Journal of Agriculture and Food Research 15 (2024) 100890

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

# A comparative study of state-of-the-art deep learning architectures for rice grain classification

Farshad Farahnakian *, Javad Sheikh, Fahimeh Farahnakian, Jukka Heikkonen

*Department of Computing, University of Turku, 20500, Turku, Finland*

ARTICLE INFO

*Keywords:*
Rice grain classification
Deep learning
Machine learning
Automatic feature extraction

ABSTRACT

Accurate and efficient automated rice grain classification systems are vital for rice producers, distributors, and traders, offering improved quality control, cost optimization, and supply chain management. They also hold the potential to aid in the development of rice varieties that are more resistant to disease, pests, and environmental stress. While most existing studies in the rice classification domain rely on traditional machine-learning techniques that necessitate feature extraction engineering processes, our research explores the effectiveness of novel deep-learning models for this task. We evaluated the performance of various contemporary deep-learning models, including Residual Network (ResNet), Visual Geometry Group (VGG) network, EfficientNet, and MobileNet. These models were tested on a dataset comprising 75,000 images, classified into five different rice categories. We assessed each model using established evaluation metrics such as accuracy, F1 score, precision, recall, and per-class accuracy. Our findings showed that the EfficientNet-based model delivered the highest accuracy (99.67%), while the MobileNet-based model excelled in the speed of classification (2556 s). We concluded that, compared to traditional machine learning methods, the models employed in our study are highly scalable and capable of managing large volumes of complex data with millions of features and samples.

## 1. Introduction

Rice, one of the world's most significant agricultural products, is crucial for human nutrition, economies, and various industrial sectors [1,2]. Classifying rice varieties, an essential part of rice supply management is often time-consuming, energy-intensive, and expensive. With over 120,000 rice varieties categorized by the International Rice Research Institute (IRRI) based on milling degree, kernel size, starch content, and flavour [3], the need for automation in rice grain classification is evident. Recent advancements in Machine Vision present an opportunity for agricultural companies and rice suppliers to utilize this technology as an effective solution. Utilizing these innovative methods could bring significant advantages to the sector, including improved quality control, cost optimization, and streamlined supply chain management.

The combination of image processing and traditional Machine Learning (ML) techniques has shown promise in various agricultural applications, including disease detection, species classification, and quality analysis [4]. However, these techniques can be computationally demanding and slow when handling large datasets [5]. Moreover, the extraction and selection of appropriate features to construct an efficient rice classification system can be laborious and time-consuming, requiring considerable human effort.

On the other hand, cutting-edge Deep Learning (DL) models have demonstrated superior performance with complex and large datasets across various domains, including maritime, robotic, military, and agriculture [6]. These models can learn and extract complex features automatically, thereby addressing the shortcomings of traditional ML methods. Unlike ML models, DL models can identify abstract patterns in data through multiple layers of artificial neural networks, leading to more accurate and efficient results. Furthermore, DL models' high adaptability allows them to learn and adjust to new and evolving datasets, making them suitable for diverse applications.

In this study, we explore the potential of these popular DL models for the rice classification task, focusing specifically on a dataset with five different classes and over 75,000 images. We compare the performance of various DL models, including Residual Network (ResNet), Visual Geometry Group (VGG) network, EfficientNet, and MobileNet, using common metrics such as precision, F1-score, recall, and accuracy. We answer this main research question in this paper: Which DL model is

* Corresponding author.
*E-mail address:* farfar@utu.fi (F. Farahnakian).

https://doi.org/10.1016/j.jafr.2023.100890
Received 1 August 2023; Received in revised form 2 November 2023; Accepted 20 November 2023
Available online 28 November 2023

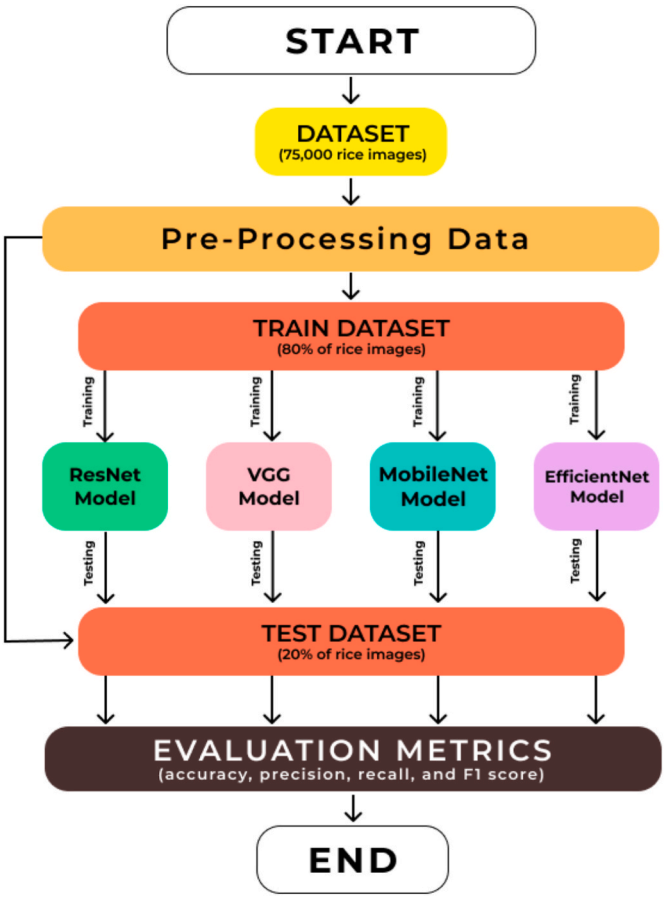


**Fig. 1.** Flowchart illustrating the methodology used in this study, which starts from the pre-processing step, and splitting data into train and test datasets to model training and evaluation process.

most effective in classifying rice, and how does its performance compare to others?

Our paper makes a significant contribution to the field of rice classification using DL models. To the best of our knowledge, this is the first study to examine the performance of multiple DL models, including ResNet50, ResNet101, VGG16, VGG19 EfficientNet, and MobileNet, on this classification task. Through a comprehensive evaluation, we provide insights into these models' performance and running time, which can inform future research in this area. Furthermore, our study contributes to advancing the state-of-the-art in rice classification, with implications for agricultural productivity, food logistics management, and food security.

The remainder of this paper is organized as follows: Section 2 presents a review of the relevant literature. Section 3 outlines our research methodology. The experimental setups like dataset and pre-processing have been explained in Section 4. Section 5 presents the results and findings of our study. Finally, Section 6 first discusses and compares this paper with other work, then summarizes the main findings and our methodology's broader implications.

