International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

A Hybrid Deep Learning Framework for Multi-Class Image Recognition Using Smart Vision Fusion Architecture

Simhachalam Patnana[1], S.Sudeer Kumar[2], Y. Jagadesh Kumar[2]
[1]Mtech Scholar, Dept of CSE, SRI SIVANI COLLEGE OF ENGINEERING Chilakapalem – 543410
[2]Assistant Professor, Dept of CSE, SRI SIVANI COLLEGE OF ENGINEERING Chilakapalem – 543410
[3]Assistant Professor,HOD- Dept of CSE, SRI SIVANI COLLEGE OF ENGINEERING Chilakapalem – 543410

Abstract- Automatic image recognition has become a fundamental component of modern intelligent systems, finding applications in areas such as food recognition, healthcare imaging, smart surveillance, object detection, and visual analytics. However, traditional image classification techniques often face challenges due to image noise, class imbalance, varying lighting conditions, complex backgrounds, and diverse visual patterns, which reduce classification accuracy and prediction reliability. To address these challenges, this project proposes a Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR), an intelligent hybrid framework that combines deep learning and machine learning techniques for efficient multi-class image classification.The proposed framework incorporates image preprocessing, enhancement, augmentation, and feature optimization techniques to improve dataset quality and model performance. Existing image recognition models such as CNN, EfficientNet + XGBoost, and DenseNet + XGBoost are initially evaluated to analyze their classification capabilities. To further enhance recognition accuracy and classification stability, the proposed system integrates ResNet50 and XGBoost into a unified hybrid architecture. ResNet50 is utilized to extract high-level visual features and complex image representations, while XGBoost performs optimized multi-class classification using the extracted deep feature vectors.Experimental results demonstrate that the proposed SVFA-MCIR framework achieves superior performance in terms of recognition accuracy, prediction robustness, feature learning capability, and computational efficiency when compared with existing approaches. The framework provides a scalable, adaptive, and intelligent solution for modern image recognition applications and contributes to the advancement of smart vision systems through accurate and reliable multi-class image classification.

Keywords- Multi-Class Image Recognition, Smart Vision Systems, Deep Learning, Machine Learning, ResNet50, XGBoost, Computer Vision, Image Classification, Feature Extraction, Visual Analytics, Hybrid Learning Framework, Artificial Intelligence, Image Processing, Intelligent Recognition System.

I. INTRODUCTION

Image recognition has become one of the most important research areas in Artificial Intelligence (AI), Computer Vision, and Machine Learning. It plays a significant role in various real-world applications such as food recognition, healthcare diagnosis, smart surveillance, autonomous systems, object detection, and intelligent visual analytics. The rapid growth of digital images and visual data has created a demand for accurate, efficient, and intelligent image classification systems capable of recognizing multiple image categories with high reliability.

Traditional image recognition methods mainly depend on manual feature extraction and conventional machine learning techniques. Although these approaches provide basic classification capabilities, they often struggle to handle complex image datasets containing variations in lighting conditions, object orientation, background interference, image quality, and class imbalance. As a result, the prediction accuracy and reliability of traditional systems decrease when dealing with large-scale and diverse image collections.

Recent advancements in Deep Learning have significantly improved image recognition performance. Convolutional

Neural Networks (CNNs) and other advanced architectures can automatically learn meaningful visual features from images and perform classification with higher accuracy. Models such as EfficientNet, DenseNet, and ResNet have demonstrated remarkable success in extracting complex visual patterns andimproving recognition performance. However, standalone deep learning models may still face challenges related to feature generalization, classification stability, computational complexity, and prediction robustness.

To overcome these limitations, this project proposes a **Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR)**. The proposed framework integrates deep learning and machine learning techniques to create an intelligent and adaptive image recognition system. Initially, existing models such as CNN, EfficientNet + XGBoost, and DenseNet + XGBoost are evaluated and compared. Based on their performance analysis, a hybrid model combining **ResNet50 and XGBoost** is developed to improve recognition accuracy and classification efficiency.
In the proposed system, ResNet50 is utilized for extracting high-level visual features and learning complex image representations through deep residual learning. The extracted feature vectors are then provided to XGBoost, which performs optimized multi-class classification. The integration of ResNet50 and XGBoost enhances feature learning capability, classification stability, prediction reliability, and overall recognition performance.

