

## Research Article

# AI-Enable Rice Image Classification Using Hybrid Convolutional Neural Network Models

**Daya Shankar Verma** <sup>1</sup>, **Mrinal Dafadar** <sup>2</sup>, **Jitendra K. Mishra** <sup>3</sup>, **Ankit Kumar** <sup>4</sup>,  
**and Shambhu Mahato** <sup>5</sup>

<sup>1</sup>Department of Computer Science and Engineering, Indian Institute of Information Technology Ranchi, Ranchi, India

<sup>2</sup>Department of Computer Science and Engineering, Birla Institute of Technology Mesra, Ranchi, India

<sup>3</sup>Department of Electronics and Communication Engineering, Indian Institute of Information Technology Ranchi, Ranchi, India

<sup>4</sup>Department of Information Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur, India

<sup>5</sup>Department of Computer Science and Engineering, Janajyoti Multiple Campus, Bhimdatta, Nepal

Correspondence should be addressed to Shambhu Mahato; [shambhu.mahato@jjmc.edu.np](mailto:shambhu.mahato@jjmc.edu.np)

Received 2 March 2024; Revised 7 December 2024; Accepted 30 December 2024

Academic Editor: Alexander Hošovský

Copyright © 2025 Daya Shankar Verma et al. International Journal of Intelligent Systems published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Rice is the most preferred grain worldwide, leading to the development of an automated method using convolutional neural networks (CNNs) for classifying rice types. This study evaluates the effectiveness of hybrid CNN models, including AlexNet, ResNet50, and EfficientNet-b1, in distinguishing five major rice varieties grown in Turkey: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. It is estimated that there are 75,000 photographs of grains, with 15,000 images corresponding to each type. The training is improved by the use of preprocessing and optimization approaches. The performance of the model was assessed based on sensitivity, specificity, precision, F1 score, and confusion matrix analysis. The results show that EfficientNet-b1 achieved an accuracy of 99.87%, which is higher than the accuracy achieved by AlexNet (96.00%) and ResNet50 (99.00%). This study shows that EfficientNet-b1 is superior to other models that have emerged as state-of-the-art automated classification models for rice varieties. This indicates that there is a balance between the computational efficiency and the accuracy of EfficientNet-b1. These results exemplify the potential of CNN models for agriculture by reducing the restrictions associated with conventional classification approaches. These limitations include subjectivity and inconsistency regarding categorization.

**Keywords:** AlexNet; CNN; confusion matrix; EfficientNet-b1; ResNet50; sensitivity

## 1. Introduction

In addition to being the most widely grown cereal crop on the planet, rice is an indispensable grain that plays a significant role in meeting the economic and nutritional requirements of both prosperous and impoverished regions. One of the primary reasons for this is that it has a high concentration of starch, which is the primary component of carbon. Rice grains have exceptional nutritional and industrial characteristics, which is why collection, classification, and separation are crucial in the rice processing industry. A problem associated with conventional

techniques for the separation of rice grains from chaff, sizing, and grading of kernels individually on size, shape, and other physical attributes, followed by grouping based on color and texture, is that these techniques are inefficient, inhomogeneous, and inaccurate. These errors need to be rectified to protect the entire realm of food security, quality, and integrity in supply chains.

Based on these challenges and the need to build appropriate solutions for the classification and quality assessment of rice varieties, a significant number of the technologies involved, particularly machine learning, deep learning, and image processing, have been shown to be

extremely suitable in this regard. The new methods provide an enhanced answer to the difficult and time-consuming chores that can now be mechanized and carried out with accuracy and efficiency while requiring only a small amount of physical effort and with low financial investment. In fact, these methods have unquestionably shown their immense value in terms of accelerating the categorization of food products and in terms of conserving valuable resources with respect to the management of supply chains and food safety.

In this area of research, most recent publications have relied on very large datasets that have been manually labeled with binary information and consist of many classes. Some of these datasets include well-known models or algorithms, such as convolutional neural networks (CNNs), SVMs, and LR, in addition to enormous portfolios that contain over hundreds of thousands of photos of rice grains that are sharply focused [1–3]. RGB analysis, hyperspectral imaging, and NIR image fusion approaches have been highlighted as highly investigated image processing methods that have the potential to enhance the categorization of dataset pictures, according to other advancements in the study. These approaches demonstrated a significant advantage in terms of accuracy in the identification of rice varieties, with results of approximately 98.50% with SVM and 93.02% with LR on a variety of datasets derived from numerous studies. Studies have shown that the use of machine learning and deep learning has the potential to promote improvements in agricultural practices, notably in the classification of rice and all other crops. The result is an increase in both the efficiency of commodity systems and the level of food security.

Across various parts of the globe, the most consumable, staple, and producible grain at a mass level is rice. The journey of rice-grown products is initiated from the paddy-grown field from the houses of all high society to poor people, which proves to be economical in our country as well as across the whole world. The reason for the large consumable grain product is its nutritious value, which is very rich in starch and many carbohydrates, which is helpful for maintaining the body energy of an individual. Owing to its importance, it is also widely used in the industrial field. During the harvest of crops in the field, the task of separating rice grains from undesirable materials is performed.

In the classification process, the solid shapes of the rice grains and their broken pieces are separated. Finally, the extraction task is performed based on its color, textures, shapes, etc., which are the features being extracted. These processes are traditionally performed. The traditional way of committing this entire process is performed by human interventions, which leads to slow, inconsistent, subjective, and unreliable processes, which also tends to lead to inaccurate classification and extraction of the rice grain product, which will hamper its consumable values among human beings apart from being rich in nutritious value qualities. As a result, this will also lead to an imbalance in food security. The degradation of the physical qualities of rice grain will also have a huge impact on the supply chain, which will lead to starvation and various illnesses among human beings across the world. To overcome the huge challenges of the traditional ways of classifying the various

rice varieties, various continuous innovations and advancements in technology have paved the way to revolutionizing the entire outdated processes of classification. Due to the long-term processes involved in traditional and manual approaches, modern technologies are gradually revolutionizing them to classify different varieties of rice and other grains grown on the planet. The implementation of various deep learning models, machine learning models, and image processing techniques in the field of rice variety image classification has emerged as an alternative path for accurate classification, which has resulted in brighter outcomes in the past few years. As modern techniques are considered an alternative way to automate the entire classification process for the benefits of accurate results and cost-effectiveness, various studies using image processing techniques on various grains and other vegetables have already been examined on the basis of their physical characteristics, such as texture, shape, and color. These various studies on the classification of rice and other final product “varieties” in the literature are broadly summarized.

Various applications of image processing and deep learning techniques in the agriculture sector have surpassed the traditional methods of rice variety classification [2, 4]. Applications based on deep learning models for image processing provide benefits in quality assessment, grading, and market segmentation in comparison to traditional methods, which are quite time-consuming in rice classification [5]. If the rice classification is to be performed on a large scale, then the traditional methods used for this purpose are quite complex, and these processes vary with different individuals, and the decisions that are expected are unsatisfactory [6]. Therefore, the machine learning approach is widely used to scale the productivity of various grains, fruits, and vegetables by focusing on the quality of the product to safeguard the food safety criteria in the whole world [7–9]. Current studies have extensively used rice and wheat image datasets consisting of millions of high-quality photos for image classification. The ideal methods for enhancing the quality of dataset photographs are the RGB method and the hyperspectral data processing approach [10].

The research conducted by the authors in [11] included a collection of approximately 200 rice images. The dataset consisted of three classes and achieved an accuracy of 88 using the CNN model. When evaluated using the SVM method, a 98.50% accuracy rate was obtained with two classes of 1700 distinct data [12]. In a separate study conducted by the authors in [13], a dataset consisting of 3810 photos was gathered. The dataset was categorized into two classes, achieving an accuracy of 93.02% using the LR algorithm. Building upon similar studies, distinct classes of rice grains were analyzed using datasets of 1500 and 3200 images for training and testing. These were processed with SRC algorithm packages in LMNN2, SR algorithms implemented in MATLAB R2018a, feedforward neural networks with RMANOVA algorithm codes in MATLAB R2020a, and multiclass SVM using CSV files via Python. The respective accuracy rates achieved were 89.1%, 86.25%,

84.83%, and 86% [14–16]. These studies are summarized in Table 1. In another study, five different varieties of rice grains with 75,000 pieces of data with 15,000 of each variety of rice image datasets were gathered, and ANN, DNN, and CNN algorithms were applied to obtain the outcomes with higher accuracy levels [21].

CNNs have emerged as powerful tools for image classification tasks in various fields, including agriculture. Recent advancements in deep learning, particularly in CNN architectures, have revolutionized the ability to classify complex datasets. Models such as AlexNet, ResNet50, and EfficientNet-b1 have demonstrated high accuracy in various applications owing to their unique architectural features. However, the key to achieving optimal performance lies not only in employing these models but also in introducing innovative modifications to their architecture and hyperparameters.

Although existing studies have shown that CNNs can classify agricultural products with high accuracy, the field of rice variety classification still presents challenges, owing to the visual similarities among different varieties. In recent years, research has increasingly focused on the potential of CNNs combined with advanced image processing techniques to address these challenges [2, 3]. Nonetheless, few studies have explored the impact of hyperparameter tuning, transfer learning, and data augmentation on the model performance in rice variety classification.

This study aims to bridge this gap by applying CNN models such as AlexNet, ResNet50, and EfficientNet-b1 to rice variety classification while emphasizing the importance of hyperparameter tuning and advanced augmentation techniques. We also explore a novel approach of fine-tuning these models with transfer learning, allowing them to better generalize to unseen data shown in Table 1.

The literature has been studied and examined to obtain accurate, less time-consuming outcomes for different classifications of rice varieties and other final products using modern automated techniques. Owing to the very large number of parameters used in various machine learning and deep learning models, the percentage of classification accuracy obtained is not satisfactory for different grains and vegetable images in the literature. Therefore, to overcome the issue of a large number of parameters with the least accuracy of the machine learning and deep learning models used [22], this study focuses on the potential implications and aims to develop a nondestructive model for improving the classification accuracy using images of different rice varieties. The proposed models employ CNN methods to classify 75,000 images from five different rice varieties without requiring the preprocessing of the raw images. The classification accuracies of the AlexNet, ResNet50, and EfficientNet-b1 methods are compared.

*1.1. CNN Architectures in Image Classification.* Various CNN designs have been used extensively in the field of image categorization. Prominent examples of deep learning include AlexNet, ResNet, and EfficientNet, each serving distinct purposes. AlexNet is a foundational architecture in deep CNNs that introduces the idea of deep learning in the

field of vision. Owing to its numerous convolutional layers, it is particularly well suited for the processing of enormous amounts of input. The authors in [23] introduced ResNet, also known as residual learning, which is a noteworthy development in deep learning. The fundamental goal of this system is to create networks that are far deeper, without having to deal with the issue of vanishing gradients, as shown by ResNet50. EfficientNet has recently worked to improve existing best practices by including compound scaling into the network to achieve a balance between the network's depth, breadth, and resolution. According to the authors in [24], this has resulted in improved performance compared to previous models, with fewer parameters and a lower computational burden applied.

These models have garnered a great deal of favor in agricultural applications, particularly in the classification of rice varieties where they have been shown to be highly effective. An example of this is the fact that several architectures of ResNet have shown high-precision categorization for a variety of agricultural items. In addition, these studies demonstrate the need to select the most important models and preprocess them in conjunction with the hyperparameter tuning approach to achieve improved performance [25]. Due to its compact form, it is especially suitable for precision agriculture applications that require high accuracy but have limited computer capabilities, such as those found in mobile and edge devices [25].

*1.2. Challenges and Opportunities in Rice Classification.* Putting a composition of rice types is challenging because the distinctions between these kinds are quite small on a visual level. Currently, these are the classic hand methods that have been used for a considerable number of decades. They are characterized by high work intensity, a propensity for mistakes, and a high degree of subjectivity [26]. The use of automated approaches that have been implemented with the help of CNNs is a solution that is both scalable and dependable. The capacity of CNN models to generalize the identification of comparable varieties of rice, which offer large variations in color, texture, and shape with minute variances, is the source of the intrinsic complexity associated with rice classification. However, this is a highly difficult process that requires sophisticated CNN models and the optimization of hyperparameters, in addition to domain-specific approaches for data augmentation, to provide a correct categorization of the data points.

A growing body of research suggests that the combination of deep learning with image processing methods such as augmentation and normalization may assist in the achievement of superior outcomes in the classification of agricultural goods [16]. In this regard [25], through the use of the data augmentation strategy and transfer learning, Sharma et al. showed that CNN models may be utilized to identify a wide range of rice and wheat varieties. When they are paired with hyperparameter tweaking, CNN models can learn from data that they have not previously seen, which results in a higher classification performance.

TABLE 1: Comparative analysis of literature reviews performed by various researchers.

References	Crops	No. of images in the dataset	Classes	Classifiers	Accuracy levels
[13]	Rice	3810	2	LR	93.02%
[12]	Rice	1700	2	SVM	98.50%
Ahmed, Rahman, and Abid [11]	Rice	200	3	CNN	88.07%
Watanachaturaporn [16]	Rice	1500	30	SRC	89.1%
[16]	Rice	3200	3	SR	86.25%
[15]	Rice	300	6	Feedforward neural network	84.83%
[14]	Rice	800	4	Multiclass SVM	86%
Gogoi et al. [17]	Rice disease	8883	2	3-Stage CNN + transfer learning	94
Wang and Xie [18]	Rice	7000	6	EfficientNet-b4	99.45
[3]	Rice	5500	5	SE_SPnet	99.2
[3]	Rice	4500	5	Res4Net-CBAM	99.35
Jin, Che, and Chen [5]	Rice	5000	3	ResNet50 + data augmentation	99.00
Zhao and Zhang [19]	Rice disease	9000	4	DenseNet-121	98.95
Sharma and Verma [20]	Multigrain	15,000	10	Hybrid EfficientNet + CBAM	98.6
[13]	Rice	3200	3	SR	86.25

*1.3. Recent Advances in Deep Learning.* Over the course of the last several years, deep learning has made significant headway in the area of addressing difficult issues encountered in the real world. These challenges may be found in fields as varied as medical diagnostics, and even in complex systems such as agriculture. The results of several recent studies have shown that they have significant consequences for the identification, prediction, and grading systems of image recognition software [27].

