Classification of Rice Using Deep Learning

TEAM ID: LTVIP2025TMID33007

Team Members: Akula Surekha,Kantha Jyothsna,Kummari Charitha,Seelam Subba Sahithi,Segireddy Babitha

Santhiram Engineering College(Autonomous),Nandyal,
Andhra Pradesh, India

***Abstract**: Rice is one of the most important staple crops in the world and serves as a staple food for more than half of the global population. It is a critical source of nutrition, providing carbohydrates, vitamins, and minerals to millions of people, particularly in Asia and Africa. This paper presents a study on using deep learning for the classification of different types of rice. The study focuses on five specific types of rice: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. A dataset of 75,000 images was collected and annotated, consisting of 15,000 images for each of the five types of rice. The deep learning system uses convolutional neural networks (CNNs) to classify images of rice grains based on their physical characteristics. The performance of the deep learning system was evaluated using metrics such as accuracy, precision, recall, and F1-score. The results show that the proposed deep learning system was able to achieve high accuracy in classifying the different types of rice, outperforming traditional image processing techniques. This study provides valuable insights into the potential of deep learning for classifying different types of rice and can be useful for applications such as quality control in rice processing industries and research in crop science. The large size of the dataset, 15,000 images for each type of rice, helped the deep learning system to learn the features of each type of rice, and increase the performance of the classifier. The attained overall accuracy of the proposed model (99.96%), Precision (99.96%), Recall (99.96%) and F1-score (99.96%). Thus, the proposed model proved that it leaned the five categories of rice and it can generalize these categories and it is effective.*

Keywords: Classification, Type of Rice, Deep Learning, CNN, ResNet50V2

1. INTRODUCTION

Rice classification is a crucial task that has a wide range of applications, including quality control in rice processing industries, research in crop science, and selection of appropriate rice varieties for different culinary uses. Traditional methods of rice classification, such as manual visual inspection, are "time-consuming and prone to human error" [1]-[3].

In recent years, deep learning has emerged as a powerful tool for image classification tasks and has been applied in various domains, including agriculture. In this study, we propose a deep learning system for classifying five different types of rice: Arborio, Basmati, Ipsala, Jasmine, and Karacadag.

The system utilizes convolutional neural networks (CNNs) to classify images of rice grains based on their physical characteristics. A dataset of 75,000 images was collected and annotated, consisting of 15,000 images for each of the five types of rice. This large dataset allows the deep learning system to learn the features of each type of rice, and increase the performance of the classifier.

The performance of the proposed system was assessed using metrics like accuracy, precision, recall, and F1-score. Furthermore, this study also explores the impact of different pre-processing techniques and CNN architectures on the classification performance.

2. BACKGROUND

2.1. Deep Learning

Deep learning is a subfield of machine learning that uses artificial neural networks (ANNs) with multiple layers to process and analyze large sets of complex data, such as images, audio, and text. The field of deep learning has its roots in the study of artificial neural networks, which began in the 1940s and 1950s. In the 1980s and 1990s, researchers began to develop more complex neural network architectures, such as "multi-layer perceptron (MLPs)" and "radial basis function networks (RBFNs)." However, it was not until the advent of powerful computational resources and large datasets in the early 2000s that deep learning truly began to take off [4].

The term "deep learning" was first coined by Rina Dechter in 1986, but it was not widely used until the early 2000s, when a resurgence of interest in neural networks and the availability of large datasets and computational resources led to the development of deep learning models. In 2006, a team led by Geoffrey Hinton at the University of Toronto developed a deep learning model called a deep belief network (DBN), which was able to achieve "state-of-the-art performance on several benchmark datasets [7]."

Since then, deep learning has been applied to a wide range of tasks, including image recognition, natural language processing, speech recognition, and game playing. The field has also seen the development of many different architectures, such as CNN and "recurrent neural networks (RNNs)," as well as new techniques for training deep networks, such as unsupervised pre-training and dropout [8].