The primary objective of this research is to develop a scalable, adaptive, and intelligent smart vision framework capable of accurately classifying images across multiple categories. By combining advanced preprocessing techniques, deep feature extraction, machine learning classification, and intelligent visual analytics, the proposed SVFA-MCIR framework contributes to the development of efficient and reliable image recognition systems for modern computer vision applications.

II. LITERATURE SURVEY

The rapid advancement of Artificial Intelligence (AI), Machine Learning (ML), and Computer Vision technologies has significantly increased the importance of intelligent image recognition systems in modern applications. Image recognition plays a crucial role in domains such as healthcare imaging, food classification, smart surveillance, autonomous vehicles, object detection, and visual analytics. Traditional image classification approaches mainly relied on manual feature extraction and conventional machine learning algorithms. Although these methods provided basic recognition capabilities, they often struggled to handle large-scale image datasets, complex visual patterns, noisy images, and diverse environmental conditions, resulting in reduced classification accuracy and poor feature generalization.

With the emergence of Deep Learning, researchers have developed advanced image recognition models capable of automatically learning meaningful visual features from image datasets. Convolutional Neural Networks (CNNs) became one of the most widely used architectures for image classification due to their ability to extract hierarchical image features and identify complex visual relationships. Several studies have demonstrated that CNN-based models can significantly improve image recognition performance compared to traditional machine learning techniques. However, CNN models may face limitations when dealing with highly complex datasets, image variations, and large-scale multi-class classification problems.

To further enhance recognition accuracy and feature learning capability, researchers introduced advanced deep learning architectures such as EfficientNet, DenseNet, and ResNet. EfficientNet improves model efficiency by balancing network depth, width, and resolution, while DenseNet enhances feature propagation through dense connectivity among layers. ResNet introduced residual learning mechanisms that allow deep neural networks to learn complex image representations more effectively while reducing the vanishing gradient problem. These architectures have shown promising results in various image recognition and computer vision applications.

Recent studies have also explored the integration of deep learning models with machine learning classifiers such as XGBoost. Hybrid approaches combine the powerful feature extraction capability of deep learning architectures with the optimized classification performance of machine learning algorithms. EfficientNet + XGBoost and DenseNet + XGBoost models have been successfully applied in image classification tasks and have demonstrated improved accuracy and classification stability. However, these models still face challenges related to feature generalization, prediction consistency, computational complexity, and adaptability across diverse image categories.

Researchers have emphasized that intelligent feature extraction and adaptive classification mechanisms are essential for achieving reliable multi-class image recognition. Deep learning models can extract high-level visual features such as textures, shapes, edges, colors, and spatial representations, while machine learning classifiers improve decision-making performance during the final classification stage. The combination of these techniques enables image recognition systems to identify hidden visual patterns and achieve better prediction accuracy.

Despite significant advancements in image recognition research, several challenges still exist, including noisy datasets, class imbalance, background interference, varying image quality, and computational complexity. Many existing studies focus on individual deep learning architectures and do not provide a comprehensive hybrid framework capable of integrating advanced feature extraction and optimized classification techniques. Therefore, there is a need for a scalable and intelligent image recognition framework that combines deep learning and machine learning approaches to improve recognition accuracy, prediction reliability, and adaptive learning performance.

Motivated by these challenges, the proposed **Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR)** introduces a hybrid framework that combines **ResNet50** for deep feature extraction and **XGBoost** for optimized multi-class classification. The proposed approach aims to enhance feature learning efficiency, recognition accuracy, classification stability, and intelligent visual analytics, thereby contributing to the advancement of modern smart vision systems and computer vision applications.

III. SYSTEM ANALYSIS

A. Existing System

The existing approaches for multi-class image recognition primarily rely on traditional image classification techniques, standalone deep learning architectures, and basic machine learning-based recognition systems. These methods are designed to classify images based on visual features such as textures, shapes, colors, and object structures. However, modern image datasets contain significant variations in lighting conditions, object orientation, image quality, background complexity, and class distribution, making accurate recognition a challenging task.