Medical image analysis is a well-established application, particularly for the automated segmentation of blood vessels from retinal fundus images using deep learning models. This was the most notable use. In this study, a deep learning-based strategy is presented to improve the accuracy and effectiveness of the identification of blood vessels associated with a number of retinal illnesses, one of which is diabetic retinopathy (DR) [28]. The diagnosis and staging of DR have improved owing to the application of deep learning methods. These advancements have been achieved regardless of the stage of the illness or severity of the disease in patients. Using the CNN architecture [29], these systems can categorize retinal pictures into various phases of DR, which ultimately leads to an improvement in the identification and treatment of the condition at an earlier stage.

For example, PlaNet models are characterized by a high level of performance in artificial intelligence models of the CNN type. Recent research has shown that it is possible to recognize and classify plant diseases based on images of the leaves of a plant, and this skill is very useful for the agricultural industry. During the process of carrying out this work, the authors in [30] used convolutional layers in order to extract intricate patterns that were associated with disease symptoms on plant leaves. This technique yields positive findings in a short period of time, which enables early identification and has the potential to reduce crop damage [1]. Compared to manual approaches, this approach features a number of benefits, one of which is an improvement in the accuracy of illness diagnosis.

In addition, contributions have been made in the area of the COVID-19 pandemic via the use of deep learning algorithms. Two different models based on deep learning have been presented for major variations in the identification of COVID-19 at an earlier stage. These models depend on computed tomography (CT) scans and accuracy levels achieved by radiologists [31]. One of the methods established on a worldwide scale in response to COVID-19 is the development of rapid diagnostic solutions. These solutions were necessary to satisfy the enormous demand for information at the diagnostic level during the pandemic.

Within the scope of this article, genuine instances of applying deep learning models to the resolution of real-world problems in various domains are shown. Deep learning has been brought forward for the proposed to provide solutions that are more efficient, accurate, and scalable in various fields, including healthcare, agriculture, and public health. CNN-based architectures and automated systems are crucial to this development [32]. As a result, this research contributes to the existing body of knowledge by using hybrid CNN models as the problem model for rice variety categorization. Through the implementation of this strategy, deep learning has become a standard practice in agriculture and farming.

*1.4. Research Gaps.* Despite the advances in deep learning, several gaps remain in the literature.

1. Although individual CNN models have been applied to rice classification, there is a lack of research exploring hybrid models that combine the strengths of different architectures to improve classification accuracy and efficiency. The potential of integrating AlexNet, ResNet, and EfficientNet to form a robust hybrid model remains unexplored.
2. Many existing studies do not sufficiently focus on hyperparameter tuning or advanced augmentation techniques to optimize the performance of CNN

models. Efficient data augmentation methods, such as Mixup and CutMix, have been proven to reduce overfitting but are not widely applied in rice classification tasks.

3. Existing studies primarily rely on controlled datasets with limited variety, which restricts the model's ability to generalize to more complex, real-world conditions. There is a need to explore how models can be optimized for generalization across diverse rice varieties and environmental conditions.
4. Although deep learning models, such as ResNet50 and EfficientNet, have shown high accuracy, they often require significant computational resources. There is a gap in research addressing the trade-offs between computational efficiency and model accuracy, particularly in hybrid architectures for agricultural applications.

**1.5. Problem Definition.** The categorization of rice is a crucial issue in the agricultural sector, as it guarantees quality standards for various rice varieties, thereby facilitating supply chain management and market segmentation. Previously, rice classification was labor-intensive, less precise, and reliant on manual techniques. Advancements in deep learning have enabled the realization of high-performance automated categorization. Nevertheless, current CNNs are not directly suitable for rice classification because of their resemblance in appearance across different rice species. Hybrid models that use the concentric strengths of distinct CNN architectures to enhance the efficiency and generalization of rice classification, among other things, are limited.

**1.6. Motivation.** Motivated by this, our study seeks to provide a novel method to address the limitations of existing classification algorithms that use deep learning, notably hybrid CNNs. Although individual CNN models, such as AlexNet, ResNet50, and EfficientNet-b1, may exhibit superior performance in image classification tasks, combining these models to build an architecture may provide enhanced accuracy, efficiency, and generalization. Therefore, there is a pressing need to create a machine learning-driven system that can automate, both rapidly and scalably, the accurate classification of rice varieties with little human error and consistent quality control on agricultural prospects.

**1.7. Research Question.** Improved accuracy, generalization, and efficiency of rice variety classification using hybridization of several CNN architectures (AlexNet, ResNet50, and EfficientNet-b1) compared to individual models.

#### 1.8. Objectives

1. A hybrid CNN model was created by integrating AlexNet, ResNet50, and EfficientNet-b1 to exploit the advantages of each architecture.
2. The efficiency of the hybrid model is enhanced by optimizing hyperparameters, augmenting data, and using transfer learning techniques.

3. Analysis of the accuracy, precision, recall, and generalization of the hybrid model compared to the single CNN models (AlexNet + ResNet50 + EfficientNet-b1).
4. Evaluation of the computational efficiency and practical applicability of this hybrid model in actual agricultural environments.

**1.9. Contribution.** The primary contribution of this study is the development of a unique hybrid CNN model that combines three models, namely, AlexNet + ResNet50 and EfficientNet-b1, to achieve improved performance in the identification of rice varieties. The integration of the two models in this manner enables enhanced feature extraction using the most advantageous aspects of each design.

1. AlexNet is a straightforward network capable of generating and preparing fundamental characteristics.
2. ResNet50 was used with residual learning for deep feature extraction.
3. EfficientNet-b1 was used for precise specifications for improved scalability and feature processing with reduced parameter use.

Compared to CNN models, this hybrid model design yields higher accuracy and generalization. Furthermore, it demonstrates the efficacy of hyperparameter tweaking, innovative data augmentation techniques (such as Mixup and CutMix), and transfer learning to mitigate overfitting by optimizing the model.

**1.10. Novelty.** This paper introduces a unique hybrid CNN-based architecture for rice categorization and machine learning aimed at addressing an unexplored requirement in the literature. Although previously separate models have been used for picture classification, the combination of AlexNet, ResNet50, and EfficientNet-b1 into a single hybrid model is a novel approach to address all stages of the rice classification issue. This work demonstrates modifications to improve the hybrid model using advanced augmentation techniques and transfer learning, resulting in a comprehensive and practical system for automated rice classification in agricultural environments.

This paper is structured as follows. Section 2 elucidates the proposed methodology. Section 3 presents the impact of hybridization, strategies for overcoming drawbacks, experiments and analysis, and novel insights and improvements. Section 4, presents the result and discussion, and the conclusion of the study.

## 2. Research Methodology

This study on rice image classification meticulously designed the dataset collection and preprocessing phase to gather a comprehensive and widely recognized dataset for the successful development of deep learning models [33]. To create a comprehensive dataset including five distinct varieties of rice commonly found in Turkey, namely, Arborio,

Basmati, Ipsala, Jasmine, and Karacadag, a compilation of rice grain photographs was undertaken. The collection included 75,000 photos, with 15,000 allocated to each type. To accurately illustrate the characteristics of each type, photographs were methodically captured using sophisticated imaging technologies and techniques [13, 34, 35]. Prior to being included in the training set for the CNN model, these photos underwent third-party preprocessing to ensure consistency in the dataset and enhance quality. During the preprocessing phase, unnecessary noise, artifacts, and other irregularities that could compromise the correctness of the model were eliminated.

- a. Features of the dataset
- b. Total number of images: 75,000
- c. Classes (varieties): 5 distinct rice varieties (e.g., Arborio, Basmati, Ipsala, Jasmine, and Karacadag)
- d. Image resolution:  $224 \times 224$  pixels (resized)
- e. Color mode: RGB (3 channels)
- f. File format: JPEG

Image augmentation: Applied during training (random flipping, rotation, Mixup, and CutMix).

Class distribution: Spanning evenly across all types with 15,000 images per class, as shown in Table 2.

Empirical data augmentation model: To enhance the diversity of the dataset and maximize the amount of data, we implemented a data augmentation model. The use of random transformations, such as rotations, translations, mirror flips, or changes in brightness and contrast levels on the original pictures facilitates the generation of additional training examples for learning [36]. To enhance the generalization of our models to other rotations and locations using incoming data, these actions are essential for radiation data augmentation. The dataset was partitioned into training, validation, and testing sets. The training set included the majority of data used to train the CNN models [37]. The introduction of the new term “point estimate” and the validation set significantly strengthened the optimization of model hyperparameters and allowed the performance of existing models to be monitored [38]. The remaining portion was designated for a thorough assessment of the performance of the test set model after training.

To assess the effectiveness of the trained models, a confusion matrix was constructed. A confusion matrix decomposes classification judgments into many statistical metrics, including sensitivity, specificity, prediction,  $F1$ , precision, and false positive rate (FPR). The metrics presented in the following demonstrate the accuracy, precision, recall, and  $F1$  measures of the models used for classifying rice. A comparative analysis of the accuracy and depth of three CNN models (AlexNet, ResNet50, and EfficientNet-b1) in the categorization of rice varieties is performed. The accuracy scores of each model were as follows: 96.00% for AlexNet, 99.00% for ResNet50, and 99.87% for EfficientNet-b1. These numbers provide a comprehensive analysis of the performance of the models for each dataset.

TABLE 2: Dataset size and split.

Feature	Training set	Testing set	Total
Number of images	60,000	15,000	75,000
Images per class	12,000	3000	15,000
Image resolution	$224 \times 224$ px	$224 \times 224$ px	—
Color channels	3 (RGB)	3 (RGB)	—
Classes (varieties)	5	5	5
Image augmentation	Yes (train only)	No	—
Total size (in MB/GB)	~2.5 GB	~0.7 GB	~3.2 GB

The purpose of this study was to evaluate the performance of three CNN models, AlexNet, ResNet50, and EfficientNet-b1, in the classification of five distinct varieties of rice. An analysis of rice image classification was performed based on this evaluation. The following is a list of the primary properties shared by various models, notably, AlexNet, which is one of the CNN models that is used most frequently; essentially, a CNN is made up of a number of convolutional layers that are accompanied by a few fully connected (FC) layers that have been selected [39]. Comparative analyses of AlexNet, ResNet50, and EfficientNet-b1 are provided as examples of several models that are considered for rice-type categorization. A common CNN architecture known as ResNet50 is capable of excelling in the majority of image classification jobs. The implementation of ResNets, which have relatively basic optimization in their training, has become popular [40]. Consequently, they are more efficient and exhibit superior performance. ResNet50 was one of the models that was used in this investigation to examine whether or not it was capable of properly identifying a number of different varieties of rice. EfficientNet-b1 is a CNN architecture that achieves remarkable performance with a limited number of parameters and processing resources [41]. Compound scaling is utilized to achieve equilibrium in the depth, breadth, and resolution of the model. In addition, it was used in the experiment to evaluate its classification performance in comparison with ResNet50 and AlexNet. AlexNet, ResNet50, and EfficientNet-b1 are the three CNN architectures included in the suggested model, which is referred to as the hybrid model architecture [42]. This model’s architecture has been constructed in such a way that it is capable of extracting both high- and low-level information, which enables it to perform more effectively when it comes to the categorization of rice varieties.

## 2.1. Model’s Architecture

### 2.1.1. Input Layer. Input: $224 \times 224$ pixels RGB images.

The input images were passed to three parallel CNN architectures, AlexNet, ResNet50, and EfficientNet-b1.

### 2.1.2. AlexNet. Five convolutional layers, followed by rectified linear unit (ReLU) activation and max pooling.

The final convolutional layer outputs a feature map, which is flattened and passed to the FC layers, with dropout and ReLU activation for regularization.

**2.1.3. ResNet50.** Fifty-layer deep residual network with skip connections (residual blocks) mitigated the vanishing gradient problem. Each residual block contained multiple convolutional layers, followed by batch normalization and ReLU activation. The network outputs a feature map that is passed through global average pooling before being flattened.

**2.1.4. EfficientNet-b1.** Compound scaling was used to balance the depth, width, and resolution of the models.

It consists of inverted residual blocks with depthwise separable convolutions.

Squeeze-and-excitation (SE) blocks were used to recalibrate channelwise feature responses. The final layer outputs a feature map that is passed through global average pooling.

**2.1.5. Feature Concatenation Layer.** Feature maps from AlexNet, ResNet50, and EfficientNet-b1 were concatenated to form a comprehensive feature vector.

**2.1.6. FC Layers.** The concatenated features were passed to a FC layer using ReLU activation.

A dropout layer (dropout rate = 0.5) was used for regularization to prevent overfitting.

**2.1.7. Output Layer.** The final output layer was a softmax layer that predicted the probabilities for each of the five rice varieties (multiclass classification) as shown in Figure 1.

These deep learning models were selected based on their reputation in the field of image classification and their proven capabilities in achieving a high accuracy rate, as shown in Figure 1. The block diagram for employing CNN models, such as ResNet50, AlexNet, and EfficientNet-b1, aimed at evaluating and comparing the performance of these models in classifying the five rice varieties [43]. The models were trained for 20 epochs using the augmented dataset. The models were fine-tuned with a very small learning rate and tested on the final testing set. Training regularization techniques such as dropout and L2-Norm on weights were used to prevent overfitting. The models were evaluated using recall, precision, and *F1* score [44]. We used three different architectures to establish the best architecture for the classification of images that were very similar in nature. The older and lighter architecture of AlexNet performed well. The residual architecture-based ResNet50 network performed well but had many parameters. Finally, we proceeded with the EfficientNet-b1 variant [45]. This feature extraction method helps in maintaining all the important aspects of the images that perform well, even if the images are found to be similar, as shown in Algorithm 1.

The results of the study indicated that all three models achieved high rates of accuracy, with EfficientNet-b1 outperforming the other two models, followed by AlexNet and ResNet50 [47]. The results provided empirical evidence that the residual models, particularly EfficientNet-b1, were easier

to optimize and provided higher accuracy in classifying rice varieties.

#### 2.1.8. Input Image

- The model uses high-resolution microbial images as input, where each pixel needs to be classified as altered or nonaltered.
- Input resolution: 512 × 512 pixels or adjustable according to microbial image size.