Deep Learning has quickly become one of the most active and exciting areas of research in machine learning and AI, with numerous papers being published every year in top-tier conferences and journals, and many companies and startups investing in the development of deep learning technologies [9].

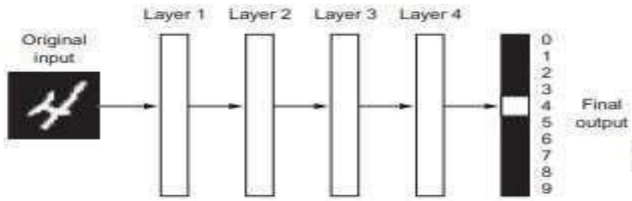

Fig. 1. A deep neural network for classification [2]

As evidenced by Figure 2, the deep neural network effectively transforms the digit image into representations that are progressively dissimilar from the original image, while concurrently becoming increasingly informative in regards to the final classification result. This process can be conceptualized as a multi-stage information distillation operation, in which the information passes through successive filters, ultimately resulting in increasingly purified representations that are highly relevant to the specific task at hand [3].

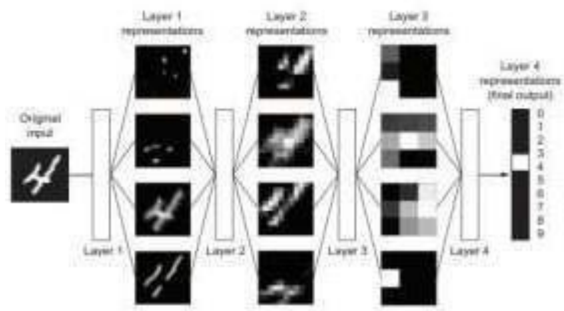

Fig. 2. Deep representations learned by a digit-classification model [4]

Deep learning is a multi-stage approach to learning data representations, which can appear like magic due to the use of simple mechanisms at a large scale [5].

2.2. Convolutional neural network (CNN):

A (CNN) is a specialized type of “deep learning neural network” that is particularly well-suited for “processing structured arrays of data,” such as images. The widespread utilization of CNNs in computer vision is a testament to their effectiveness, with the architecture having become the state-of-the-art for a vast array of visual applications, including image classification. Additionally, CNNs have also demonstrated success in natural language processing, particularly in the realm of text classification [6]-[8].

A CNN is an advanced “deep learning neural network” that is specifically designed for image processing. It is particularly good at detecting patterns in an input image, such as circles, lines, faces, eyes and gradients. This feature can makes CNN highly effective in the applications of computer vision. Not like traditional computer vision algorithms, CNN can operate straight on an images raw data without the need for pre-processing [9].

CNNs are feed-forward neural networks that consist of multiple layers, with some models having up to 20 or 30 layers. The power of CNNs is due to the use of convolutional layers, which are capable of recognizing increasingly complex shapes when stacked on top of each other. A few convolutional layers can recognize handwritten digits, while 25 layers can distinguish human faces. The structure of the convolutional layers in a CNN mimics the human visual cortex, enabling the network to process images in a way that is similar to human visual processing [10].

2.2.1. CNN Design

A CNN is a “multi-layered feed-forward neural network” made up of sequential “hidden layers.” This design enables CNNs to acquire “hierarchical features. The hidden layers” are often composed of activation layers, convolutional layers, and in some cases pooling layers [11].