## 2. Related work

Studies on agricultural product classification, particularly pertaining to rice species, have commonly employed image processing techniques to extract geometric parameters such as length and perimeter, fracture rate, and the presence of cracks in rice grains. The selected features are subsequently used to train machine learning (ML) algorithms for the classification task. The ML and DL models frequently used in these studies include Random Forest (RF), K-Nearest Neighbor (KNN), Multi-Layer Perception (MLP), Logistic Regression (LR), Support Vector Machine (SVM), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Deep Neural Network (DNN).

In [9], a total of 3810 images of two rice species were analyzed, with seven morphological characteristics identified for each grain. These features were then used to train models using ML algorithms including LR, MLP, SVM, Decision Trees (DT), RF, Naive Bayes (NB), and KNN, with performance measurements recorded. Similarly [7], used the same 75,000-image dataset selected for our study to construct a second dataset with 106 features, including 12 morphological, 4 shape, and 90 colour features. They trained DNN and ANN models using these features to classify rice varieties. They also compared these models' performance with a CNN-based classification model trained directly on the images.

In another study [8], 106 features were extracted from five different colour spaces, and ML models such as KNN, DT, LR, MLP, RF, and SVM were used for the rice classification task. As traditional ML models may struggle with large datasets due to their computational demands and slow processing times, some studies have limited their datasets. For instance Ref. [9], used a dataset of only 1700 images from two classes to identify the location and type of rice chalkiness, while [10] used a dataset of 5000 images from three classes to train a DNN architecture for improving rice classification accuracy.

Upon reviewing the literature, it becomes evident that the majority of studies have employed traditional ML methods that necessitate feature extraction. While these methods have demonstrated effectiveness, the laborious and time-consuming process of feature extraction can limit the scalability of these models, especially with large, high-dimensional datasets. DL models, in contrast, provide an alternative that negates the need for feature extraction. These models can learn directly from raw data, simplifying the classification process and reducing computational costs.

## 3. Methodology

This section provides a detailed description of the experimental design, including the dataset and evaluation metrics. Moreover, we also briefly described the DL models and their architecture used in this study separately. A schematic representation of the entire methodology is provided in Fig. 1, which illustrates the flow of operations. It begins with the acquisition of the rice grain image dataset, followed by data pre-processing. Then, the preprocessed data is fed into the deep learning models, namely ResNet, VGG network, MobileNet, and EfficientNet, for training. Upon model training, the evaluation of their performance is carried out using several metrics such as accuracy, precision, recall, and F1-score.

### 3.1. Review of deep-learning models

The following subsections provide an exhaustive exploration of the deep learning architectures utilized in our study. These models include ResNets, VGG Network, EfficientNet, and MobileNet.

**Residual Network (ResNet):** ResNet, a variant of supervised, feed-forward deep neural networks, introduces a novel paradigm wherein each layer learns a transformation in reference to the layer's input, as opposed to learning independent mappings [12]. This architecture enables the training of exceptionally deep neural networks with reduced training error. Notably, ResNet's design effectively mitigates the vanishing gradient problem [13], a common issue faced during the training of deep networks using gradient-based learning methods. A standard ResNet block, as shown in Fig. 2, comprises two convolutional layers, integrating the input with the residual function's output for subsequent use in the following blocks.

**Visual Geometry Group (VGG) Network:** The VGG network illustrated in Fig. 3, originating from the Visual Geometry Group at Oxford University, is acclaimed for its simplicity and efficiency [14]. It primarily employs a series of 3x3 convolutional layers stacked in increasing