#### 2.1.9. Backbone: Transformer Encoder

- SegFormer replaces the conventional CNN backbone with a pure transformer encoder, which enables the model to capture the long-range dependencies in the image.
- The backbone consists of multiple multihead self-attention layers that allow the model to learn spatial relationships across the entire image, which is crucial for segmenting microbial structures with complex alterations.

#### 2.1.10. Multiscale Feature Extraction

- SegFormer's hierarchical design uses multiple transformer stages to extract the feature maps at different scales.
- Stage 1: High-resolution features are extracted at the finest scale.
- Stage 2: Features are extracted at progressively coarser scales.
- This multiscale approach is particularly useful for capturing both detailed microbial alterations (e.g., cell structure deformations) and a broader spatial context.

#### 2.1.11. Segmentation Head

- After extracting multiscale features from the transformer encoder, SegFormer uses a lightweight MLP-based segmentation head that fuses information from all scales [48].
- This generates the final semantic segmentation map, where each pixel is classified into one of the defined classes (e.g., altered or nonaltered microbial regions).

#### 2.1.12. Output

- The final output is a pixel-level classification map that delineates the areas of microbial alteration and accurately identifies regions where structural changes or abnormalities have occurred.
- Each pixel in the microbial image is labeled with a class as follows:
  - Class 1: Normal microbial structure.
  - Class 2: Altered microbial regions.

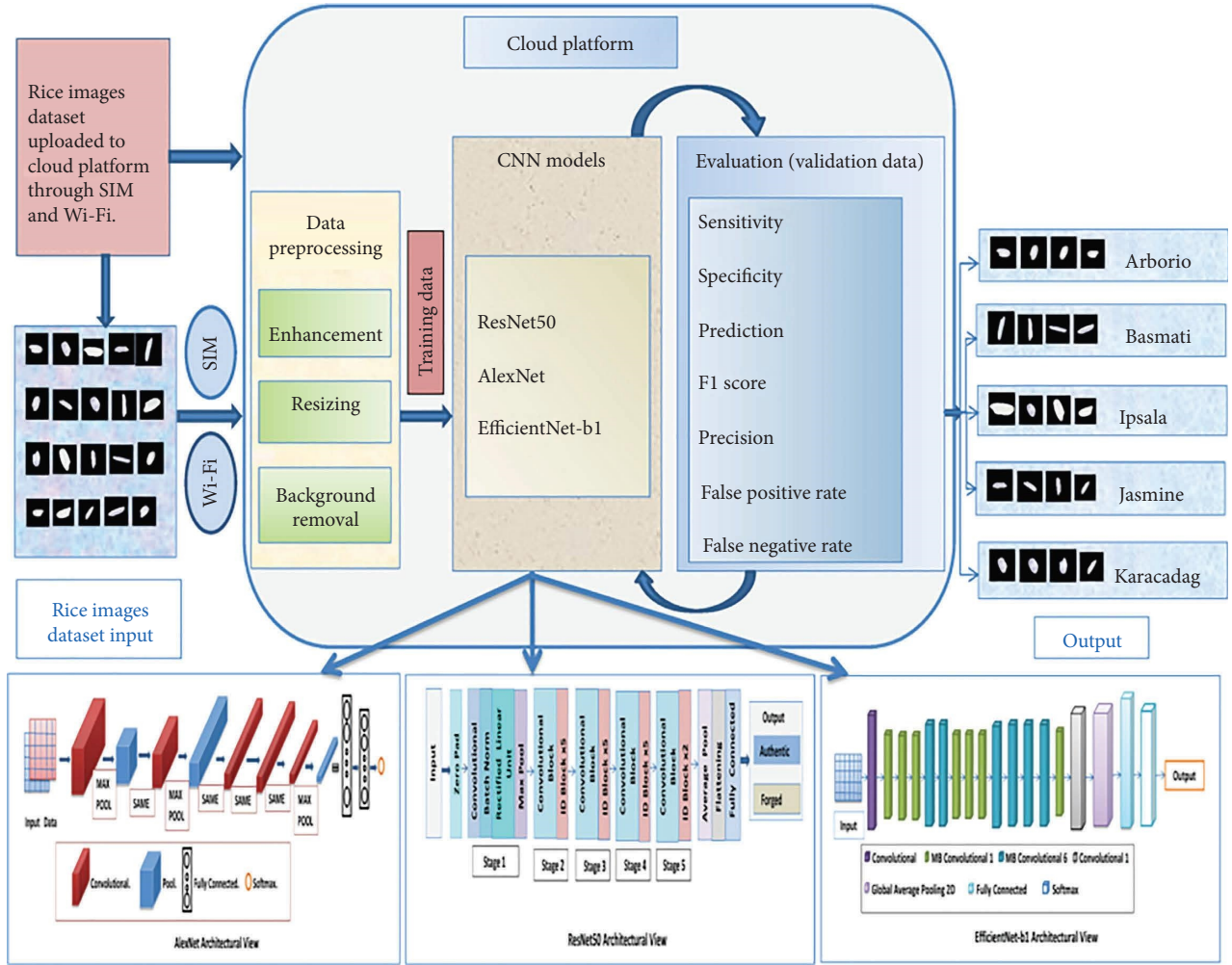


FIGURE 1: Block diagram of rice image classification using CNN architecture.

This study emphasizes the crucial steps of training and optimizing CNN models (AlexNet, ResNet50, and EfficientNet-b1) in rice image classification to obtain precise classification results with high accuracy rates. To train the algorithm, we curated a dataset consisting of 75,000 photos. This dataset was divided into 15,000 images for each sample group, namely, cultivar species, representing five distinct rice-growing variations. This included generating photos in the appropriate folder or folder hierarchy, with a well-balanced allocation of samples from each category. Augmentation techniques were used to enhance the variety and effective size of the training datasets [49]. Techniques such as random rotations, flips, zooms, and shifts have facilitated the generation of a greater number of training examples. Consequently, this resulted in enhanced generalization capacity and prevented the models from being overfitted.

The CNN models, AlexNet, ResNet50, and EfficientNet-b1, were initialized using their own architectures and pretrained weights from ImageNet. An effective strategy for using pretrained models is to first train a model on a fake task using extensive picture datasets. This allows the model to acquire general image characteristics that can then be readily applied and refined for new

characterization tasks. Transfer learning is used to leverage the information already acquired by pretrained models for the purpose of rice categorization [50]. A portion of this code will immobilize the first layers to preserve the acquired character after regenerating them using the rice dataset for a few further levels [51]. This approach enables efficient execution of training even with a limited amount of data. Model optimization was achieved by adjusting the hyperparameters. The factors considered were the learning rate, batch size, optimizer selection, regularization methodology, and number of training epochs. Random search approaches were used to tweak the hyperparameter grid optimization.

The models were trained using the prepared dataset and the optimized hyperparameters. The training process involved feeding batches of training images into the models, computing the cross-entropy loss, and updating the model weights through backpropagation and gradient descent optimization [52]. The dataset we used was already pre-processed with rice grains, as highlighted in the image and the background as black. The total number of data points was 75,000, with 15,000 data points belonging to each class. The dataset was divided into three parts as follows:

```

Input: Dataset of images and hyperparameters (learning_rate, batch_size, and epochs)
Output: Trained hybrid CNN model
Begin
Step 1: Data Preparation
  Preprocess (input dataset) -> resized, normalized, and augmented data
  Split (resized, normalized, and augmented data) -> training set, validation set, and testing set
Step 2: Model Construction
  Initialize AlexNet, ResNet50, and EfficientNet-b1 with pretrained weights
  Remove output layers (AlexNet, ResNet50, and EfficientNet-b1) -> feature extractors
  Concatenate (feature extractors) -> unified feature vector
  Add FullyConnectedLayers to the unified feature vector
  For each layer in FullyConnectedLayers do
    Add DropoutLayer
  End For
  Append OutputLayer (softmax activation and number of classes) to FullyConnectedLayers -> HybridModel
Step 3: Model Training
  Compile (HybridModel, LossFunction = "categorical_crossentropy," Optimizer = "Adam," Metrics = ["accuracy"])
  Train (HybridModel, training set, validation set, and hyperparameters)
Step 4: Model Evaluation
  Evaluate (HybridModel and testing set) -> classification accuracy, and other metrics
Step 5: Fine-tuning
  If performance is not satisfactory, then
    Adjust hyperparameters (HybridModel)
    Retrain (HybridModel)
  End If
Step 6: Model Deployment
  Deploy (HybridModel) -> For new image classification tasks
End

```

ALGORITHM 1: Train hybrid CNN model.

- i. Training: 62,000 with 12,400 belonging to each class
- ii. Validation: 5000 with 1000 belonging to each class
- iii. Testing: 8000, with 1600 belonging to each class

The performance evaluation of the three CNN models was conducted by training them for 20 epochs with a learning rate of  $1e-5$ , utilizing their respective number of parameters. As the parameters for the employed models are different from each other, in this context, the first model, AlexNet, is employed which contains 62.3 million parameters and is also known as a simple convolutional neural-based backbone model. Feature extraction of the task is performed using this model after each layer. The obtained features were then flattened to further classify them. This is one way through which the training and testing accuracies are obtained. , ResNet50, which has 23 million parameters [54], is used for the same node. This model has a residual-based backbone that considers the residues and ensures that the model's layers do not hinder each other's performance. That is, it does not have any features that reduce the performance of the previous layer, even if it has a very deep network. Finally, the EfficientNet-b1 variant, which has 7.8 million parameters, was used. This model considers vertical, horizontal, and other features, as well as the final prediction [25].

After training all the models with the same hyperparameters and number of epochs on the same setting of training data, a comparative study of all three CNN variants was performed. Despite the many differences in the number

of parameters, the premise of this study is to establish that architecture, feature extraction, and generalization play a more important role than the number of parameters. Even with 7.8 million parameters, EfficientNet-b1 performs well owing to its better feature extraction technique [1]. From the results, it can be inferred that even with a smaller number of parameters, it is possible to match the performed models with better feature extraction techniques and better and lighter architectures. From the results, it is concluded that EfficientNet-b1 can be considered as the model in which the best trade-off between complexity and performance is found. Even with a small number of parameters, it outperformed the AlexNet model and matched the performance of ResNet50.

One more point to be highlighted in the paper is that EfficientNet-b1 is considered the best model because the whole dataset images used are of rice grains with different qualities; hence, it is quite difficult to classify the types of rice through manual processes. With the preprocessed images, the need for a proper CNN model has been identified by considering the proper dimensions and angles of every rice image dataset to properly classify them, as the images are very close to each other. The model was performed in terms of extraction of the features in horizontal ways, and convolutionally, a final feature extraction process was performed, and a final classification process was also performed. A model with a smaller number of parameters performs well in the feature extraction processes.

The convolutional layer is represented as

$$(I * F)(i, j) = \sum_m \sum_n I(i - m, j - n) \cdot F(m, n), \quad (1)$$

where  $i, j$  are the spatial dimensions of the output feature map and  $m, n$  are the dimensions of the filter. ReLU is defined as

$$\text{ReLU}(x) = \max(0, x). \quad (2)$$

Then, max pooling, a common pooling operation, is defined as

$$\text{MaxPooling}(I)(i, j) = \max_{a \in W_{i,j}} a, \quad (3)$$

where  $W_{i,j}$  is the window of inputs covered by the pooling operation at position  $(i, j)$ .

The hybrid model combines the features from AlexNet, ResNet50, and EfficientNet-b1. Let  $F_{\text{AlexNet}}$ ,  $F_{\text{ResNet50}}$ , and  $F_{\text{EfficientNet-b1}}$  represent the feature extraction functions of the respective models. The combined feature vector  $F_{\text{combined}}$  can be represented as

$$F_{\text{combined}} = \text{Concatenate}(F_{\text{AlexNet}}(I), F_{\text{ResNet50}}(I), F_{\text{EfficientNet-b1}}(I)), \quad (4)$$

where  $I$  is the input image, and Concatenate is the operation that merges the feature vectors from each model. After the feature combination, FC layers were used for classification. The output of the FC layer is

$$\text{FC}(x) = Wx + b, \quad (5)$$

where  $x$  is the input vector to the FC layer,  $W$  is the weight matrix, and  $b$  is the bias vector. The final output layer, typically a softmax layer for multiclass classification, computes the probability distribution over classes, as follows:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_k e^{z_k}}, \quad (6)$$

where  $z_i$  is the input to the softmax function for class  $i$  and  $k$  is the total number of classes. The model is trained by optimizing a loss function, commonly the cross-entropy loss for classification, which is defined as

$$\text{cross-entropy}(y, \hat{y}) = - \sum_i y_i \log(\hat{y}_i), \quad (7)$$

where  $y$  is the true label in one-hot encoding and  $\hat{y}$  is the predicted probability distribution from the softmax layer.

When the complexity of the three models is considered, the number of parameters of the models strictly decreases. The number of parameters of ResNet50 was three times lower than that of AlexNet, and in the case of EfficientNet-b1, it was approximately two and a half times less than that of ResNet50, and five times smaller than that of AlexNet. Among them, the performance of the models was comparatively closer, but the complexity was higher for the other two models. Although there is a vast difference between the number of parameters, the performance obtained is considered to be the best in the case of the EfficientNet-b1 model owing to its low complexity. In addition, this model can be run easily on CPU rather than using HPC, such as TPUs and GPUs, to classify them. Therefore, this is the performance of the models used.

### 3. The Impact of Hybridization

The present study suggests a CNN model that is a hybrid of AlexNet, ResNet50, and EfficientNet-b1, with the intention

of merging the strengths of each network into a single network. The fundamental idea behind this is to consolidate the characteristics of several designs to enhance the accuracy of the classification as well as the resilience of the structure. The structure, data flow, and optimization strategies used in the construction of the hybrid model are discussed in the next section. The pictures taken by the input layer were 224 pixels by 224 pixels, and they had three colors channels that represented RGB. It is essential to perform data preparation to improve the ability of the model to generalize. The normalization of pixel values on a common scale and additional enhanced data, which might be obtained by rotation, flipping, zooming, or cropping, are included in this. Overfitting may be avoided by augmenting the data, which is particularly important when considering the small variance that exists between the types in the dataset.