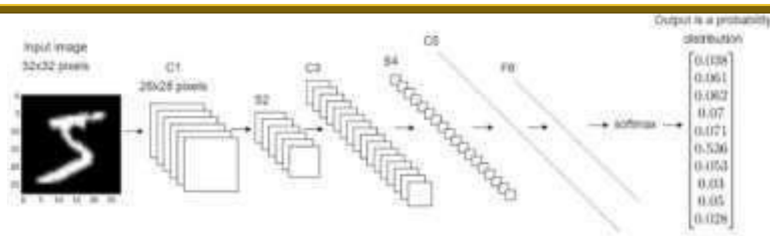

Fig. 3. CNN Design

2.2.2. Architecture of ResNet50V2:

ResNet50V2 is a version of the ResNet (Residual Network) architecture, a “deep convolutional neural network” designed for image classification tasks. The V2 version introduces changes to the original ResNet50 architecture, such as the use of pre-activation units in the residual blocks, as well as batch normalization applied before the activation function. The architecture consists of multiple residual blocks, each containing multiple convolutional layers and identity shortcuts, allowing for deeper networks with reduced training error. The ResNet50V2 model has 50 weight layers and is typically trained on large image datasets such as ImageNet [12].

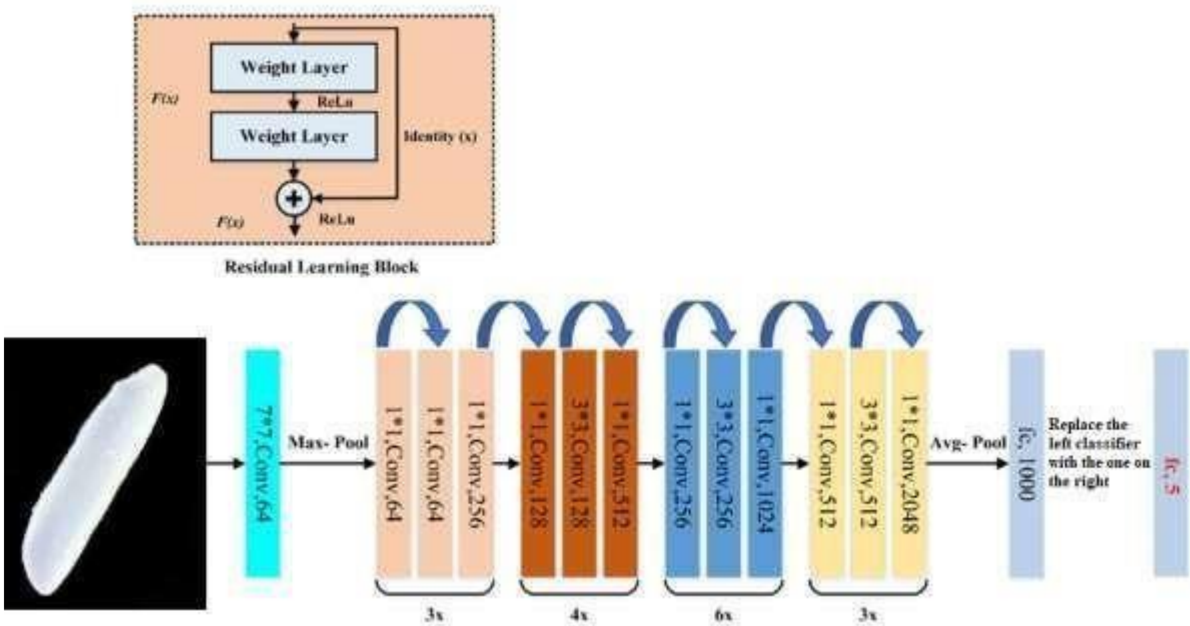

Fig. 4. ResNet50V2 Architecture

The 50-layer ResNetV2employs a “bottleneck design in its building block,” utilizing 1x1 "bottleneck" convolutions to decrease parameters and matrix multiplications, resulting in quicker training per layer. Instead of a two-layer design, it features a stack of three layers [13].

The 50-layer ResNetV2 architecture consists of the followings:

- **“A 7×7 kernel convolution** alongside 64 other kernels with a 2-sized stride”.
- **“A max pooling layer** with a 2-sized stride.”
- **“9 more layers**—3×3,64 kernel convolution, another with 1×1,64 kernels,” and a third with “1×1,256 kernels. These 3 layers are repeated 3 times. “
- **“12 more layers** with 1×1,128 kernels, 3×3,128 kernels, and 1×1,512 kernels, iterated 4 times.”
- **“18 more layers** with 1×1,256 cores, and 2 cores 3×3,256 and 1×1,1024, iterated 6 times.”
- **“9 more layers** with 1×1,512 cores, 3×3,512 cores, and 1×1,2048 cores iterated 3 times.”

