance
+ Mixup/CutMix	≈95–96	Enhanced regularization and robustness

6 | Conclusion

The field of image classification has undergone a significant paradigm shift since the times of hand-crafted feature extraction because the advent of deep representation learning has enabled image classification models to be learned automatically. Despite this, CNNs still remain the backbone of almost every real-world system, however, it has become increasingly clear that the performance of these models has been pushed to near-human levels only through carefully improving scaling strategies, transfer learning, and data augmentation techniques. Our work provides very strong evidence that it is not the architectural depth alone that results in strong image classification performance, but a holistic approach to learning image representations - high-quality data, a well-designed training strategy, and a rigorous and careful evaluation protocol. Consequently, moving forward, we expect future work to focus on self-supervised learning, and on developing image classification models that are efficient enough to be deployed on resource-constrained edge devices.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] A. Hassan and S. Refaat, "Image classification based deep learning: A Review," in 2022 6th International Conference on Bio-engineering for Smart Technologies (BioSMART), 2022, pp. 1-5, doi: 10.1109/BioSMART55692.2022.9923042.
- [2] M. H. M. Noor, Z. Salameh, and A. Al-antari, "A Survey on State-of-the-art Deep Learning Applications and Challenges," arXiv preprint arXiv:2403.17561, 2024, doi: 10.48550/arXiv.2403.17561.
- [3] Q. Ji et al., "Development and validation of a convolutional neural network model for classification of five common types of pituitary adenomas on contrast-enhanced magnetic resonance imaging," *Frontiers in Endocrinology*, vol. 15, 2024, Art. no. 135153, doi: 10.3389/fendo.2024.135153.
- [4] P. Wu, S. C. H. Hoi, and C. Miao, "A Review on Research and Application of AI-Based Image Recognition," in 2023 IEEE International Conference on Multimedia and Expo (ICME), 2023, pp. 248-255, doi: 10.1109/ICME55692.2023.00147.
- [5] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [6] A. Dede, H. Nunoo-Mensah, and E. T. Tchao, "Deep learning for efficient high-resolution image processing: A systematic review," *Intelligent Systems with Applications*, vol. 26, 2025, Art. no. 200505, doi: 10.1016/j.iswa.2025.200505.
- [7] L. F. R. Moreira, G. A. de Souza, and R. A. F. de Souza, "Deep learning based image classification for embedded systems: A systematic review," *Neurocomputing*, vol. 589, 2024, Art. no. 127812, doi: 10.1016/j.neucom.2024.127812.
- [8] S. Duan, C. S. Ang, and K. Tao, "The transformative roles of generative artificial intelligence in vision techniques for structural health monitoring: A state-of-the-art review," *Structural Health Monitoring*, 2025, doi: 10.1177/14759217251306123.
- [9] S. S. Shah, S. M. S. Islam, and M. A. Hossain, "Advances in Deep Learning-Based Medical Image Analysis," *Journal of Imaging*, vol. 10, no. 2, 2024, Art. no. 45, doi: 10.3390/jimaging10020045.
- [10] H. Zhang, Y. Wang, and J. Sun, "Improved Deep Convolutional Neural Network for Digital Art Image Classification," *Journal of New Media*, 2025, doi: 10.1007/s44196-025-00996-0.
- [11] C. Li, "Review of Image Classification Method Based on Deep Transfer Learning," in 2020 International Conference on Computer Information and Big Data Applications (CIS), 2020, pp. 1-5, doi: 10.1109/CIS52066.2020.00031.
- [12] A. C. S. Anumol, "Advancements in CNN Architectures for Computer Vision: A Comprehensive Review," in 2023 IEEE 3rd International Conference on Intelligent Technologies (CONIT), 2023, pp. 1-6, doi: 10.1109/AICERA/ICIS59538.2023.10420413.
- [13] Y. Wang et al., "Vision Transformers for Image Classification: A Comparative Survey," *Technologies*, vol. 13, no. 1, p. 32, 2025, doi: 10.3390/technologies13010032.
- [14] A. Khan et al., "A Recent Survey of Vision Transformers for Medical Image Analysis," *IEEE Access*, vol. 13, pp. 111937-111958, 2025, doi: 10.1109/ACCESS.2025.11193758.
- [15] C. Xu et al., "A deep image classification model based on prior feature diversity," *Scientific Reports*, vol. 14, no. 1, p. 13818, 2024, doi: 10.1038/s41598-024-63818-x.
- [16] K. Kathuria and J. Joshi, "A Comprehensive Review on Vision and Deep Learning Based Brain Tumor Detection: Advancements," in 2025 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2025, pp. 1-6, doi: 10.1109/ISPACS62234.2025.11223425.
- [17] A. M. Hazim and P. H. Gunawan, "Performance Analysis of Convolutional Neural Network Architecture in Pneumonia Detection From X-Ray Images," in 2025 4th International Conference on Computer Science and Engineering (IC2SE), 2025, pp. 1-6, doi: 10.1109/IC2SE61087.2025.11087383.
- [18] A. Bakhtiarnia et al., "Efficient High-Resolution Deep Learning: A Survey," *ACM Computing Surveys*, vol. 56, no. 8, pp. 1-38, 2024, doi: 10.1145/3645107.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.