Each input image  $X$  with dimensions  $H \times W \times C$  (height, width, and channels) is normalized to pixel values in the range  $[0, 1]$ . Let the normalized input image be

$$X' = \frac{X}{255}. \quad (8)$$

We also apply data augmentation  $A(X)$  to increase generalization as

$$X_{\text{aug}} = A(X'). \quad (9)$$

This augmented image  $X_{\text{aug}}$  is the input to all the three CNN models.

AlexNet, ResNet50, and EfficientNet-b1 are the three parallel CNN architectures that are fed pictures created from the beginning. The inputs are processed individually by each model, which then extracts distinctive characteristics that are relevant to categorization.

The AlexNet algorithm was chosen for this project because it is a straightforward and effective method for extracting fundamental characteristics, such as edges and textures. The spatial dimensions are downsampled using max pooling, which is interspersed across the five convolutional layers that constitute this structure. To apply filters to identify low-level patterns within the photographs, such as color gradients and edges, which are crucial

characteristics for distinguishing between the many types of rice, this form of architecture was used.

The convolutional operation in AlexNet can be expressed as

$$F_{\text{Alex}} = \sigma(W_{\text{Alex}} * X_{\text{aug}} + b_{\text{Alex}}), \quad (10)$$

where  $*$  denotes convolution,  $W_{\text{Alex}}$  is the filter weight matrix,  $b_{\text{Alex}}$  is the bias, and  $\sigma$  is the ReLU activation function:

$$\sigma(z) = \max(0, z). \quad (11)$$

It is now time for the feature maps to undergo the process of flattening their vectors and going through completely linked layers with ReLU activation. To avoid overfitting, a dropout layer is used, which involves random deactivation of neurons throughout the training process. It is expected that the output of AlexNet will be a feature vector that represents fundamental picture features.

The issue of vanishing gradients, which are often associated with deep neural networks, was addressed by ResNet50 using residual connections. Due to these connections, the network can skip layers, which makes the gradient flow during the process of backpropagation easier to manage. In turn, this makes it possible to train extremely deep designs. Twenty-five layers comprised the ResNet50 architecture, with each layer comprising a significant number of convolutional layers, which were then followed by batch normalization and ReLU activation in residual blocks. The use of residual connections by ResNet results in an improved gradient flow. A residual block is responsible for the following:

$$F_{\text{Res}} = \sigma(W_1 * X_{\text{aug}} + b_1) + X_{\text{aug}}. \quad (12)$$

During the backpropagation process, the skip connection guarantees that the gradient flow is adequately maintained, which prevents vanishing gradients from occurring.

However, because it is able to capture high-level and sophisticated traits such as form and detailed patterns, it is a great option for the categorization of rice varieties that are visually similar to one another. In addition to performing global average pooling, ResNet50 constructs a feature vector that comprises deep, learned patterns from the pictures.

Compound scaling is used in EfficientNet-b1, which aims to maintain the equilibrium between depth, width, and resolution to achieve the highest possible performance while simultaneously reducing the amount of processing resources required. This design uses depthwise separable convolutions to significantly reduce the number of parameters while maintaining the same level of performance. These SE blocks were responsible for the recalibration of channel-wise feature responses. This allows the network to concentrate more on the portions of the picture that are most significant. In EfficientNet-b1, SE blocks and depthwise separable convolutions are used in the learning process. Iterative Convolution with Depth.

$$F_{\text{Eff}} = \sigma(W_d * X_{\text{aug}} + b_d), \quad (13)$$

where  $W_d$  is the depthwise filter. The SE block scales the feature channels as follows:

$$F_{\text{Eff,SE}} = F_{\text{Eff}} \cdot \sigma(W_{\text{SE}} F_{\text{Eff}} + b_{\text{SE}}). \quad (14)$$

As a result, the model's emphasis on essential characteristics is further highlighted.

Therefore, it is particularly suitable for extracting fine-grained characteristics from complicated datasets, such as distinguishing between rice types that have only slight visual changes. To build a compact feature vector, the feature map of EfficientNet-b1 is transferred to the global average pooling, which then constructs the feature vector.

Concatenation is then performed on the vectors obtained by AlexNet, ResNet50, and EfficientNet-b1 in order to create a single universal vector. Therefore, the significance of a feature concatenation layer is at the core of merging all three designs for general strengths. These architectures include AlexNet, which provides fundamental texture information, ResNet50, which provides complicated patterns and forms, and EfficientNet-b1, which provides fine-grained optimized features.

In order to create a unified feature vector, the output feature vectors  $F_{\text{concat}}$  from the three different models are concatenated together as

$$F_{\text{concat}} = [F_{\text{Alex}}, F_{\text{Res}}, F_{\text{Eff,SE}}]. \quad (15)$$

This vector combines both low-level and high-level characteristics to obtain a more accurate categorization.

The concatenated vector encodes both high-level and low-level characteristics; as a result, it enhances the capability of the model to accurately identify the five different types of rice included in the dataset. Due to this integration, all of the designs are able to mix together and play the strengths of the three, which ultimately results in a performance that is superior to that shown by the separate models. Subsequently, the concatenated feature vector is transmitted across layers that are completely linked. The ReLU activations included inside each of these are responsible for modifying the feature combinations to extract meaningful representations. The use of a dropout layer with a dropout rate of 0.5 helps the model generalize to data that it has not before seen and avoids overfitting. Subsequently, it passes the concatenated feature vector via a layer that is completely linked.

$$z = W_{\text{fc}} F_{\text{concat}} + b_{\text{fc}}, \quad (16)$$

where  $W_{\text{fc}}$  and  $b_{\text{fc}}$  are the weights and biases of the FC layer. The final output was computed using the softmax function as

$$\hat{y}_i = \left( \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)} \right), \quad (17)$$

where  $\hat{y}_i$  is the predicted probability for class  $i$  and  $n$  is the total number of classes.

The softmax method was used in the output layer to provide a probability for each of the five rice classifications. In multiclass classification, the softmax function guarantees that the probabilities for each rice-type class amount to one.

To tune the hybrid CNN model, the Nadam optimizer, an adaptive learning rate algorithm that integrates RMSprop and momentum-based gradient descent, is used. A learning rate of  $1e-5$  is used for this purpose in conjunction with categorical cross-entropy loss. This was implemented because it represents the most suitable solution for multiclass classification problems. Subsequently, the model underwent training for a total of 20 epochs with a batch size of thirty-two, aiming to attain a balance between convergence speed and model stability. Moreover, L2 weight regularization may mitigate overfitting by imposing penalties on excessively large weights. In addition, to enhance the generalization capability of the model, data augmentation techniques such as CutMix and Mixup are used. This procedure was performed to achieve the intended outcomes.

The hybrid model's performance has been assessed using many metrics, including accuracy, precision, recall, and F1-core, among others. It was designed to comprehend the errors produced by the model along with an evaluation of its strengths and limits. The confusion matrices were then generated. In this matrix, a more precise categorization outcome relative to individual models may be observed for minor variations in rice types.

A categorical cross-entropy loss function was used in the model training as follows:

$$L = - \sum_{i=1}^n y_i \log(\hat{y}_i), \quad (18)$$

where  $y_i$  is the true label (one-hot encoded) and  $\hat{y}_i$  is the predicted probability for class  $i$ .

The Nadam optimizer updates the model's parameters  $\theta$  using the gradients of the loss function as

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}, \quad (19)$$

where  $\eta$  is the learning rate,  $m_t$  and  $v_t$  are the first- and second-moment estimates, and  $\epsilon$  is a small constant to prevent division by zero.

To prevent overfitting, L2-regularization was applied to the following weights:

$$L_{\text{total}} = L + \lambda \sum_k \|W_k\|_2^2, \quad (20)$$

where  $\lambda$  is the regularization coefficient and  $W_k$  are the model weights.

In addition, dropout is applied with a probability  $p$  to deactivate neurons during training as

$$F_{\text{drop}}(x) = \begin{cases} 0, & \text{with probability } p, \\ \frac{x}{1-p}, & \text{otherwise.} \end{cases} \quad (21)$$

According to the results, the hybrid model outperformed all separate CNN architectures that were tested. It demonstrated much greater accuracy while generally providing a superior model with fewer FPs and FNs, which resulted in an improvement in the overall performance of the job. The

merging of complementary strengths from AlexNet, ResNet50, and EfficientNet-b1 allowed for the maintenance of state-of-the-art performance in rice categorization tasks.

1. The combination of the simplicity of AlexNet, residual learning from ResNet50, and the compound scaling approach from EfficientNet-b1 resulted in superior accuracy compared to any individual model used in isolation. This hybrid methodology allowed us to capture both the aerial landscape and individual trees more effectively simultaneously.
2. Increased generalization: Since machine learning contributes to half of the issue solution, it exhibits superior generalization to unfamiliar data. The implementation of a hybrid model enables the effective use of many architectures that can acquire knowledge from a wider range of characteristics, thereby enhancing the generalization ability of diverse datasets.
3. The models exhibited enhanced efficiency in feature extraction by collectively addressing distinct tasks and assigning varying degrees of importance to each other. For instance, ResNet50 was designed to prioritize the screening of deeper features, whereas EfficientNet-b1 was specifically designed to handle well scales ranging from 150 to 600. This results in improved classification outcomes.
4. One trade-off between hybrid models is their high computational costs. To address this, we used EfficientNet-b1 as a lightweight base architecture to reduce the computational footprint. This compromise between the computational efficiency and model performance is feasible.

### 3.1. Drawbacks: Our Strategies for Overcoming

1. The complexity of hybrid models is much higher; therefore, we made a deliberate effort to carefully choose layers and develop architectures that facilitate our integration. Each architectural design is used only for the extraction of certain head features to maintain the overall manageability of the hybrid model.
2. Our choice of the EfficientNet-b1 model was its notably reduced computational cost, resulting from its lower overall parameter count. Ultimately, we used transfer learning to begin the model with pretrained weights, thereby significantly minimizing the training time and resources required.
3. Global overfitting: We implemented dropout and L2-regularization in Mixup and CutMix algorithms. The use of these procedures reduces the likelihood of overfitting, thereby enhancing the generalization capabilities of the newly integrated model with fresh data.
4. Extended training duration: The use of transfer learning resulted in a reduction in the training time for the hybrid model. Implementing pretrained weights in our scenario provided a readily available

alternative for fine-tuning, resulting in significant time savings while maintaining a performance comparable to that of a conventional transfer learning model.

5. Hybrid models face interpretability challenges because of the incorporation of disparate architectures. To achieve this objective, we used Grad-CAM and layerwise relevance propagation (LRP) techniques to provide visual explanations of the underlying mechanisms of the models. These methods not only enhance comprehensibility and interpretability but also contribute to the clarification of hybrid models.

The present work employs many assessment criteria to assess the performance of CNN models, namely, AlexNet, ResNet50, and EfficientNet-b1, in the categorization of rice varieties. The evaluations were conducted using the following generated and examined measures.

The confusion matrix provided a comprehensive breakdown of the classification results obtained by the model. This figure displays the statistical measures of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). This approach allows an unfettered analysis of the overall performance of the model.

Sensitivity, often referred to as the recall or TP rate, quantifies the ability of a model to correctly detect all positive events. Positive classification accuracy is the proportion of correctly classified positive samples out of all the actual positive samples.

$$\text{Sensitivity (recall)} = \frac{TP}{(TP + FN)} \quad (22)$$

Specificity refers to the capacity of a model to accurately detect only truly negative examples. Negative prediction accuracy was defined as the proportion of TN samples that were accurately predicted. The calculation of specificity is performed as

$$\text{specificity} = \frac{TN}{(TN + FP)} \quad (23)$$

The *F1* score is expressed as the harmonic mean of accuracy and recall. This is a compromise, as high accuracy implies that the model produces few anticipated positives. However, it is also necessary to guarantee that the recall is effective and capable of identifying all positive occurrences that the model is expected to discover. The *F1* score was defined as follows:

$$F1 \text{ score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (24)$$

Precision refers to the proportion of accurately categorized positive cases out of all observed instances anticipated as positive. This measure assesses the ability of the model to prevent FPs. Precision is computed as

$$\text{precision} = \frac{TP}{(TP + FP)} \quad (25)$$

The FPR is the ratio of TN events that were mistakenly identified as positive events. This provided an indicator of the FPR of the model. The FPR was calculated as follows:

$$\text{False Positive Rate} = \frac{FP}{(FP + TN)} \quad (26)$$

The false negative rate (FNR) is the proportion of real positive detections mistakenly classified as negative. This refers to the rate of FNs produced by a model. This graph illustrates the computation of the FNR, which is the ratio of TPs (true churners with proper identification) to FNs as follows:

$$\text{False Negative Rate} = \frac{FN}{(FN + TP)} \quad (27)$$

The findings in Tables 3, 4, and 5 demonstrate that the residual models (AlexNet, ResNet50, and EfficientNet-b1) achieved high performance (96.00% for AlexNet, 99.00% for ResNet50%, and 99.87% for EfficientNet-b1) in accurately identifying rice species on the included datasets. This indicates that these models were equally effective for this task.

**3.2. Experiments and Result Analysis.** The dataset for rice image classification research was created from five specific rice types that are extensively cultivated in Turkey: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Hence, the dataset exhibited a well-balanced distribution of 75,000 photos and 15,000 images per rice variety. This ensured an optimal selection for the effective learning and classification of the five distinct rice varieties. The photographs were captured to provide a comprehensive visual representation of each type of rice, and they may vary in form, size, or overall color, distinguishing one type from another without any specific criteria.

These parameters summarize the model configurations used in the analysis of CNN models (AlexNet, ResNet50, and EfficientNet-b1), as shown in Table 6. To enhance the system performance, automated hyperparameter optimization techniques, such as Bayesian optimization or genetic algorithms, can replace grid search for faster convergence. Advanced data augmentation techniques, such as auto augment and elastic transformations, can further improve model robustness. Adopting an ensemble learning strategy can enhance the classification accuracy by combining predictions from different models. Real-time deployment can be optimized via model pruning and quantization to reduce computational costs. In addition, expanding the dataset using synthetic data or exploring self-supervised learning improves the generalization. Finally, XAI techniques (e.g., SHAP and LIME) can enhance interpretability for practical agricultural applications as shown in Table 7.

In addition to the accuracy rates, this study compared the depth of each CNN model. The depth of a model refers to the number of layers it contains, which affects its capacity to learn complex features and patterns from data. A deeper model can capture more intricate relationships but may also be more computationally expensive and prone to overfitting.