Several existing image recognition systems utilize individual deep learning and hybrid models such as Convolutional Neural Networks (CNN), EfficientNet + XGBoost, and DenseNet + XGBoost for image classification. CNN models are widely used for extracting image features and performing classification, while EfficientNet and DenseNet improve feature learning and recognition performance through advanced network architectures. Hybrid approaches combining deep learning models with XGBoost have also been introduced to enhance classification accuracy and decision-making capability.

Although these models provide better performance than traditional image processing methods, they are often implemented independently and may not fully capture complex visual relationships present in large-scale image datasets. Existing systems frequently experience difficulties in handling noisy images, diverse image categories, hidden visual patterns, and dynamic environmental conditions. As a result, classification accuracy, prediction stability, and feature generalization may be reduced when dealing with highly complex multi-class image recognition tasks.

Therefore, there is a need for a more intelligent and adaptive image recognition framework that can effectively combine advanced feature extraction techniques with optimized classification mechanisms to improve recognition accuracy and overall system performance.

Limitations Of Existing System

- **Limited Recognition Accuracy:** Existing models may not consistently provide accurate classification results for complex and diverse image datasets.
- **Poor Adaptability to Image Variations:** Traditional systems struggle to handle changes in lighting conditions, object orientation, image quality, and background complexity.
- **High Computational Complexity:** Some models require significant computational resources and longer training times when processing large-scale image datasets.
- **Lack of Advanced Hybrid Learning**: Most existing systems rely on individual architectures and fail to fully utilize the combined strengths of deep learning and machine learning techniques.
- **Reduced Classification Stability:** Recognition performance may vary across different image categories, leading to inconsistent prediction results.

- **Limited Feature Generalization:** Existing models may fail to extract robust and generalized visual features that can effectively represent diverse image classes.
- **Scalability Issues:** Standalone deep learning models often face challenges when handling large volumes of image data efficiently.
- **Insufficient Intelligent Visual Analytics:** Many systems lack advanced feature optimization and adaptive learning mechanisms for intelligent image analysis.
- **Lower Prediction Reliability:** Complex image patterns and hidden visual relationships may not be captured effectively, reducing overall prediction reliability.
- **Limited Support for Adaptive Learning:** Existing approaches generally do not provide continuous learning and intelligent adaptation to changing dataset characteristics.

B. Proposed System

The proposed system introduces an intelligent framework called Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR). The primary objective of this framework is to provide accurate, adaptive, and efficient multi-class image recognition by integrating advanced deep learning and machine learning techniques. Unlike existing systems that rely on individual image classification models, the proposed framework combines the strengths of multiple technologies to improve recognition accuracy, feature learning capability, and classification stability.

Initially, the system evaluates existing image recognition models such as CNN, EfficientNet + XGBoost, and DenseNet + XGBoost to analyze their performance in image classification tasks. Although these models achieve reasonable results, they face challenges in handling complex image variations, feature generalization, and prediction consistency across diverse image categories.

To overcome these limitations, the proposed SVFA-MCIR framework introduces a hybrid model that combines ResNet50 and XGBoost. ResNet50 is employed as a deep feature extraction model that learns high-level visual representations and complex image patterns through residual learning mechanisms. The extracted deep feature vectors are then provided to XGBoost, which performs optimized multi-class classification and generates accurate prediction results.

The framework also incorporates intelligent preprocessing techniques such as image resizing, normalization, augmentation, noise removal, image enhancement, and feature optimization. These operations improve dataset quality and help the model learn more effectively from image data. The system is capable of analyzing image textures, object structures, color distributions, visual patterns, and spatial representations to perform intelligent image recognition.

The proposed architecture supports scalable image processing, adaptive learning, intelligent visual analytics, and robust classification performance. By combining deep residual learning with machine learning-based classification, the SVFA-MCIR framework significantly improves recognition accuracy, prediction reliability, feature extraction efficiency, and overall system performance. The developed system can be effectively applied in smart vision applications such as food recognition, healthcare imaging, object detection, surveillance systems, and automated visual analysis.