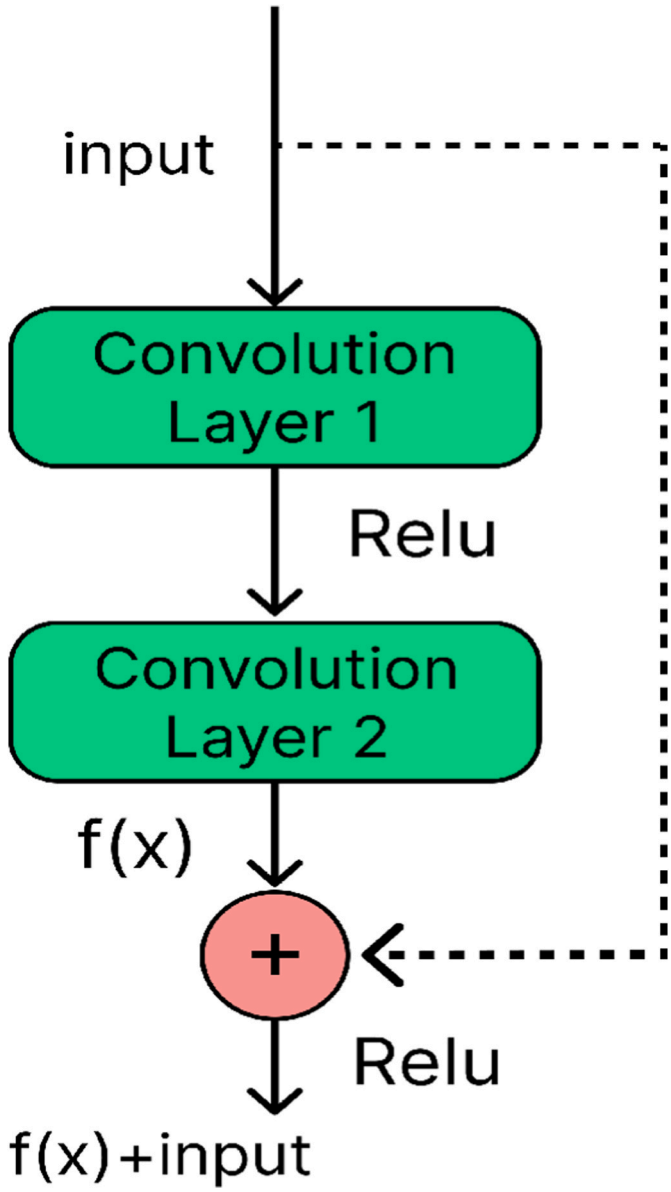


**Fig. 2.** A block of a ResNet model.

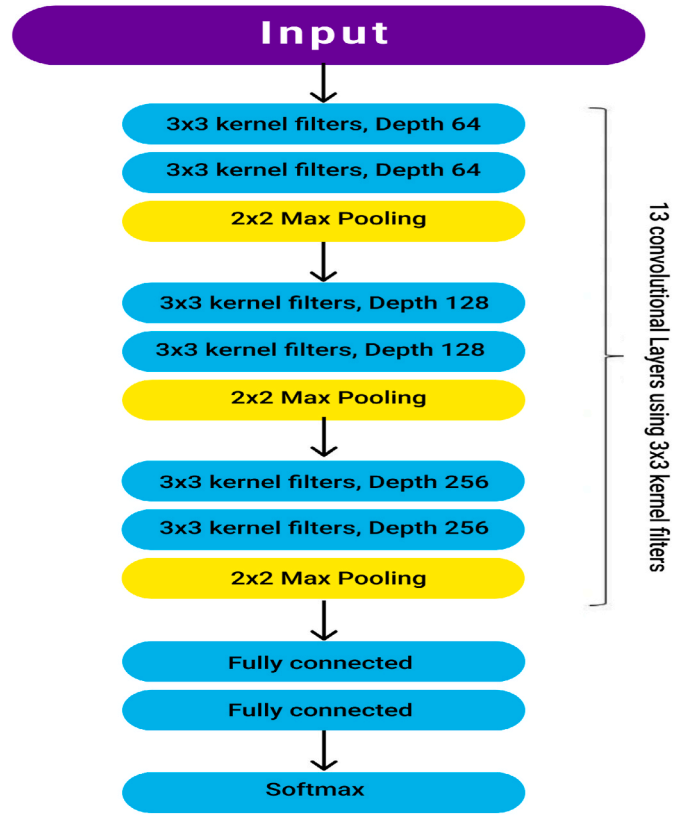


**Fig. 3.** Architecture of the VGG model.

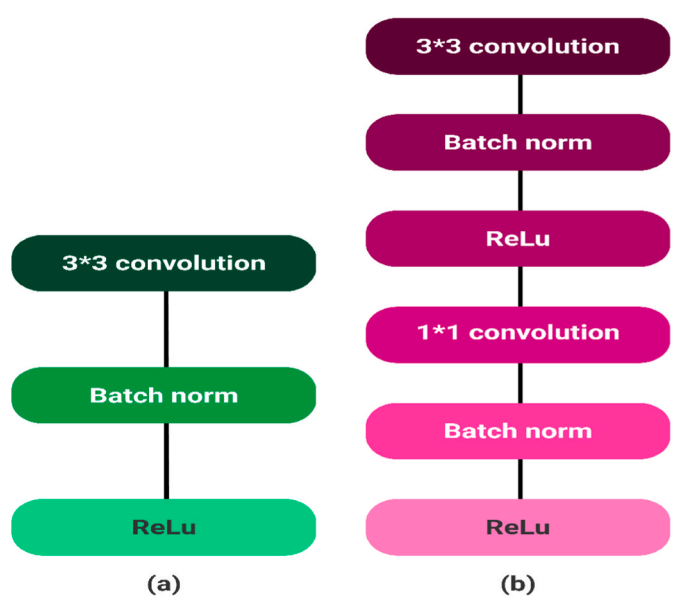


**Fig. 4.** (a) Standard convolution network (b) Depth-wise separable convolution with depth-wise and pointwise layers.

depth, reducing the number of parameters while allowing for the extraction of more complex features.

**MobileNet:** Designed for mobile and embedded vision applications, MobileNet, a suite of Convolutional Neural Networks (CNNs), is apt for various computer vision tasks, including object and face detection, along with logo or text recognition [15]. MobileNet stands out for its compactness, owing to fewer parameters, making it suitable for mobile applications. It is less complex compared to other models and requires fewer computational operations, leading to improved accuracy, decreased memory consumption, and reduced computational time [16]. A feature of MobileNet is its use of depthwise separable convolutions, leading to a significant reduction in the model's parameter count [16]. These convolutions involve a two-step process of filtering and combination, distinguishing MobileNet from conventional CNNs (Fig. 4).

**EfficientNet:** Scaling up a convolutional network is a common strategy to enhance accuracy on benchmark datasets. However, the traditional approaches, such as width-wise, depth-wise, and image resolution techniques, usually require careful manual tuning and are time-intensive. EfficientNet [16], designed as a systematic solution to this challenge, enables effective scaling of CNN models. Through comprehensive evaluation of different scaling methods [17], EfficientNet demonstrates that balancing the three dimensions-width, depth, and image resolution-with a set of coefficients can improve the model's performance, as illustrated in Fig. 5.

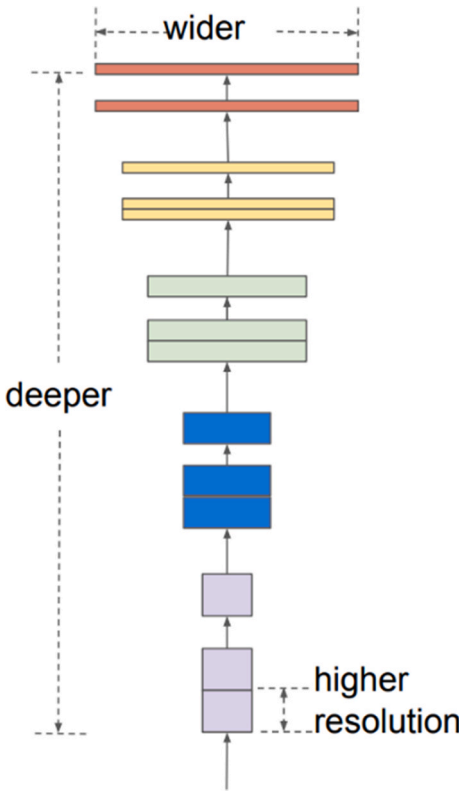


**Fig. 5.** Compound scaling technology in EffiecientNet model [18].

## 4. Experimental setup

### 4.1. Dataset

The dataset employed in this study comprises images of five distinct varieties of rice commonly cultivated in Turkey: Arborio, Basmati, Ipsala, Jasmine, and Karacadag [11]. It includes a total of 75,000 rice grain images, with each variety represented by 15,000 images. These images are in RGB format and boast a resolution of 250 x 250 pixels, each capturing a single grain of rice. This dataset was chosen due to its comprehensive representation of rice varieties commonly found in the Turkish agricultural sector, offering a practical context to evaluate the performance of DL models. Furthermore, the large volume of images and the diversity of rice types provide a challenging environment for DL models, facilitating a rigorous comparison of their performance. Fig. 6 provides some examples of each rice grain class from the dataset.