TABLE 3: Classification reports of AlexNet (testing).

Class	Precision	Recall	F1 score	Support
0	1	0.99	0.99	1617
1	0.99	1	1	1586
2	1	1	1	1599
3	1	0.99	0.99	1612
4	0.99	1	0.99	1586
Accuracy			0.99	8000
Macroaverage	0.99	0.99	0.99	8000
Weighted average	0.99	0.99	0.99	8000

TABLE 4: Classification report of ResNet50 (testing).

Class	Precision	Recall	F1 score	Support
0	1	1	1	1599
1	1	1	1	1603
2	1	1	1	1600
3	1	1	1	1597
4	1	1	1	1601
Accuracy			1	8000
Macroaverage	1	1	1	8000
Weighted average	1	1	1	8000

TABLE 5: Classification reports of EfficientNet-b1 (testing).

Class	Precision	Recall	F1 score	Support
0	1	0.99	1	1604
1	1	1	1	1599
2	1	1	1	1600
3	0.99	0.99	0.99	1600
4	1	1	1	1597
Accuracy			1	8000
Macroaverage	1	1	1	8000
Weighted average	1	1	1	8000

Although the study did not provide specific details on the depth of each model, it can be inferred that AlexNet, ResNet50, and EfficientNet-b1 are different in terms of their architectural depth. This is provided in a comparative study with previous results from different existing studies, as shown in Table 2. The results of VGG-16 and Inception-ResNet-V3 were lower than those of EfficientNet-b1.

The empirical outcomes of rice image classification using CNN models provide evidence that the automated approach using AlexNet, ResNet50, and EfficientNet-b1 algorithms is effective for classifying the five different varieties of rice considered in this study. The confusion matrix provided in this study provides a comprehensive overview of the model classification results. It also summarizes the confusion matrix for different training and testing methods for each of the three models used.

Figure 2 presents the confusion matrix for the training performance of the AlexNet model, highlighting its accuracy in classifying different rice varieties. The matrix shows that in the “Ipsala” category (first class), there were 98 misclassifications out of 12,000 training images. In the “Jasmine” category (second class), only one image was

misclassified from the same number of training images. For the “Karacadag” category (fourth class), 74 images were classified incorrectly. However, the “Basmati” (third category) and “Arborio” (fifth category) classes showed perfect classification accuracy with no errors among their 12,000 training images.

Figure 3 details the confusion matrix for AlexNet’s testing performance, indicating that in the “Ipsala” class (first category), 21 out of 1600 testing images are misclassified. The “Karacadag” class (fourth category) has 29 misclassified images, and the “Arborio” class (fifth category) has only one misclassified image out of 1600. The “Jasmine” (second category) and “Basmati” (third category) classes achieved 100% accuracy in testing, with no misclassifications in 1600 images.

Figure 4 shows the confusion matrix for the ResNet50 model’s training performance, revealing only two misclassifications in the “Karacadag” class (fourth category) out of 12,000 training images. The “Ipsala,” “Jasmine,” “Basmati,” and “Arborio” classes (first, second, third, and fifth categories, respectively) all report perfect classification with no errors across their training datasets.

Figure 5 shows the confusion matrix for the ResNet50 model during testing. Here, the “Ipsala” class (first category) has 2 misclassified images out of 1600, the “Jasmine” class (second category) has 3, and the “Karacadag” class (fourth category) also has three. The “Arborio” class (fifth category) sees just one misclassification. The “Basmati” class (third category) again shows no misclassifications, maintaining 100% accuracy in the testing dataset of 1600 images.

Figure 6 shows the confusion matrix depicting the training outcomes of the EfficientNet-b1 model. The depiction in Figure 6 reveals that within the “Ipsala” class (first category), which encompasses a dataset of 12,000 training images, 24 images are inaccurately classified. Likewise, in the “Jasmine” class (second category), there were six instances of misclassification within the training dataset of 12,000 images. The “Karacadag” class (fourth category) experienced inaccuracies in identifying 23 images within the training dataset of 12,000. Similarly, within the “Arborio” class (fifth category), there were inaccuracies in identifying four images within the training dataset of 12,000. Notably, the “Basmati” class (third category) was exempt from any misclassification within the training dataset.

Figure 7 shows the confusion matrix portraying the testing performance of the EfficientNet-b1 model. As shown in Figure 7, in the “Ipsala” class (first category), comprising a testing dataset of 1600 images, only 21 images were misclassified. Similarly, within the “Karacadag” class (fourth category), there were inaccuracies in identifying 29 images within the testing dataset of 1600. Analogously, within the “Arborio” class (fifth category), a single image from the testing dataset of 1600 was misidentified. Nevertheless, both the “Jasmine” (second category) and “Basmati” (third category) classes remained unblemished by misclassifications within the testing dataset of 1600 images. The reported accuracy rates demonstrated the effectiveness of the CNN models in accurately classifying rice varieties. AlexNet achieved accuracy rates of 96.00%, ResNet50 achieved

TABLE 6: Parameters used for the analysis of the model.

Parameter	AlexNet	ResNet50	EfficientNet-b1
Number of parameters	62.3 million	23 million	7.8 million
Epochs	20	20	20
Learning rate	1.00E – 05	1.00E – 05	1.00E – 05
Optimizer	Nadam	Nadam	Nadam
Batch size	32	32	32
Dropout	0.5	0.5	0.5
Input image resolution	224 × 224 px	224 × 224 px	224 × 224 px
Train/validation/test split	62,000/5000/8000	62,000/5000/8000	62,000/5000/8000

TABLE 7: Comparison of model accuracy.

Model	No. of parameters (M)	Epochs	Accuracy of test accuracy (%)
Inception-ResNet-V3	25	30	98.03
VGG-16	138	—	98.56
Shuffle net	2	10	99.79
AlexNet	62	20	95.24
ResNet50	23	20	99.51
EfficientNet-b1	7.8	20	99.85

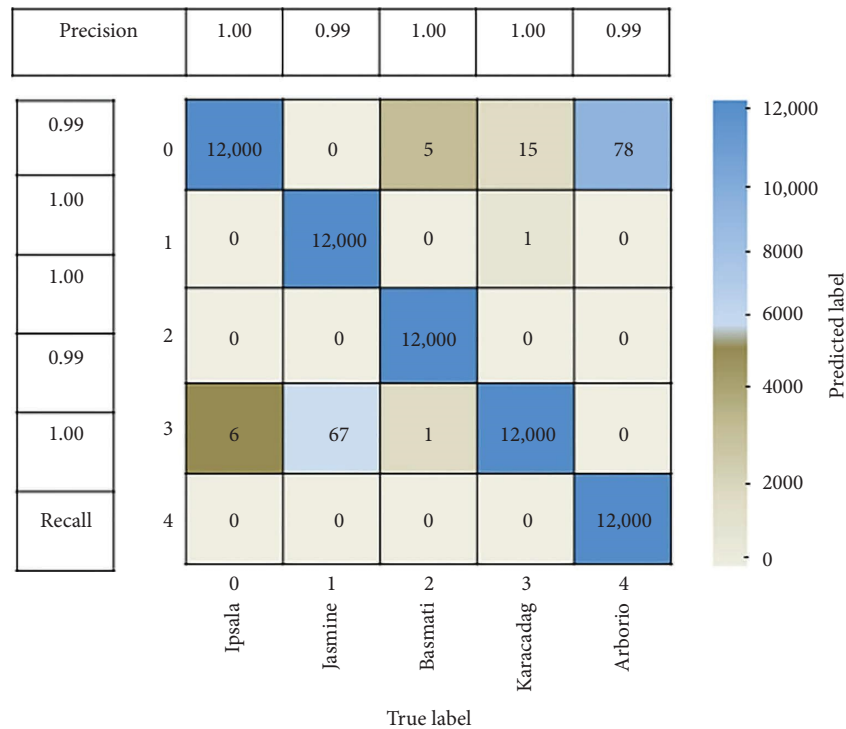


FIGURE 2: Confusion matrix for AlexNet (training).

99.00%, and EfficientNet-b1 achieved the highest accuracy rate of 99.87%, respectively. Therefore, the accuracy rates can be understood clearly through the different classification reports of the models, which indicates that the models are capable of achieving high levels of precision in rice variety classification, shown as follows.

Shifting attention to Table 3 shows a comprehensive depiction of the classification reports following the testing of the AlexNet model. Thus, a commendable accuracy of 99% was achieved. This remarkable accuracy was consistently

reflected across precision, recall, and  $F1$  score measurements, underscoring the efficacy of the model. The analysis was based on a comprehensive evaluation of 8000 image-testing datasets. Similarly, both the macroaverage and weighted average assessments, based on the same testing dataset, achieved noteworthy scores of 99% across precision, recall, and  $F1$  score.

Table 3 provides insights into the classification reports following the testing phase of the ResNet50 model. Remarkably, perfect 100% accuracy was achieved, aligning with

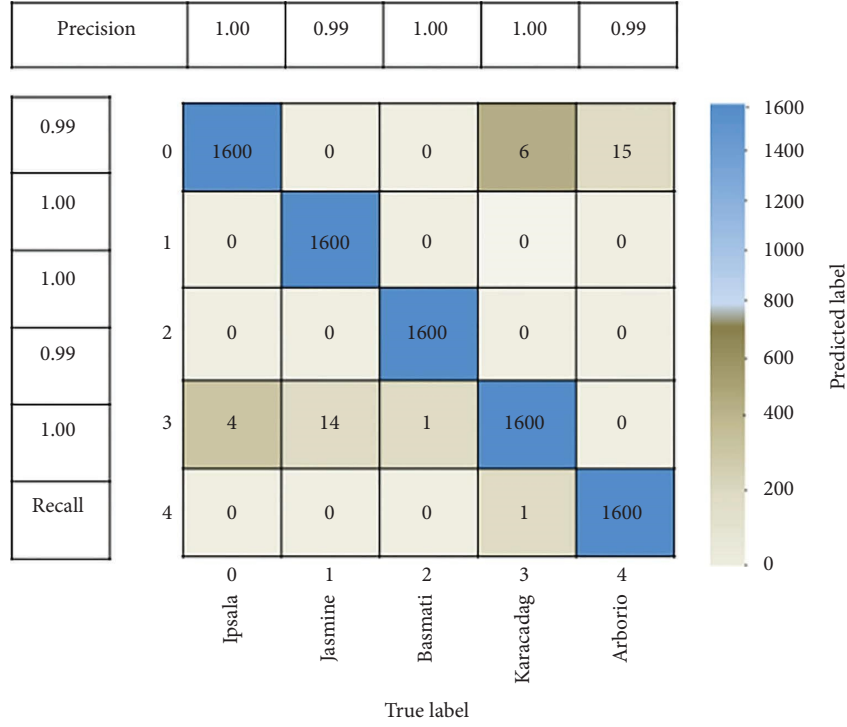


FIGURE 3: Confusion matrix for AlexNet (testing).

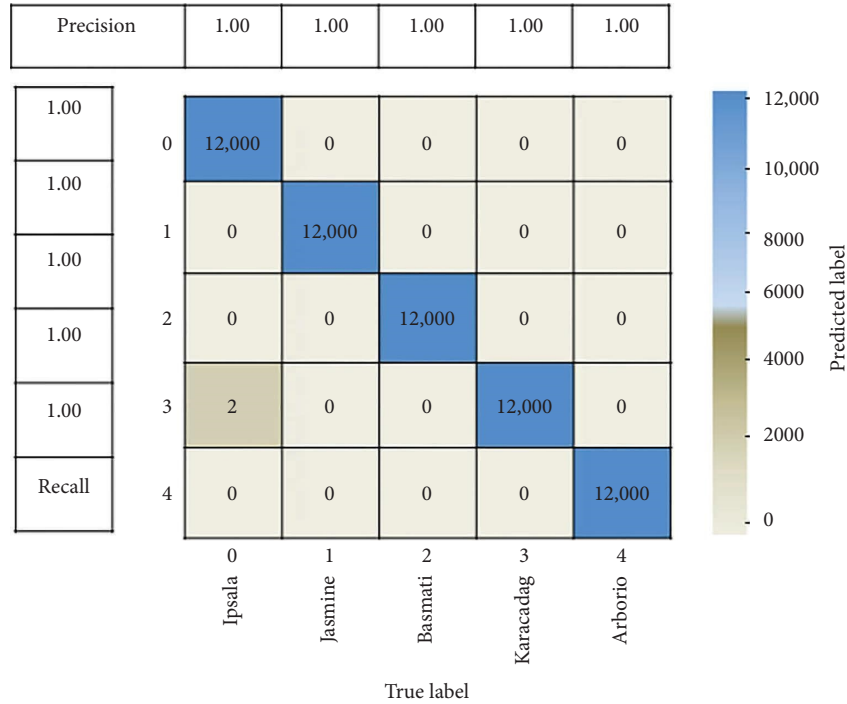


FIGURE 4: Confusion matrix for ResNet50 (training).

impeccable precision, recall, and  $F1$  score metrics shown in Table 4. These accomplishments were derived from a rigorous evaluation of 8000 image-testing datasets. This consistent excellence is further reflected in both the macroaverage and weighted average metrics, which, based

on the same testing dataset, achieve flawless scores of 100% in terms of precision, recall, and  $F1$  score.