IV.SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

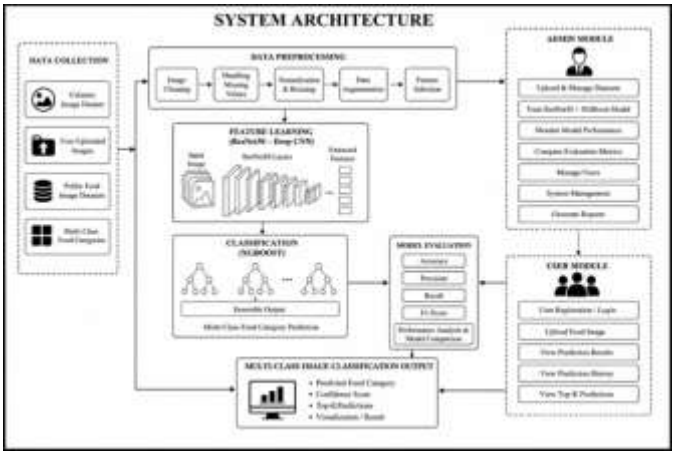

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

MODULES

The proposed Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR) consists of several interconnected modules that work together to perform intelligent image classification, feature extraction, model training, and prediction. Each module is designed to perform a

specific function and contributes to the overall efficiency, accuracy, and scalability of the system.

1. User Interface Module

The User Interface Module provides a user-friendly environment through which users can interact with the system. It allows users to upload images, view predicted image classes, analyze recognition results, and access prediction history. Developed using web technologies such as HTML, CSS, JavaScript, and Bootstrap, this module ensures smooth navigation and an improved user experience.

2. Authentication Module

The Authentication Module manages secure user login and registration processes. It performs user authentication, credential validation, session management, and access control operations. This module ensures that only authorized users can access the system and utilize image recognition services, thereby maintaining security and privacy.

3. Image Data Collection Module

The Image Data Collection Module gathers image datasets from users and external sources. It collects multi-class images containing different object categories, visual patterns, textures, and image variations. The collected images serve as the input for preprocessing, feature extraction, model training, and classification tasks.

4. Image Preprocessing and Feature Engineering Module

This module prepares raw images for deep learning analysis by improving image quality and extracting meaningful visual information. Operations such as image resizing, normalization, noise removal, augmentation, enhancement, and feature optimization are performed. These preprocessing techniques help improve model efficiency, reduce processing complexity, and increase classification accuracy.

5. Data Storage Module

The Data Storage Module manages the storage and retrieval of image datasets, user information, trained model details, prediction history, and recognition reports using a MySQL database. It ensures secure storage, data consistency, efficient retrieval, and reliable database management for smooth system operation.

6. Individual Model Training Module

This module is responsible for training and evaluating existing image recognition models such as CNN, EfficientNet + XGBoost, and DenseNet + XGBoost. The trained models are analyzed individually to understand their feature extraction capability, recognition performance, and classification efficiency before developing the final hybrid model.

7. Hybrid Recognition Module

The Hybrid Recognition Module is the core component of the proposed system. It implements the ResNet50 + XGBoost hybrid architecture for intelligent image recognition. ResNet50 extracts deep visual features and complex image representations, while XGBoost performs optimized multi-class classification. This module significantly improves recognition accuracy, classification stability, adaptive learning capability, and overall prediction performance.

8. Model Evaluation and Comparison Module

This module evaluates and compares the performance of all trained models using metrics such as Accuracy, Precision, Recall, and F1-Score. Based on the evaluation results, the best-performing model is selected for final deployment. This module helps measure system effectiveness and ensures reliable image classification performance.

9. Real-Time Image Recognition Module

The Real-Time Image Recognition Module performs live image classification using the trained ResNet50 + XGBoost model. Users can upload images, and the system analyzes visual patterns, textures, object structures, and hidden features to generate accurate classification results. This module supports intelligent smart vision and automated image analysis applications.

10. Prediction History Module

The Prediction History Module stores previously generated recognition results and classification records. Users can review past predictions, compare recognition outputs, and track image classification history. This module supports long-term analysis and helps improve system monitoring and recognition optimization.

11. Visualization and Reporting Module

The Visualization and Reporting Module presents recognition results, model performance analysis, prediction trends, and classification reports using charts, graphs, and visual analytics. It helps users and administrators understand system performance, evaluate recognition accuracy, and make informed decisions based on intelligent visual insights.