### 4.2. Pre-processing

Pre-processing is a vital step in image analysis as it can significantly enhance image quality, thereby improving the robustness and performance of the subsequent classification task. For this study, we employed one ubiquitous technique in image pre-processing: contrast enhancement. It is utilized to increase the discernibility of the details in the image by amplifying the difference in intensity between lighter and darker areas. This technique enhances the definition and clarity of the image, thereby making it easier for the model to identify and learn the intricate differences between various rice varieties. By implementing the pre-processing technique, we have been able to ensure that the deep learning models employed in this study are provided with high-quality, meaningful data, which in turn leads to more accurate and reliable classification results.

**Table 1**
Model parameters and time.

| Model | Time | Trainable |
|---|---|---|
| VGG16 | 7857 s | 6,456,325 |
| VGG19 | 9358 s | 6,456,325 |
| ResNet50 | 5902 s | 25,723,909 |
| ResNet101 | 9803 s | 25,723,909 |
| EfficientNet | 4304 s | 4,049,571 |
| MobileNet | 2556 s | 12,878,853 |

### 4.3. Implementation details

The implementation of this study was carried out using the Keras library in a Python environment. Computations were performed on a platform equipped with an 11th Generation Intel Core i7-11800H and an NVIDIA GeForce RTX 3050 Ti with 4 GB of GDDR6 memory. For the training of deep learning models, the Adam optimizer [19] was utilized, with a fixed learning rate of 0.0001. The models were trained in mini-batches, each consisting of 16 images, over the course of 100 epochs. During this training process, strategies such as early stopping and model checkpointing were employed to prevent overfitting and to save the most effective model, respectively. The available dataset was divided in such a way that 80% of it was dedicated to training and the remaining 20% was used for testing, thereby ensuring that the performance of the models was evaluated on unseen data.

### 4.4. Evaluation metrics

The performance of the classification models in this study has been evaluated using four common metrics, which are detailed below.

- **Accuracy**: This metric quantifies the proportion of correct predictions made by the model across all classes. It is calculated by dividing the number of correct predictions by the total number of predictions. The range of accuracy is between 0 and 1, with a score of 1 representing perfect prediction accuracy. Accuracy is calculated using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where *TP*, *TN*, *FP*, and *FN* are the number of true positives, true negatives, false positives, and false negatives, respectively.

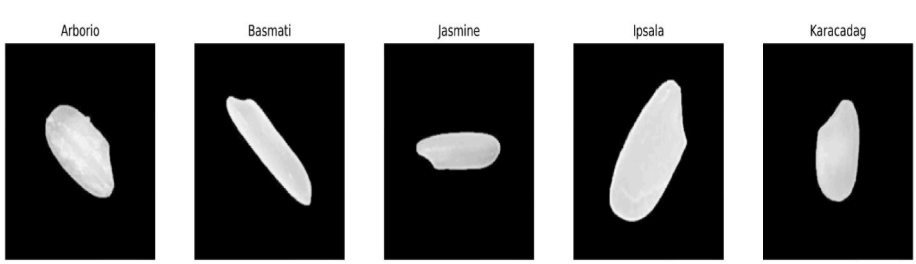


**Fig. 6.** Example of rice grain classes: (a) Arborio (b) Basmati (c) Ipsala (d) Jasmine (e) Karacadag.

**Table 2**
Model performance based on evaluation metric.

| Model | Accuracy | F1-score | Precision | Recall | Per-Class Accuracy | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Arborio | Basmati | Jasmine | Ipsala | Karacadag |
| VGG16 | 99.53 | 0.9953 | 0.99 | 0.99 | 99.5 | 99.53 | 99.43 | 99.66 | 99.53 |
| VGG19 | 99.57 | 0.9957 | 0.99 | 0.99 | 99.53 | 99.53 | 99.56 | 99.7 | 99.53 |
| ResNet50 | 99.53 | 0.9953 | 0.99 | 0.99 | 99.23 | 99.6 | 99.5 | 99.76 | 99.56 |
| ResNet101 | 99.51 | 0.9951 | 0.99 | 0.99 | 99.3 | 99.43 | 99.43 | 99.83 | 99.56 |
| EfficientNetB0 | 99.67 | 0.99673 | 0.99 | 0.99 | 99 | 99.8 | 99.76 | 99.93 | 99.86 |
| MobileNet | 98.86 | 0.9886 | 0.98 | 0.98 | 98 | 98.8 | 99.06 | 99.63 | 99.8 |

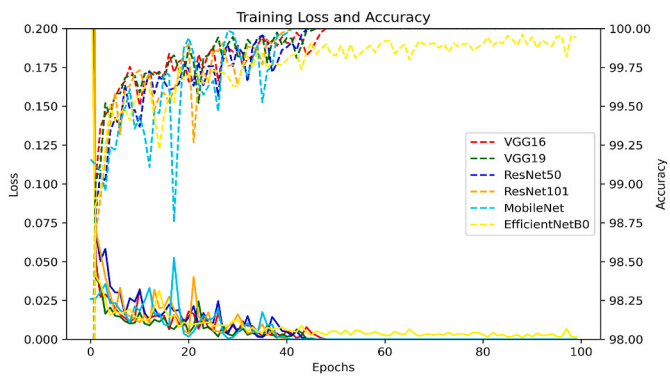


**Fig. 7.** Training loss and accuracy over 100 epochs for all models.

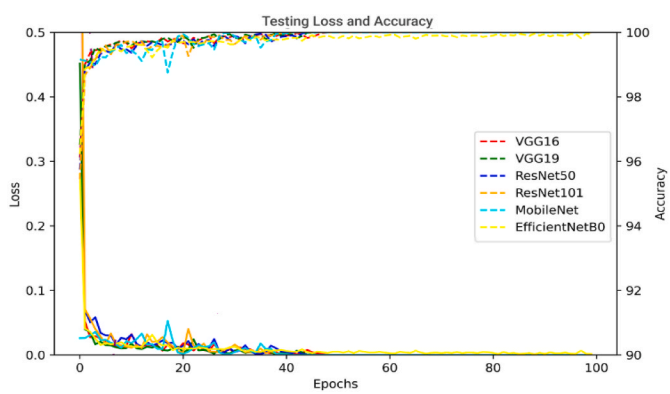


**Fig. 8.** Testing loss and accuracy over 100 epochs for all models.

- **Precision**: Precision measures the ratio of true positive predictions (correctly predicted positive instances) to the sum of true positives and false positives (all instances predicted as positive). A perfect precision score of 1 indicates that the model has a low false-positive rate. The formula for precision is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

where $TP$ and $FP$ are the numbers of true positives and false positives, respectively.

- **Recall**: Also known as sensitivity or the true positive rate, recall calculates the proportion of true positive predictions out of the total actual positive instances. A recall score close to 1 signifies that the model has a low false-negative rate. Recall metric is calculated using the formula 3:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

where $TP$ and $FN$ are the numbers of true positives and false negatives, respectively.