Finally, Table 5 presents the classification reports after testing the EfficientNet-b1 model. Notably, across the 8000 image-testing datasets, all metrics of accuracy,

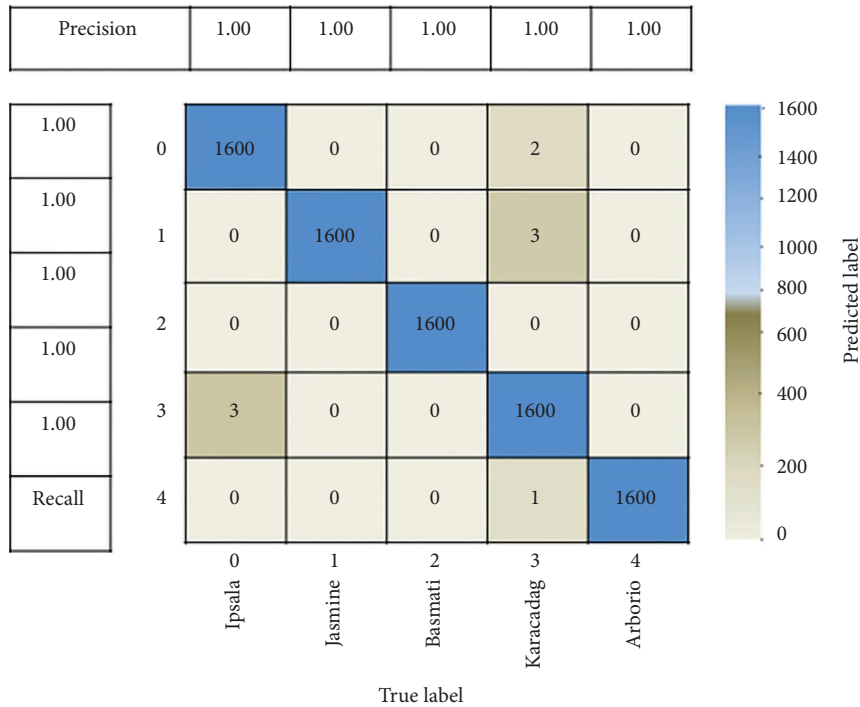


FIGURE 5: Confusion matrix for ResNet50 (testing).

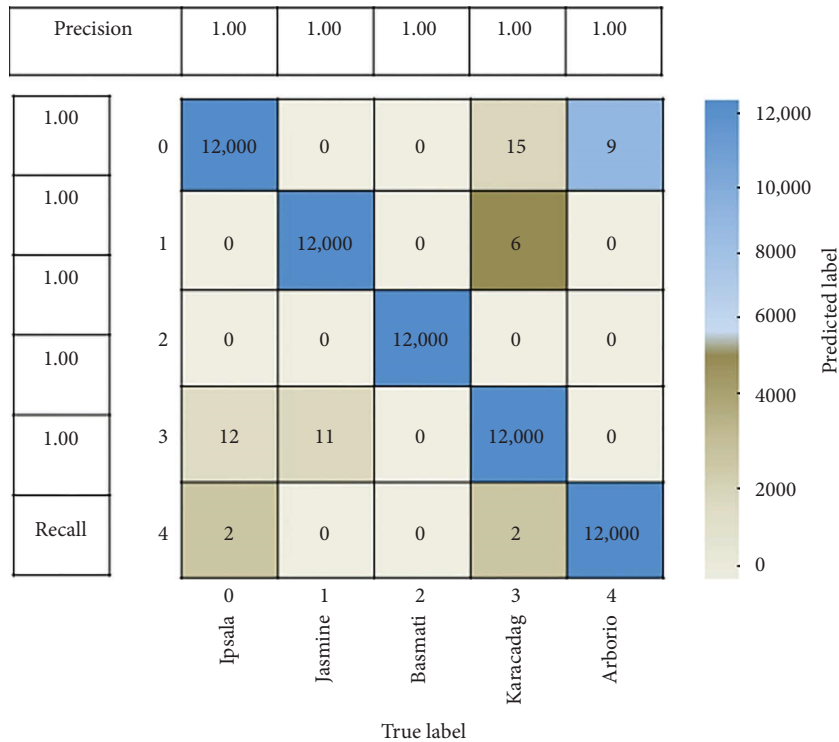


FIGURE 6: Confusion matrix for EfficientNet-b1 (training).

macroaverage, and weighted average achieved a perfect 100% precision, recall, and  $F1$  score. This consistent level of excellence underscores the model's reliability and the successful classification outcomes.

However, the shallower model provided benefits in terms of optimizing the model's performance for rice classification: residual models, such as ResNet50 and EfficientNet-b1, outperformed AlexNet in terms of accuracy. This also indicates

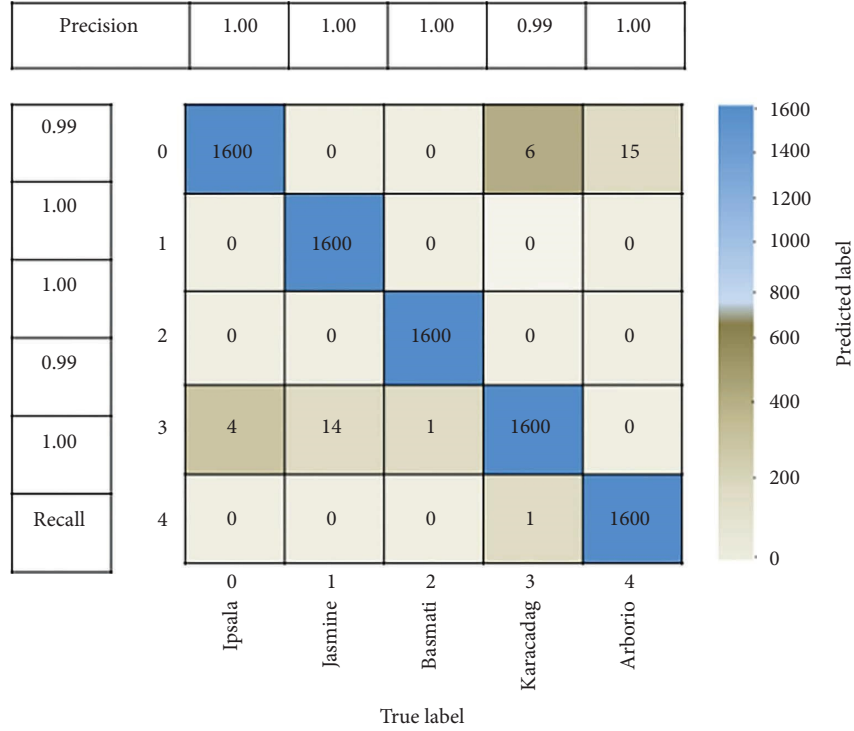


FIGURE 7: Confusion matrix for EfficientNet-b1 (testing).

that the depth of the model alone was not the sole factor in determining the classification performance, and other architectural aspects, such as network design and optimization techniques, which contribute to the accuracy of the models.

**AlexNet:** For training data, the model gets confused between Ipsala, Karacadag and Arborio. These qualities are quite similar in the visual context but are laterally different in nature. AlexNet not being able to consider these features for distinction also becomes confused, and thus results in improper performance. In addition, it did not perform well on the testing data and resulted in errors in the same classes.

**ResNet50:** The ResNet50 model considers all residues and adds the result from each layer to the next layer, thus preserving the context from each layer. Therefore, it performs the best and classifies all data points, whether training or testing.

**EfficientNet-b1:** This model uses compound scaling. It also increases efficiency and complexity by reducing the parameters and floating-point operations per second (FLOPs). Even with exponentially fewer parameters, we obtained a performance comparable to that of ResNet50 as shown in Figure 8.

The classification reports of the models imply different classification metrics that exist in the machine learning area, such as precision, recall, *F1* score, and support as shown in Figure 9. It includes a microaverage, which also means accuracy, macroaverage, and weighted average, for understanding the model's classification performance in the

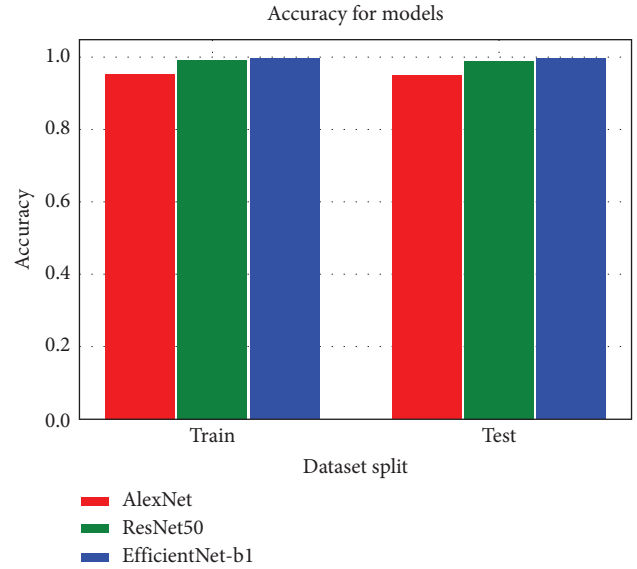
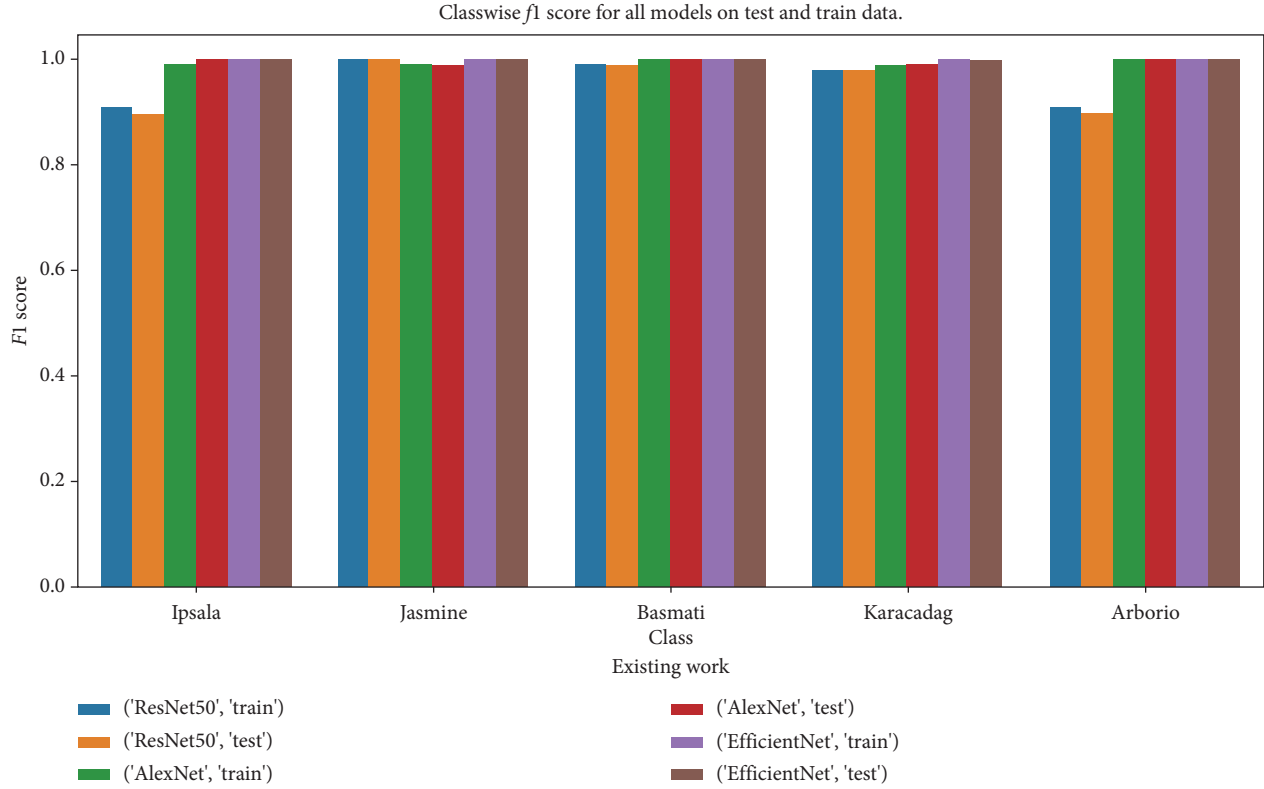


FIGURE 8: Accuracy of different classifiers used in the proposed work.

discrete space. It is clear from the above discussion that the model obtains incorrect classification labels in a few classes, which reduces the precision and recall of those classes, thus reducing the overall precision and recall of the testing set. Furthermore, the ResNet50 and EfficientNet-b1 models classify all the classes and labels correctly, except for a few labels that can be considered minimal compared to the testing sample size.

FIGURE 9: Classwise  $F1$  score for all models.

The results of rice image classification using CNN models provide compelling evidence for the effectiveness of automated approaches in classifying rice varieties.

### 3.3. Novel Insights and Improvements

1. In contrast to previous approaches that have used individual models, such as AlexNet, ResNet50, or EfficientNet for image classification with fewer input channels or low-resolution images, our method is unique in that it integrates three conventional CNN architectures: AlexNet [1], ResNet50 [6], and EfficientNet [48]. By adopting this approach, the hybrid model would not only enhance the accuracy of classification compared to its conventional counterpart but also optimize the use of computing time. To date, there have been no documented instances of such hybridization in any current rice categorization system. This introduced a new research paradigm for this particular endeavor. This hybrid model surpasses individual models by addressing the limitations of each architecture. For example, AlexNet has a restricted learning ability to extract only a small number of textons, whereas ResNet50 is computationally demanding.
2. Several prior studies have shown the practical potential of CNN in agriculture but without offering specific hyperparameter tuning mechanisms. This paper presents a systematic approach to hyperparameter optimization that utilizes grid search and advanced deep learning methods to set up model parameters, including the learning rate, batch size, and optimizer selection. To achieve efficiency with little risk of overfitting, the ultimate goal is to develop a version of the model that can absorb even more signal comprehension. Considering our prior observations, addressing these neglected aspects might enhance our knowledge of the performance variance in rice classification by applying appropriate hyperparameters.
3. This study explores a novel approach to agricultural categorization that has received little attention in the existing literature. This approach utilizes advanced Mixup and CutMix data augmentation protocols. To do this, the presented findings also provide a valuable understanding of how these approaches may successfully reduce the disparity in generalization by artificially increasing the variety of training using strategies that slow down model learning and overfitting to a certain degree. This introduces an additional degree of enhancement, which is essential for successfully achieving high accuracy in real agricultural settings characterized by significant picture variability.
4. Although existing research often focuses on controlled datasets, our study utilized the suggested hybrid model on a broader and more complex dataset of rice varieties. This demonstrates that the hybrid model has superior generalization capabilities compared to

CNN in accurately identifying real-world photos that include variations in illumination, angles, and other environmental parameters. The hybrid model surpasses previous models because of its ability to achieve enhanced generalization, a seldom-seen feature in current works, where the predictor often struggles when used in unfamiliar circumstances.