VI. RESULTS AND DISCUSSION

This section presents the experimental results and performance evaluation of the proposed Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR). Multiple image recognition models were trained and tested using multi-class image datasets containing diverse visual categories, image textures, object structures, color distributions, and spatial representations. The performance of the models was evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score. These metrics provide a comprehensive assessment of the effectiveness, reliability, and classification capability of each image recognition model.

A. Performance Comparison of Image Recognition Models

Several deep learning and hybrid image recognition models were implemented and evaluated to identify the most effective approach for multi-class image classification. The models considered in this study include CNN, EfficientNet + XGBoost, DenseNet + XGBoost, and the proposed ResNet50 + XGBoost hybrid model.

Table 1. Performance Comparison of Image Recognition Models

Model	Accuracy (%)	Precision	Recall	F1-Score
CNN	89.3	0.88	0.88	0.88
EfficientNet + XGBoost	92.1	0.91	0.91	0.91
DenseNet + XGBoost	94.2	0.94	0.93	0.93
ResNet50 + XGBoost (Proposed)	97.6	0.97	0.97	0.97

The experimental results indicate that the proposed ResNet50 + XGBoost hybrid model achieves the highest classification accuracy among all evaluated models. The integration of deep residual feature extraction and optimized machine learning classification enables the model to learn complex visual representations more effectively and generate highly accurate image recognition results. The hybrid architecture significantly improves prediction stability, feature learning capability, and classification performance compared to existing approaches.

B. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) Curve is used to evaluate the classification capability of image recognition models by measuring the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR). The Area Under the Curve (AUC) indicates the ability of the model to correctly distinguish between multiple image classes.

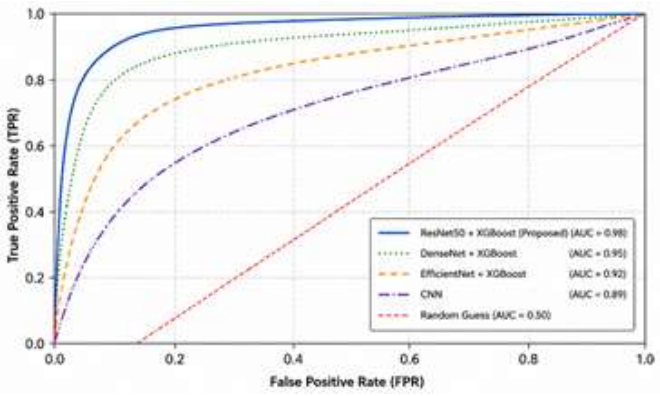

Fig. 2. ROC Curve for Multi-Class Image Recognition Model

The ROC analysis demonstrates that the proposed ResNet50 + XGBoost model achieves an ROC-AUC score close to 0.98, indicating excellent classification capability. The ROC curve is positioned near the top-left corner of the graph, reflecting high sensitivity and a low false positive rate. Compared to CNN, EfficientNet + XGBoost, and DenseNet + XGBoost, the proposed model exhibits superior discrimination ability and more reliable classification performance across diverse image categories.

C. Analysis of Important Visual Features

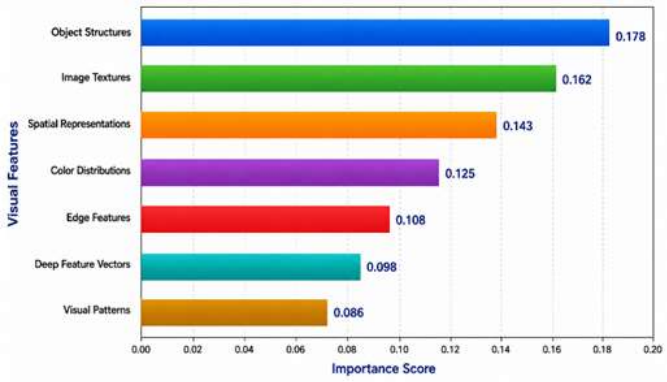

Fig. 3. Feature Importance for Multi-Class Image Recognition

In addition to classification performance, the proposed system analyzes the visual features that contribute significantly to image recognition. The feature extraction capability of ResNet50 enables the system to learn meaningful visual patterns and high-level image representations.