- **F1-score**: This metric combines precision and recall into a singular measure, providing a balanced assessment of these two metrics. Also known as the harmonic mean of precision and recall, the F1-score ranges from 0 to 1, with 1 being the optimal value. It effectively represents the trade-off between precision and recall. This metric is calculated using the following formula:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

where Precision and Recall are as defined above.

- **Per-class Accuracy (PA)**: It calculates the accuracy for each individual class separately. This is done by dividing the number of correctly classified instances of a particular class by the total number of instances of that class. Furthermore, it is often used in multi-class classification problems, and it ranges between 0 and 1. In mathematical terms, for a given class 'i', the per-class accuracy can be calculated as follows:

$$PA_i = (TP_i)/(TP_i + FN_i) \tag{5}$$

where: $TP_i$ is the number of true positives for class 'i', and $FN_i$ is the number of false negatives for class 'i'.

Although accuracy provides an overall measure of model performance, it can be misleading when dealing with imbalanced classes. Therefore, we also consider precision, which provides insight into the rate of false-positive errors, and recall, which indicates the rate of false-negative errors. These metrics are particularly important in our study, where missing a particular rice class (false negative) or misclassifying a rice class (false positive) could have significant implications. The F1 score offers a balance between precision and recall and is particularly useful when the cost of false positives and false negatives are very different. In our case, it would signify the trade-off between misclassifying a rice variety and missing a rice variety.

## 5. Experimental results

### 5.1. Running time vs performance

The experiment results pertaining to the running times and performance of various deep learning models are summarized in Table 1 and Table 2, respectively. These tables collectively provide a comprehensive evaluation of the models' capabilities in terms of their classification accuracy, precision, recall, F1 score, per-class accuracy, training times, and trainable parameters.

From a computational standpoint, as summarized in Table 1, MobileNet demonstrated the fastest training time among the evaluated models. This suggests that MobileNet provides a balance between computational efficiency and model performance, making it an attractive choice for applications where computational resources or training time may be a limiting factor. However, as indicated by its slightly lower precision and recall scores, this increased speed may come at the cost of a small compromise in model performance.

As demonstrated in Table 2, EfficientNet emerged as the top-performing model in terms of overall accuracy and F1 score. This

indicates that EfficientNet outperformed the other models in both correctly classifying the rice grain images and in providing a balance between precision and recall. Interestingly, all models, with the exception of MobileNet, achieved a high score of 0.99 for both precision and recall, suggesting that these models were highly effective in correctly identifying the rice classes with minimal false positives and negatives. The per-class accuracy sub-columns in Table 2 offer a more detailed analysis of the models' performance at the level of individual rice classes. VGG19 proved to be the best model for classifying the Arborio rice class. However, EfficientNet was superior in classifying the Basmati, Jasmine, Ipsala, and Karacadag rice classes, further reinforcing its status as the top-performing model.

As depicted in Figs. 7 and 8, the evolution of training and testing accuracy and loss over 100 epochs for all models is exhibited. It is apparent from Figs. 7 and 8 that most of the models achieve peak accuracy concurrently with a minimized loss on both the training and testing datasets at around epoch 40. This observation signifies that the models are capable of efficiently learning the feature representations from the dataset within this number of iterations, and further training beyond this point does not notably improve their performance.

### 5.2. Confusion matrix

The confusion matrix is utilized in this study as an additional performance evaluation tool, allowing for a detailed analysis of the tested models' performance in the rice grain classification task. The metrics derived from the confusion matrix include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). As depicted in Fig. 9, all models exhibit robust performance in accurately predicting the labels, with minimal errors reported. For instance, the MobileNet model has 53 instances and the EfficientNet model has 24 instances where the Arborio class has been incorrectly predicted as the Karacadag class. This analytical approach allows for an intricate understanding of where the models may struggle, thus providing insights into potential areas of improvement.