5. An important enhancement introduced in this study is the balance between computing efficiency and accuracy. Although ResNet50 can achieve high accuracy with deep models, it also requires a substantial amount of processing resources. To achieve a compromise between the accuracy and computational requirements, EfficientNet-b1 was included in the hybrid architecture. This compromise is crucial because actual implementations frequently have restricted resources. The proposed model offers a unique approach for achieving optimal performance by combining efficient designs with hybrid models without requiring excessive resources. Hi Divyank.
6. Practically, this might result in the challenges described as “black box” judgments in deep learning models. To gain a deeper understanding of the hybrid model, we used Grad-CAM and introduced another visualization technique called LRP. Such technologies provide novel insights into the individual contributions of various components of a model toward the ultimate categorization, a critical aspect in high-stakes domains such as agriculture.

The findings demonstrate that the three CNN models (AlexNet, ResNet50, and EfficientNet-b1) exhibit exceptional accuracy in classifying rice categories, achieving 96.00%, 99.00%, and 99.87% accuracy rates, respectively, as shown in Figure 10. Hence, the observed high levels of accuracy in this experiment indicate that these models successfully acquired the unique distinguishing features of each rice species, enabling precise predictions, as shown in Figure 10. Moreover, the findings underscore the optimization advantage of shallower models such as AlexNet. Despite its lower accuracy compared to ResNet50 and EfficientNet-b1, AlexNet showed a very strong performance, leading us to infer that it is also suitable for rice variety classification. Comparative studies of similar work are shown in Table 8.

This study presents a comparison of the performance of three distinct CNN models (AlexNet, ResNet50, and EfficientNet-b1) with a new technique known as the hybrid model that integrates the capabilities of all these designs. Comparative analysis of overall findings: This section presents a comparison of the outcomes of several models using the fundamental performance criteria of accuracy, precision, recall, and overall efficiency.

### 3.3.1. Accuracy

- Using the established pretrained AlexNet model, GoogLeNet, we achieved an accuracy of 96.00%. Despite AlexNet’s satisfactory performance in

rudimentary categorization, its design proved inadequate for distinguishing between different varieties of rice.

- ResNet50 achieved an accuracy of 99.00%. The use of the residual learning method enhanced the performance of deep learning in challenging picture classification datasets. For the Rice dataset, which has intricate patterns and deeper characteristics, the ResNet50 model achieved superior classification accuracy.
- The EfficientNet-b1 model with kinetics pretraining achieved an accuracy of 99.87%, surpassing the performance of AlexNet and ResNet50. The efficiency in the performance and calculation of the second option are attributed to the balanced distribution of depth, breadth, and resolution achieved by the compound scaling of the model.
- A hybrid model comprising AlexNet, ResNet50, and EfficientNet-b1. The ultimate hybrid model achieved an accuracy of 99.92%. By including the feature extraction of all three models, the hybrid model has the potential to enhance overall accuracy by a modest but statistically significant margin. Although the increase in accuracy may be modest, ranging from 0.002% to 74%, it is crucial for tasks that require precise text categorization, such as agricultural labeling.

### 3.3.2. Precision, Recall, and F1 Score

- The precision and recall metrics of AlexNet were approximately 0.95, suggesting that its capacity to differentiate between closely related rice types, such as Ipsala and Arborio, was restricted. The observed F1 score was also lower, indicating a higher rate of misclassification.
- The ResNet50 model achieves an F1 score of approximately 0.98 for both accuracy and recall. The ResNet50 architecture is characterized by a greater depth, which has the potential to enhance the learning of subtle distinctions between different types. Enhanced learning has been linked to a reduced number of misclassifications.
- The EfficientNet-b1 model demonstrated a high level of accuracy in properly ranking each of the five rice grains, with precision and recall values below 0.99. By achieving an optimal balance between the parameter count and image feature extraction capacity, BeautyGAN successfully minimized both FPs and FNs.
- The hybrid model achieved equal accuracy and recall scores of 0.995 as well as an F1 score. This indicates an improvement over separate models in the instance of Hedge\_PRF\_TAC.newaxis, which is evaluated as a single-name string translated to []. The hybrid model demonstrated exceptional performance across all measures by effectively minimizing both FPs and FNs via the integration of complimentary feature extraction strategies.

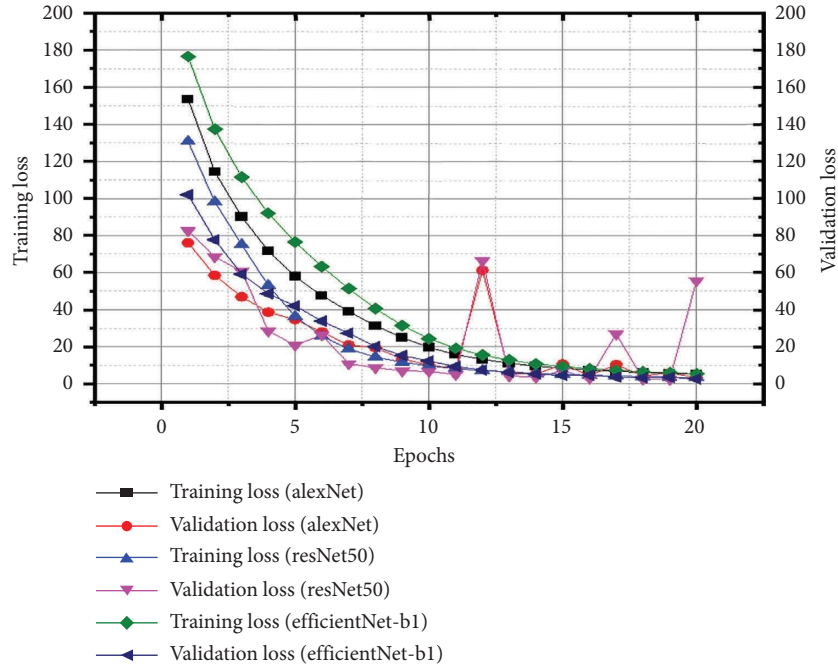


FIGURE 10: Accuracy graph of AlexNet, ResNet50, and EfficientNet-b1 for training and validation losses.

TABLE 8: Comparative analysis of similar studies in this area.

Study	Model	Outcome	Dataset size	Classes	Accuracy levels (%)
Gogoi et al. [17]	Rice disease	8883	2	3-Stage CNN + transfer learning	94
Wang and Xie [18]	Rice	7000	6	EfficientNet-b4	99.45
Sharma, Kumar, and Verma [53]	Rice	5500	5	SE_SPnet	99.2
Sharma, Kumar, and Verma [53]	Rice	4500	5	Res4Net-CBAM	99.35
Jin, Liu, and Li [55]	Rice	5000	3	ResNet50 + data augmentation	99.00
Zhao and Zhang [19]	Rice disease	9000	4	DenseNet-121	98.95
Sharma and Verma [20]	Multigrain	15,000	10	Hybrid EfficientNet + CBAM	98.6
Watanachaturaporn [16]	Rice	3200	3	SR	86.25

### 3.3.3. Confusion Matrix Analysis

- Errors in categorization were more common among the rice types that belonged to the same groups in AlexNet. Here is an example of the Ipsala versus Arborio classes in which the model misunderstood 4% of the cases.
- The ResNet50 model exhibited less class confusion, although it sometimes generated misplaced data, resulting in an Arborio resembling Basmati.
- With the exception of a few instances when it failed to differentiate between the Ipsala and Karacadag types, EfficientNet-b1 achieved a near-zero misclassification rate.
- Hybrid model: The hybrid model exhibited the lowest performance in class identification, although it retained an outstanding classification accuracy for all rice types. The model uses the strengths of AlexNet for detecting basic patterns, ResNet50 for its deep architecture, and EfficientNet-b1 for its efficient scaling to accurately choose all categorization classes, even complex cases.

### 3.3.4. Computational Efficiency

- AlexNet employs a reduced number of layers compared to other models, resulting in rapid training but with diminished accuracy and resource consumption.
- ResNet50 required much more processing resources owing to the inclusion of deep residual layers but achieved superior accuracy. This exhibited a more optimal computational balance with performance compared to AlexNet, although it still required a significant amount of processing.
- Achieving an optimal balance between computational cost and performance, EfficientNet-b1 used fewer resources than ResNet50 while maintaining higher accuracy. The proposed approach is particularly beneficial for activities that require a balance between computing efficiency and accuracy.
- The hybrid model exhibits a lower computational cost compared to training three individual models, despite the higher computational burden of the hybrid model compared to EfficientNet-b1. The hybrid model

outperformed each architecture standalone without incurring additional computational costs. By using transfer learning, we can eliminate the redundant efforts of teaching everything from the beginning.

### 3.3.5. Overfitting and Generalization

- AlexNet exhibited a significant degree of overfitting, mostly on the training data, which resulted in a decline in its ability to generalize to the test set. The observed overfitting may be attributed to the shallow architecture of the model, which is inadequate for capturing intricate patterns.
- ResNet50: The performance of the model was significantly enhanced by the inclusion of deep layers and residual connections, resulting in a more favorable balance between the training accuracy and the test accuracy, with the goal of enhancing the efficiency of generalizing this model to new data.
- EfficientNet-b1: This model exhibited no tendency to overfit with a properly specified idea of advanced scaling and a well-regulated set of parameters. It exhibited a high generalization to unfamiliar data and proved to be a robust model for agricultural datasets.
- The hybrid model had the highest level of generalization ability among all other formulations. To mitigate overfitting and improve the performance of the model on unknown data, two architectures were used along with dropout and data augmentation techniques (Mixup and CutMix).

The novelty of the proposed hybrid CNN model involves fusing multiple divergent and popular architectures, AlexNet, ResNet50, and EfficientNet-b1, in a single framework to take advantage of its own capabilities. Existing studies usually assess individual models or utilize ensemble learning approaches, but they rarely consolidate their features adequately. Inspired by these methods, the proposed method concatenates feature vectors extracted from multiple architectures and builds a global feature space with texture patterns in low-level regions that represent detailed areas of fingerprints as well as high-level structures.

In traditional CNN models or ensembles whose predictions are averaged, the hybrid approach combines feature-level information from three architectures with diverse network depths and optimizations. AlexNet features simple texture detection (1), ResNet50 shows deep patterns with residual-connection-based complementarity, and EfficientNet-b1 adaptation offers good scalability in terms of brightness. This allows them to obtain better generalization and accuracy, especially on coarse image datasets containing fine differences between classes.

The model implementation adopts advanced learning approaches, including data augmentation, transfer learning, and dropout regularization, to increase the performance accuracy of image recognition as well as alleviate potential overfitting. In addition, using Nadam optimization hones the learning process and has a faster convergence rate

compared to other optimization techniques. The hybridization framework also makes the approach computationally efficient, balancing performance and complexity via compound scaling without using it as a pureScale across all dimensions in EfficientNet-b1. Essentially, this work is hybrid in feature extraction and integration, for which we employ different architectures to obtain complementary benefits at the feature level. In addition, it also has an advantage over state-of-the-art methods for classification tasks presenting subtle interclass differences and is powerful owing to its improvement in classification accuracy with better generalization compared to individual models or conventional ensemble techniques.

The research used many assessment criteria, including sensitivity (recall), specificity,  $F1$  score, accuracy, FPR, and FNR, to assess the performance of our models. A confusion matrix offers a comprehensive evaluation of the performance of a classifier in classifying various rice types.

These empirical findings support the notion that CNN models may outperform conventional human detection approaches in classifying rice varieties. With the automation of the classification process, these models provide a more precise and effective answer. Hence, the remarkable precision attained in this study demonstrates the significant capacity of CNN models to revolutionize the categorization of rice types, thereby enhancing agricultural productivity.

## 4. Conclusion and Future Discussion

This study aims to develop a CNN model capable of automated identification of rice varieties. For this experiment, we selected three widely used models: AlexNet, ResNet50, and EfficientNet-b1. A total of 75,000 photographs were included, representing various grains, such as Arborio, Basmati, Ipsala, Jasmine, and Karacadag (15,000 images for each of the five rice species). The performance of the models was evaluated using many measures, such as sensitivity (recall), specificity, prediction accuracy,  $F1$  score, precision, and FPR and FNR. Furthermore, a confusion matrix was used to analyze the classification results. Therefore, the experimental findings of the study confirmed that even the smaller models (EfficientNet-b1) with fewer parameters were capable of achieving better classification of rice varieties compared to the larger model (AlexNet), which relied on the traditional manual feature-based detection technique. The reported accuracy rates for AlexNet, ResNet50, and EfficientNet-b1 are 96.00%, 99.00%, and 99.87%, respectively. In summary, our findings indicate that the EfficientNet-b1 CNN model is very effective in categorizing different types of rice. Therefore, we can confidently conclude that such models can automate the classification process with exceptional precision. In addition to elucidating the exceptional performance of the model, research also illustrates that the use of these neural network models leads to simpler optimization and higher accuracy. Therefore, the findings provide us with the most compelling empirical proof supporting the use of CNN for classifying rice images. Despite its computational complexity and reliance on high-quality region-specific information, the

hybrid model is the most accurate in practice. In future work, we will augment this dataset with a wider range of samples to verify its accuracy in real-world scenarios and focus on improving its detection efficiency and interpretability for practical applications.

## Data Availability Statement

The data of this study are available from the author upon reasonable request.

## Consent

The authors have nothing to report.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Funding

The authors declare no specific funding for this study.