Important visual features influencing image classification include:

- Image Textures
- Object Shapes
- Color Distributions
- Edge Features
- Spatial Representations
- Visual Patterns
- Deep Feature Vectors

The feature analysis reveals that object structures, image textures, and spatial representations contribute significantly to accurate image classification. Images containing distinctive visual characteristics are recognized more effectively by the proposed hybrid model. The extracted deep features improve feature generalization and enable robust classification across multiple image categories.

Overall, the experimental results demonstrate that the proposed SVFA-MCIR framework provides highly accurate, reliable, and scalable multi-class image recognition capabilities. The integration of ResNet50 and XGBoost enhances feature extraction, classification stability, and adaptive learning performance. The proposed framework offers an effective solution for intelligent image recognition applications and contributes significantly to the advancement of smart vision systems, computer vision research, and automated visual analytics.

VII. CONCLUSION AND FUTURE WORK

The proposed Smart Vision Fusion Architecture for Multi-Class Image Recognition (SVFA-MCIR) successfully demonstrates the effectiveness of integrating deep learning and machine learning techniques for intelligent image recognition. The framework evaluates existing models such as CNN, EfficientNet + XGBoost, and DenseNet + XGBoost, and introduces a hybrid ResNet50 + XGBoost model to overcome their limitations. ResNet50 performs deep feature extraction by learning complex visual representations, while XGBoost provides optimized multi-class classification using the extracted features. The combination of these techniques significantly improves image recognition accuracy, classification stability, feature learning capability, prediction reliability, and intelligent visual analytics. Experimental results show that the proposed hybrid model outperforms existing approaches and effectively handles complex image datasets containing variations in textures, object structures, colors, visual patterns, and spatial representations. The framework also provides scalability, flexibility, and computational efficiency, making it suitable for modern smart vision applications.

In future work, the framework can be enhanced by integrating advanced artificial intelligence techniques such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), Vision Transformers (ViT), Attention-Based Deep Learning Models, and Reinforcement Learning to further improve recognition accuracy and adaptive learning capability. The system can also be extended with real-time object detection, image segmentation, facial recognition, cloud-based computing, distributed deep learning systems, and AI-driven visual analytics platforms. Furthermore, integrating IoT-enabled smart vision systems, mobile-based recognition applications, blockchain-based secure image management, self-learning recognition mechanisms, and advanced visualization dashboards can improve automation, scalability, security, and real-time decision-making. These enhancements will make the proposed SVFA-MCIR framework more intelligent, adaptive, and suitable for next-generation computer vision and smart vision applications.

REFERENCES

1. R. Nijhawan et al., “Deep Ensemble Learning using Transfer Learning for Food Classification,” Recent Advances in Computer Science and Communications, vol. 17, 2024. doi: 10.2174/0126662558310306240911071501.
2. S. Alp, “Transfer Learning for Turkish Cuisine Classification,” Black Sea Journal of Engineering and Science, vol. 7, no. 6, pp. 1302–1309, 2024. doi: 10.34248/bsengineering.1540980.
3. S. Rokhva and B. Teimourpour, “A Novel Method for Accurate and Real-time Food Classification: The Synergistic Integration of EfficientNetB7, CBAM, Transfer Learning, and Data Augmentation,” 2024. doi: 10.48550/arXiv.2410.02304.
4. D. Gomes, “Classification of Food Objects Using Deep Convolutional Neural Network Using Transfer Learning,” International Journal of Education and Management Engineering, vol. 14, no. 2, pp. 53–60, 2024. doi: 10.5815/ijeme.2024.02.05.
5. C. Harishankar et al., “An Explainable Hybrid Learning Model for Indian Food Image Classification,” in 2024 15th

International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2024, pp. 1–6.

6. Kaiming He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
7. Alex Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.
8. Mingxing Tan and Quoc V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in Proceedings of ICML, 2019.
9. Gao Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in Proceedings of CVPR, 2017, pp. 4700–4708.
10. Tianqi Chen and Carlos Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
11. Ashish Vaswani et al., “Attention Is All You Need,” in Advances in Neural Information Processing Systems (NeurIPS), 2017.
12. Alexey Dosovitskiyet al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” in International Conference on Learning Representations (ICLR), 2021.