### 5.3. Qualitative results

In order to substantiate our quantitative findings, we conducted a

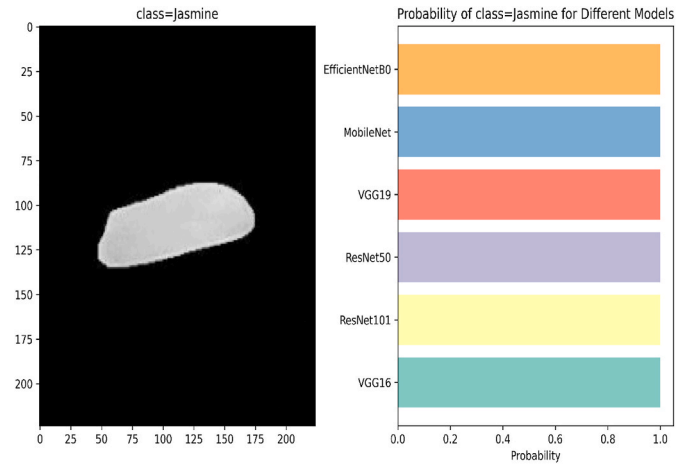


**Fig. 10.** Qualitative analysis of all DL models for Jasmine rice class.

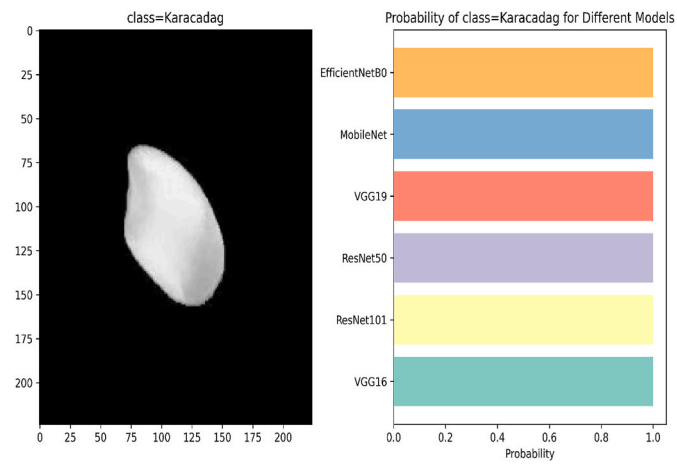


**Fig. 11.** Qualitative analysis of all DL models for Karacadag rice class.

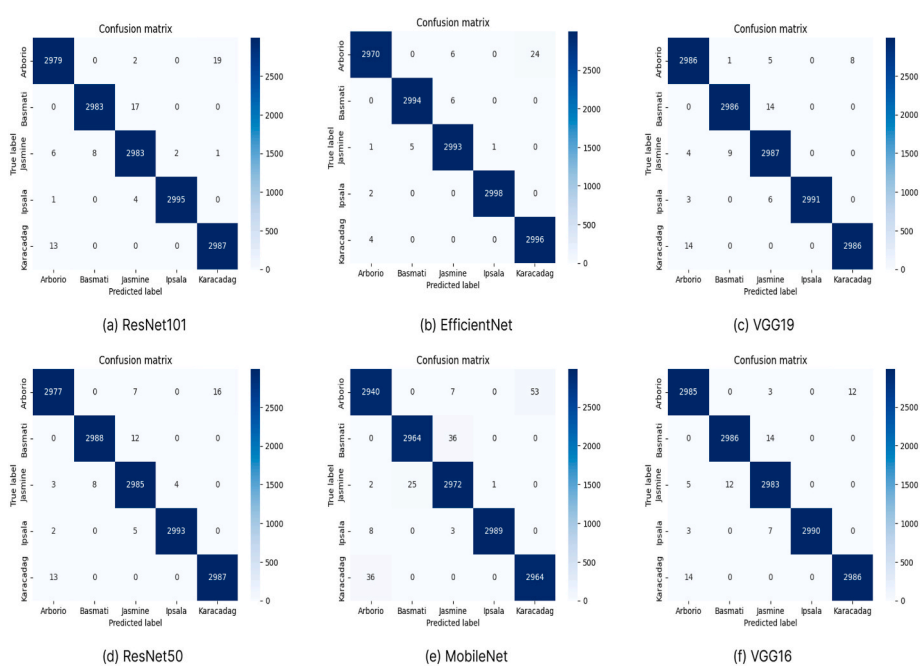


**Fig. 9.** Confusion matrix for all models.

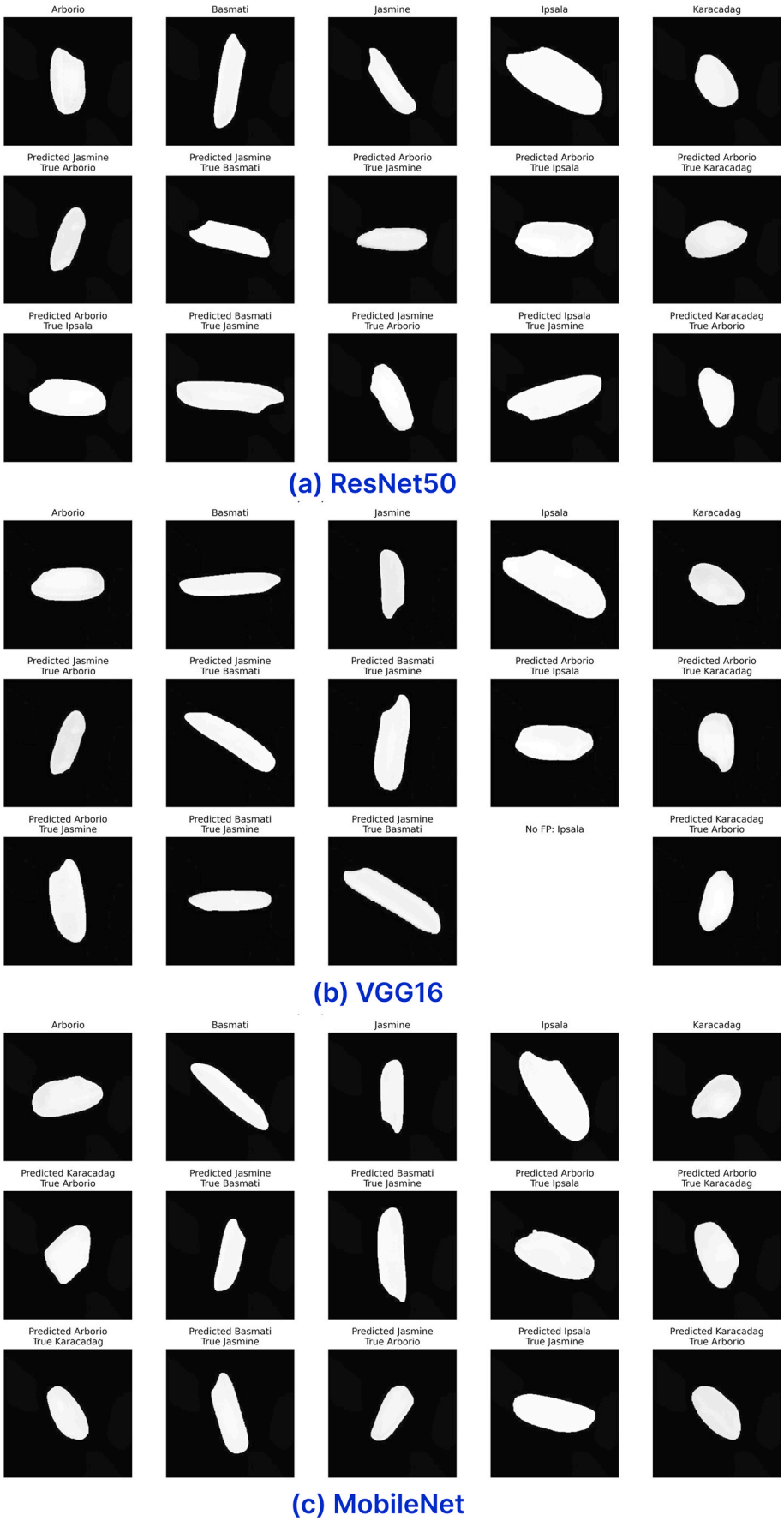


**Fig. 12.** Misclassification examples for (a) ResNet50 (b) VGG16 and (c) MovbileNet.

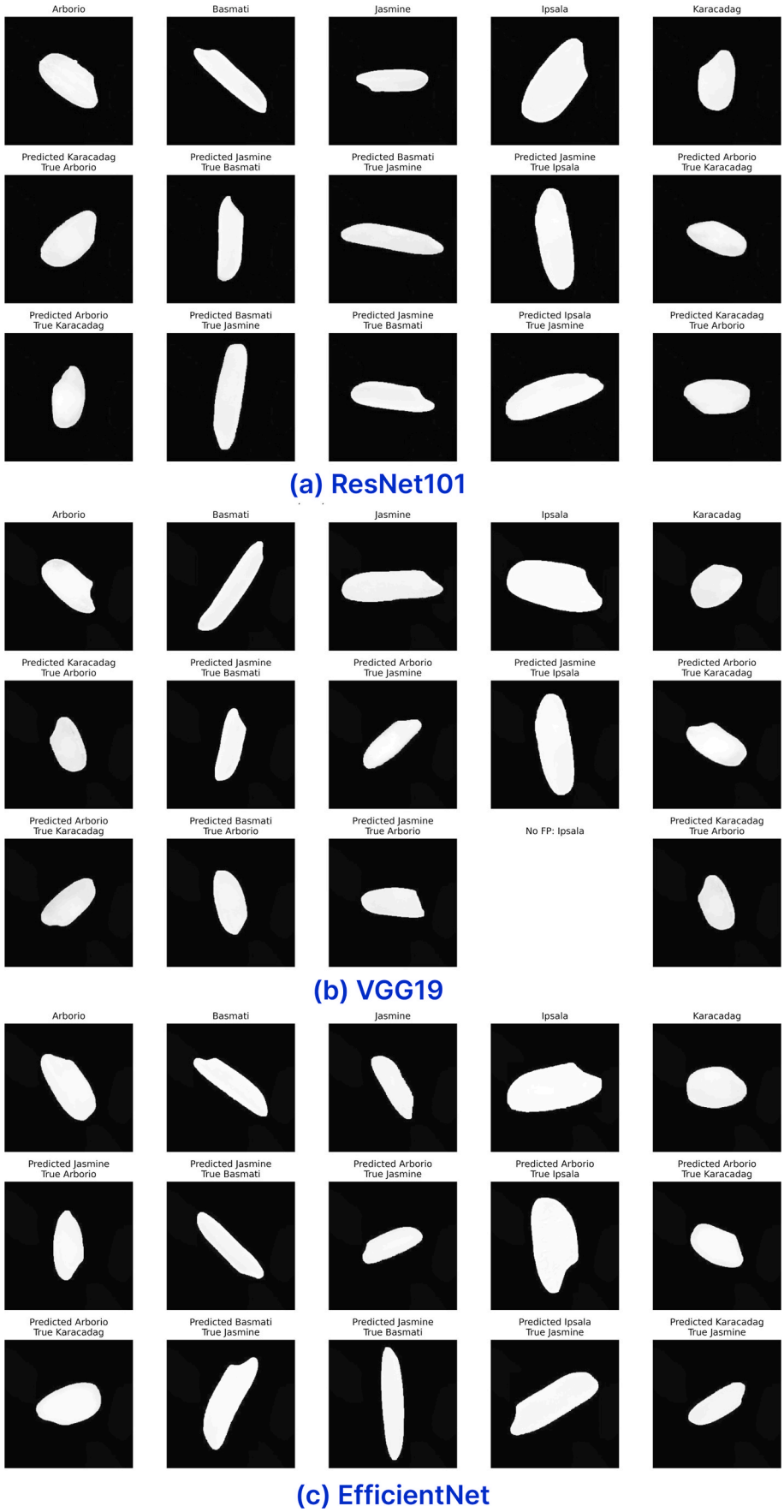


**Fig. 13.** Misclassification examples for (a) ResNet101 (b) VGG19 and (c) EfficientNet.

**Table 3**
Model performance based on evaluation metric.

| Scenario | Accuracy | F1-score | Precision | Recall | Per-Class Accuracy | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Arborio | Basmati | Jasmine | Ipsala | Karacadag |
| AlexNet | 0.9699 | 0.9699 | 0.9699 | 0.9699 | 0.9617 | 0.9730 | 0.9525 | 0.9874 | 0.9749 |
| FC6 | 0.9688 | 0.9688 | 0.9689 | 0.9688 | 0.9592 | 0.9580 | 0.9613 | 0.9893 | 0.9761 |
| FC7 | 0.9694 | 0.9694 | 0.9694 | 0.9694 | 0.9599 | 0.9687 | 0.9563 | 0.9893 | 0.9730 |
| FC8 | 0.9694 | 0.9694 | 0.9694 | 0.9694 | 0.9599 | 0.9687 | 0.9563 | 0.9893 | 0.9730 |

qualitative analysis aimed at providing a more granular understanding of the performance exhibited by the DL models in the rice classification task. This exercise necessitated a detailed examination of the model's ability to predict the correct class of individual test images. We selected two test images, each representing a different rice variety: Jasmine and Karacadag. These images were then individually fed into the trained DL models, which subsequently generated probability prediction scores. Each score represents the model's certainty or confidence in predicting that the image belongs to a specific rice variety. The probability prediction scores of the Jasmine and Karacadag rice images are presented in Figs. 10 and 11 respectively. In these figures, the x-axis denotes the probability score, ranging from 0 to 1, while the y-axis represents the type of DL model. Each line on the graph corresponds to a different model and its prediction for the given rice variety. A close inspection of the charts reveals that all models demonstrate high confidence in their predictions, providing probability scores that are proximate to 1.0 for the correct class of rice species. This is consistent for both the Jasmine and Karacadag varieties, which validates the robustness of the models across different rice classes.