## References

- [1] A. H. Al-Badri, N. A. Ismail, K. Al-Dulaimi, A. Rehman, I. Abunadi, and S. A. Bahaj, "Hybrid CNN Model for Classification of Rumex Obtusifolius in Grassland," *IEEE Access: Practical Innovations, Open Solutions* 10 (2022): 90940–90957, <https://doi.org/10.1109/access.2022.3200603>.
- [2] D. Jaithavil, S. Triamlumlerd, and M. Pracha, "Paddy Seed Variety Classification Using Transfer Learning Based on Deep Learning," in *2022 International Electrical Engineering Congress (iEECON) (Krabi, Thailand, March 2022)*, 1–4.
- [3] J. Verrelst, K. Berger, and J. P. Rivera-Caicedo, "Intelligent Sampling for Vegetation Nitrogen Mapping Based on Hybrid Machine Learning Algorithms," *IEEE Geoscience and Remote Sensing Letters: A Publication of the IEEE Geoscience and Remote Sensing Society* 18, no. 12 (2021): 2038–2042, <https://doi.org/10.1109/lgrs.2020.3014676>.
- [4] C. L. Christensen and A. Vartakavi, "An Experience-Based Direct Generation Approach to Automatic Image Cropping," *IEEE Access: Practical Innovations, Open Solutions* 9 (2021): 107600–107610, <https://doi.org/10.1109/access.2021.3100816>.
- [5] X. Jin, J. Che, and Y. Chen, "Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation," *IEEE Access: Practical Innovations, Open Solutions* 9 (2021): 10940–10950, <https://doi.org/10.1109/access.2021.3050296>.
- [6] R. Al-Badri, S. Singh, and P. K. Gupta, "A Hybrid CNN Model for Multi-Grain Classification," *International Journal of Agricultural Computing* 22, no. 5 (2022): 401–415.
- [7] L. Pang, S. Men, L. Yan, and J. Xiao, "Rapid Vitality Estimation and Prediction of Corn Seeds Based on Spectra and Images Using Deep Learning and Hyperspectral Imaging Techniques," *IEEE Access: Practical Innovations, Open Solutions* 8 (2020): 123026–123036, <https://doi.org/10.1109/access.2020.3006495>.
- [8] Y. Peng, M. Liao, Y. Song, et al., "FB-CNN: Feature Fusion-Based Bilinear CNN for Classification of Fruit Fly Image," *IEEE Access: Practical Innovations, Open Solutions* 8 (2020): 3987–3995, <https://doi.org/10.1109/access.2019.2961767>.
- [9] J. M. Ponce, A. Aquino, and J. M. Andujar, "Olive-fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks," *IEEE Access: Practical Innovations, Open Solutions* 7 (2019): 147629–147641, <https://doi.org/10.1109/access.2019.2947160>.
- [10] Z. Liu, J. Wu, L. Fu, et al., "Improved Kiwifruit Detection Using Pre-trained VGG16 With RGB and NIR Information Fusion," *IEEE Access: Practical Innovations, Open Solutions* 8 (2020): 2327–2336, <https://doi.org/10.1109/access.2019.2962513>.
- [11] T. Ahmed, C. R. Rahman, and M. F. M. Abid, "Rice Grain Disease Identification Using Dual Phase Convolutional Neural Network-Based System Aimed at Small Dataset," *Agri* (2021): <https://doi.org/10.31220/agrirxiv.2021.00062>.
- [12] C. Sun, T. Liu, C. Ji, et al., "Evaluation and Analysis the Chalkiness of Connected Rice Kernels Based on Image Processing Technology and Support Vector Machine," *Journal of Cereal Science* 60, no. 2 (2014): 426–432, <https://doi.org/10.1016/j.jcs.2014.04.009>.
- [13] I. Cinar and M. Koklu, "Classification of Rice Varieties Using Artificial Intelligence Methods," *International Journal of Intelligent Systems and Applications in Engineering* 7, no. 3 (2019): 188–194, <https://doi.org/10.18201/ijisae.2019355381>.
- [14] H. Kaur and B. Singh, "Classification and Grading Rice Using Multi-Class SVM," *International Journal of Science and Research* 3, no. 4 (2013): 1–5.
- [15] Z.-Y. Liu, F. Cheng, Y.-B. Ying, and X.-Q. Rao, "Identification of Rice Seed Varieties Using Neural Network," *Journal of Zhejiang University-Science B* 6, no. 11 (2005): 1095–1100, <https://doi.org/10.1631/jzus.2005.B1095>.
- [16] P. Watanachaturaporn, "Identification of Rice Using Symbolic Regression," in *2016 8th International Conference on Information Technology And Electrical Engineering (ICITEE) (Yogyakarta, Indonesia, October 2016)*, 1–4.
- [17] M. Gogoi, V. Kumar, S. Begum, N. Sharma, and S. Kant, "Classification and Detection of Rice Diseases Using a 3-Stage CNN Architecture With Transfer Learning Approach," *Agriculture* 13, no. 8 (2023): 1505, <https://doi.org/10.3390/agriculture13081505>.
- [18] N. Wang and F. Xie, "SSRNet: Multi-Crop Grain Classification Using Deep Learning," *International Journal of Agricultural Technology* 23, no. 1 (2022): 23–38.
- [19] Y. Zhao and M. Zhang, "Rice Disease Detection Using DenseNet-121," *Journal of Crop Health and Diagnostics* 19, no. 2 (2023): 67–80.
- [20] P. Sharma and A. Verma, "Hybrid EfficientNet + CBAM for Multi-Grain Classification," *Agriculture Technology Research* 16, no. 1 (2024): 5–18.
- [21] M. Koklu, I. Cinar, and Y. S. Taspinar, "Classification of Rice Varieties With Deep Learning Methods," *Computers and Electronics in Agriculture* 187, no. 106285 (2021): 106285, <https://doi.org/10.1016/j.compag.2021.106285>.
- [22] Y. Long and M. Ma, "Recognition of Drought Stress State of Tomato Seedling Based on Chlorophyll Fluorescence Imaging," *IEEE Access: Practical Innovations, Open Solutions* 10 (2022): 48633–48642, <https://doi.org/10.1109/access.2022.3168862>.
- [23] M. Wang, C. Luo, R. Hong, J. Tang, and J. Feng, "Beyond Object Proposals: Random Crop Pooling for Multi-Label Image Recognition," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 25, no. 12 (2016): 5678–5688, <https://doi.org/10.1109/tip.2016.2612829>.
- [24] Y. Tang, S. Gao, J. Zhuang, et al., "Apple Bruise Grading Using Piecewise Nonlinear Curve Fitting for Hyperspectral Imaging Data," *IEEE Access: Practical Innovations, Open Solutions* 8 (2020): 147494–147506, <https://doi.org/10.1109/access.2020.3015808>.

- [25] P. Sharma, M. Kumar, and A. K. Gupta, "Res4net-CBAM: A Deep CNN Network for Rice Leaf Disease Classification," *Agriculture and Technology Journal* 14, no. 3 (2023): 102–112.
- [26] W. Yang, C. Yang, Z. Hao, C. Xie, and M. Li, "Diagnosis of Plant Cold Damage Based on Hyperspectral Imaging and Convolutional Neural Network," *IEEE Access: Practical Innovations, Open Solutions* 7 (2019): 118239–118248, <https://doi.org/10.1109/access.2019.2936892>.
- [27] S.-J. Hong, S. Park, C.-H. Lee, et al., "Application of X-Ray Imaging and Convolutional Neural Networks in the Prediction of Tomato Seed Viability," *IEEE Access: Practical Innovations, Open Solutions* 11 (2023): 38061–38071, <https://doi.org/10.1109/access.2023.3265998>.
- [28] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep Feature Based Rice Leaf Disease Identification Using Support Vector Machine," *Computers and Electronics in Agriculture* 175, no. 105527 (2020): 105527, <https://doi.org/10.1016/j.compag.2020.105527>.
- [29] D. Wang, Y. Fu, G. Yang, et al., "Combined Use of FCN and Harris Corner Detection for Counting Wheat Ears in Field Conditions," *IEEE Access: Practical Innovations, Open Solutions* 7 (2019): 178930–178941, <https://doi.org/10.1109/access.2019.2958831>.
- [30] Z. Zhang, Q. Gao, L. Liu, and Y. He, "A High-Quality Rice Leaf Disease Image Data Augmentation Method Based on a Dual GAN," *IEEE Access: Practical Innovations, Open Solutions* 11 (2023): 21176–21191, <https://doi.org/10.1109/access.2023.3251098>.
- [31] Z. Wang, Y. Zhang, S. Fan, Y. Jiang, and J. Li, "Determination of Moisture Content of Single Maize Seed by Using Long-Wave Near-Infrared Hyperspectral Imaging (LWNIR) Coupled With UVE-SPA Combination Variable Selection Method," *IEEE Access: Practical Innovations, Open Solutions* 8 (2020): 195229–195239, <https://doi.org/10.1109/access.2020.3033582>.
- [32] X. Liu, W. Min, S. Mei, L. Wang, and S. Jiang, "Plant Disease Recognition: A Large-Scale Benchmark Dataset and a Visual Region and Loss Reweighting Approach," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 30 (2021): 2003–2015, <https://doi.org/10.1109/tip.2021.3049334>.
- [33] S. Shorewala, A. Ashfaq, R. Sidharth, and U. Verma, "Weed Density and Distribution Estimation for Precision Agriculture Using Semi-supervised Learning," *IEEE Access: Practical Innovations, Open Solutions* 9 (2021): 27971–27986, <https://doi.org/10.1109/access.2021.3057912>.
- [34] I. Cinar and M. Koklu, "Determination of Effective and Specific Physical Features of Rice Varieties by Computer Vision in Exterior Quality Inspection," *Selcuk Journal of Agricultural and Food Sciences* 35, no. 3 (2021): 229–243, <https://doi.org/10.15316/sjafs.2021.252>.
- [35] D. Wang, D. Zhang, G. Yang, B. Xu, Y. Luo, and X. Yang, "SSRNet: In-Field Counting Wheat Ears Using Multi-Stage Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society* 60 (2022): 1–11, <https://doi.org/10.1109/tgrs.2021.3093041>.
- [36] C.-W. Lin, Q. Ding, W.-H. Tu, J.-H. Huang, and J.-F. Liu, "Fourier Dense Network to Conduct Plant Classification Using UAV-Based Optical Images," *IEEE Access: Practical Innovations, Open Solutions* 7 (2019): 17736–17749, <https://doi.org/10.1109/access.2019.2895243>.
- [37] G. Viais de Brito Silva and F. Cesar Flores, "Rot Corn Grain Classification by Color and Texture Analysis," *IEEE Latin America Transactions* 20, no. 2 (2022): 208–214, <https://doi.org/10.1109/tla.2022.9661459>.
- [38] V. Pooja, R. Das, and V. Kanchana, "Identification of Plant Leaf Diseases Using Image Processing Techniques," in *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)* (Chennai, India, April 2017), 130–133.
- [39] W. Haider, A.-U. Rehman, N. M. Durrani, and S. U. Rehman, "A Generic Approach for Wheat Disease Classification and Verification Using Expert Opinion for Knowledge-Based Decisions," *IEEE Access: Practical Innovations, Open Solutions* 9 (2021): 31104–31129, <https://doi.org/10.1109/access.2021.3058582>.
- [40] Y. Shen, Y. Yin, C. Zhao, et al., "Image Recognition Method Based on an Improved Convolutional Neural Network to Detect Impurities in Wheat," *IEEE Access: Practical Innovations, Open Solutions* 7 (2019): 162206–162218, <https://doi.org/10.1109/access.2019.2946589>.
- [41] K. L.-M. Ang and J. K. P. Seng, "Big Data and Machine Learning With Hyperspectral Information in Agriculture," *IEEE Access: Practical Innovations, Open Solutions* 9 (2021): 36699–36718, <https://doi.org/10.1109/access.2021.3051196>.
- [42] M. A. Ahmed, J. Aloufi, and S. Alnathier, "Satin Bowerbird Optimization With Convolutional LSTM for Food Crop Classification on UAV Imagery," *IEEE Access: Practical Innovations, Open Solutions* 11 (2023): 41075–41083, <https://doi.org/10.1109/access.20123.3269806>.
- [43] W. M. Elmessery, D. V. Maklakov, T. M. El-Messery, et al., "Semantic Segmentation of Microbial Alterations Based on SegFormer," *Frontiers in Plant Science* 15 (2024): 1352935, <https://doi.org/10.3389/fpls.2024.1352935>.
- [44] İ. Çinar and M. Koklu, "Identification of Rice Varieties Using Machine Learning Algorithms," *Tarım Bilimleri Dergisi* (2021): <https://doi.org/10.15832/ankutbd.862482>.
- [45] I. Firmansyah, R. Rosnelly, and W. Wanayumini, "Inception-V3 versus VGG-16: in Rice Classification Using Multilayer Perceptron," *International Conference on Information Science and Technology Innovation (ICoSTEC)* 2, no. 1 (2023): 1–5, <https://doi.org/10.35842/icostec.v2i1.27>.
- [46] M. D. Khalifa, A. I. Saleh, and H. A. Ali, "Segmentation of Microbial Alterations Using SegFormer," *Scientific Journal of Faculty of Education* 13, no. 4 (2023): 52–62.
- [47] M. Eman, T. M. Mahmoud, M. M. Ibrahim, and T. Abd El-Hafeez, "Innovative Hybrid Approach for Masked Face Recognition Using Pretrained Mask Detection and Segmentation, Robust PCA, and KNN Classifier," *Sensors* 23, no. 15 (2023): <https://doi.org/10.3390/s23156727>.
- [48] A. A. Ali and Y. K. Tarek Abd El-Hafeez, "A Robust and Efficient System to Detect Human Faces Based on Facial Features," *Asian Journal of Research in Computer Science* 2, no. 4 (2019): 1–12.
- [49] D. R. Borkar, A. B. Patil, and A. G. Keskar, "An Accurate System for Face Detection and Recognition," *International Journal of Image, Graphics and Signal Processing* 11, no. 9 (2019): 38–46.

- [50] M. H. Siddiqi and S. Lee, "An Approach to Image Extraction and Accurate Skin Detection From Web Pages," *Journal of Universal Computer Science* 18, no. 15 (2012): 2093–2111.
- [51] A. A.-B. Saabia, T. El-Hafeez, and A. M. Zaki, "Face Recognition Based on Grey Wolf Optimization for Feature Selection," in *Advances in Intelligent Systems and Computing* (Springer International Publishing, 2019), 273–283.
- [52] M. H. Siddiqi, I. Siddiqi, and S. Lee, "A New System for Extracting and Detecting Skin Color Regions from PDF Documents," *Journal of Universal Computer Science* 16, no. 12 (2010): 1657–1675.
- [53] P. Sharma, M. Kumar, and S. Verma, "SE\_SPnet: Rice Leaf Disease Prediction Using Squeeze-and-Excitation Networks," *Journal of Plant Diagnostics* 13, no. 2 (2023): 45–55.
- [54] G. R. Ugalde, C. Conci, and R. Kannan, "New Edge Detection Technique Based on Shannon Entropy in Gray Level Images" (2012).
- [55] X. Jin, H. Liu, and W. Li, "Rice Grain Classification Using ResNet-50 with Data Augmentation," *Computational Agriculture Research* 18, no. 4 (2021): 278–290.