In an effort to highlight instances of incorrect class prediction by the models, we have assembled a selection of misclassification examples encountered during the training of the models. Figs. 12 and 13 were constructed to illustrate some of the misjudgments made by the models. Each figure comprises three rows dedicated to different types of cases for each model: the first row illustrates true positive cases, the second presents false positive cases, and the third exhibits false negative cases. Taking for example, VGG16 and VGG19, it can be observed that these models do not manifest any false negative errors for the Ipsala rice type, as validated by the confusion matrix depicted in Fig. 9. Nevertheless, it must be noted that other models did demonstrate at least one error when evaluated against the test dataset.

## 6. Discussion and conclusion

In the present research, we explore the robust capabilities of advanced deep learning models, namely VGG16, VGG19, ResNet50, ResNet101, EfficientNet, and MobileNet, for their classification potential across five distinctive rice varieties. This exploration is facilitated through experiments conducted on a comprehensive dataset, encompassing 75,000 images of rice grains. To underscore the superiority of DL models, we draw a comparative analysis with preceding studies that have relied on traditional machine learning models or novel techniques in machine learning and deep learning like transfer learning. The first work selected for comparison is Cinar et al. [11]. The researchers utilized a relatively smaller subset from the same dataset for feature extraction. This served as a basis for the development of models using traditional machine learning techniques including LR, MLP, SVM, DT, RF, NB, and k-NN. Their highest-performing model was based on Logistic Regression, which achieved an accuracy of 93.02%, a figure that pales compared to the performance exhibited by our deep learning models.

The results garnered from our study significantly underscore the efficacy of the deployed deep learning models. These models, by virtue of processing the entire dataset and eliminating the necessity for feature engineering, demonstrate notable advantages over their traditional counterparts. These attributes can be attributed to their superior performance metrics. The empirical evidence gleaned from our study reinforces the idea that deep learning models provide highly effective solutions for classification tasks, particularly in sectors such as agricultural image analysis, where datasets are often large and intricate.

The second comparison is drawn from a study that focused on classification tasks using the same rice dataset [20]. The methodology outlined in the aforementioned study revolved around the principles of transfer learning, albeit with a distinctive twist. Traditionally, transfer learning, which capitalizes on the pre-existing knowledge of models trained on extensive datasets such as ImageNet, offers a promising avenue for achieving computational efficiency and superior performance on domain-specific tasks. However, the methodology under comparison diverged from this norm, employing the AlexNet architecture, trained afresh on the rice dataset, and subsequently extracting features from its fully connected (FC)layers (FC6, FC7, FC8) for Support Vector Machine (SVM)-based classification.

In replicating this approach, our results revealed accuracies of 96.99% for the AlexNet, and between 96.88% and 96.94% for the feature extraction scenarios (FC6, FC7, FC8). A review of the per-class accuracies further underscores the consistency in performance across the rice varieties. For instance, the AlexNet exhibited accuracies of 96.17% for Arborio, 97.30% for Basmati, 95.25% for Jasmine, 98.74% for Ipsala, and 97.49% for Karacadag. Similar figures, with slight variations, were observed for the feature extraction scenarios. The details of the obtained result are mentioned in Table 3. While these figures are commendable, they fall short when juxtaposed against the performance metrics achieved by our deep learning models, with the best model boasting an accuracy of 99.87. Furthermore, the methodological choice of training AlexNet from scratch and then harnessing its layers for feature extraction, rather than capitalizing on a pre-trained model, raises questions on the optimal utilization of transfer learning's potential.

However, it is crucial to recognize the limitations inherent in this study. Despite our promising outcomes, the results were derived under specific conditions. These conditions encompass the employment of a set learning rate and certain architectural decisions for the deep learning models. Our experiments utilized a single dataset, and although it was comprehensive, results might differ with data from different sources or with varied image quality. Additionally, the performance of the models might be sensitive to hyper-parameters, and the ones chosen for this study, although effective, might not be universally optimal. Variations in performance might also arise when different pre-processing methods or data augmentation techniques are introduced. It's also worth noting that while our models performed admirably, they haven't been tested against adversarial attacks or evaluated for robustness in scenarios with noisy data.

Looking forward, our research aspirations encompass the expansion of our methodologies to cater to other agricultural commodities and more intricate classification challenges. Although we've touched upon the capabilities of transfer learning in this study, other facets, such as domain adaptation or semi-supervised learning, might provide avenues for further enhancements in model performance. In essence, this investigation underscores the profound potential of deep learning models within agricultural image analysis, setting the stage for more sophisticated and efficacious endeavors.

## Declaration of competing interest

This piece of the submission is being sent via mail.

## Data availability

The authors do not have permission to share data.

## Acknowledgment

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

## References

[1] World Food and Agriculture – Statistical Yearbook 2022, Food and Agriculture Organization (FAO), 2022, https://doi.org/10.4060/cc2211en.
[2] N.K. Fukagawa, L.H. Ziska, Rice: importance for global nutrition, J. Nutr. Sci. Vitaminol. 65 (Supplement) (2019) S2–S3, https://doi.org/10.3177/jnsv.65.S2.
[3] G.S. Khush, Origin, dispersal, cultivation and variation of rice, Plant Mol. Biol. 35 (1–2) (1997) 25–34.
[4] M.T. Habib, D.M. Raza, M.M. Islam, D.B. Victor, M.A.I. Arif, Applications of computer vision and machine learning in agriculture: a state-of-the-art glimpse, in: 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2022, pp. 1–5, https://doi.org/10.1109/ICITIIT54346.2022.9744150.
[5] S.V. Mahadevkar, et al., A review on machine learning styles in computer vision—techniques and future directions, IEEE Access 10 (2022) 107293–107329, https://doi.org/10.1109/ACCESS.2022.3209825.
[6] L.F.P. Oliveira, A.P. Moreira, M.F. Silva, Advances in agriculture robotics: a state-of-the-art review and challenges ahead, Robotics 10 (2) (2021) 52, https://doi.org/10.3390/robotics10020052.
[7] M. Koklu, I. Cinar, Y.S. Taspinar, Classification of rice varieties with deep learning methods, Comput. Electron. Agric. 187 (2021 Aug 1), 106285.
[8] Trivedi, Kumar Naresh, et al., Identification of rice varieties using machine learning algorithms, in: 2022 International Conference on Data Analytics for Business and Industry (ICDABI), 2022, pp. 266–271.
[9] C. Sun, T. Liu, C. Ji, M. Jiang, T. Tian, D. Guo, L. Wang, Y. Chen, X. Liang, Evaluation and analysis the chalkiness of connected rice kernels based on image processing technology and support vector machine, J. Cereal. Sci. 60 (2) (2014 Sep 1) 426–432.
[10] P. Lin, X.L. Li, Y.M. Chen, Y. He, A deep convolutional neural network architecture for boosting image discrimination accuracy of rice species, Food Bioprocess Technol. 11 (2018 Apr) 765–773.
[11] I. Cinar, M. Koklu, Classification of rice varieties using artificial intelligence methods, International Journal of Intelligent Systems and Applications in Engineering 7 (3) (2019 Sep 30) 188–194.
[12] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2015, https://doi.org/10.48550/ARXIV.1512.03385.
[13] S. Basodi, C. Ji, H. Zhang, Y. Pan, Gradient amplification: an efficient way to train deep neural networks, Big Data Mining and Analytics 3 (2020) 196–207, https://doi.org/10.26599/BDMA.2020.9020004.
[14] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 Sep 4 arXiv preprint arXiv:1409.1556.
[15] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017, https://doi.org/10.48550/ARXIV.1704.04861.
[16] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, 2016, https://doi.org/10.48550/ARXIV.1610.02357.
[17] M. Tan, Q.V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2019, https://doi.org/10.48550/ARXIV.1905.11946.
[18] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, International conference on machine learning (2019 May 24) 6105–6114. PMLR.
[19] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2014 Dec 22 arXiv preprint arXiv:1412.6980.
[20] N.K. Naik, P.K. Sethy, A.G. Devi, S.K. Behera, Hybrid enhanced featured AlexNet for milled rice grain identification, Ingénierie Des. Systèmes Inf. 28 (3) (2023) 663–668, https://doi.org/10.18280/isi.280315.