After the 50 layers, there is an additional:

- **"Average pooling**, followed by a fully connected layer with 1000 nodes, using the softmax activation function."

2.3. Transfer Learning:

Transfer learning is a machine learning technique where a model trained on one task is used as the starting point for a model on a second, related task. This is done in order to leverage the knowledge gained from solving the first task, which can help the model to learn the second task more quickly and effectively. The idea behind transfer learning is that many features learned by a model on one task can be useful for other tasks, especially if the tasks are related [24].

For example, a model trained to recognize objects in images might be used as the starting point for a model that recognizes scenes in images, because many of the features that are useful for recognizing objects (e.g. edges, textures, etc.) will also be useful for recognizing scenes. Transfer learning can be especially effective when the data for the second task is limited, as it can help the model to avoid overfitting and generalize better to new examples [15].

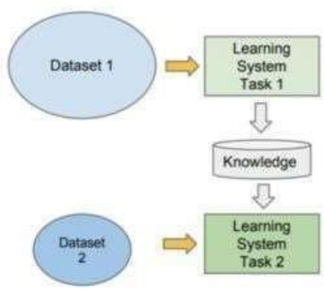

Fig. 5. Transfer learning

2.4. ImageNet:

ImageNet is a large database of labeled images that is widely used for training and evaluating computer vision models. The database was created by researchers at Stanford University and is designed to provide a benchmark for computer vision algorithms. It contains over 14 million images belonging to more than 20,000 different categories, including common objects such as animals, vehicles, and household items [16].

ImageNet is particularly well-known for its annual competition, called the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which evaluates the performance of computer vision models on a set of image classification and object detection tasks. The ILSVRC has been a driving force behind the development of many state-of-the-art computer vision models, including convolutional neural networks (CNNs), and has helped to advance the field of computer vision [17]-[19].

The ImageNet database and the ILSVRC competition have been instrumental in driving the development of computer vision, and many of the techniques and models developed for ImageNet have been applied to a wide range of real-world problems, such as object recognition, face detection, and scene understanding [20].

3. STUDY OBJECTIVES:

- To develop a deep learning system for the classification of five different types of rice: Arborio, Basmati, Ipsala, Jasmine, and Karacadag.
- To assess the performance of the deep learning system for rice classification using a dataset of images of the five different types of rice.
- To investigate the potential of the proposed deep learning approach for applications such as quality control in rice processing industries and research in crop science.
- To provide insights into the optimal CNN architectures and pre-processing techniques for rice classification using deep learning.
- To contribute to the advancement of deep learning techniques in the field of agriculture technology.

Overall, the study aims to demonstrate the potential of deep learning for the classification of different types of rice and provide valuable insights for future research in this area. It also aims to improve the effectiveness and accurateness of rice classification process, and contribute to the field of agriculture technology.

4. LITERATURE REVIEW

In the study of [1] the authors used a data set that consists of 2000 images of two classes of rice (whole rice and broken rice). They used three algorithms of deep and machine learning: Convolutional Neural Network (CNN), k-Nearest Neighbors (KNN) and Support Vector Machines (SVM) method with HOG features. The accuracies attained in CNN, KNN, and SVM are 93.85%, 84.30% and 85.06% respectively.

In the study [2], the authors adopted 6 machine learning algorithms for the Classification of Rice Varieties: Support Vector Machine (SVM), Decision Tree (DT), Linear Regression, Random Forest (RF), and Multilayer Perceptron (MLP). The authors used 3810 rice grain images for the classification task. For the evolution of these algorithm, they used precision, Confusion matrix, F1 score and accuracy. The accuracies of the 6 algorithms attained LR (93.2%), MLP (92.86%), SVM (92.83%), RF (92.39%) and DT (92.49%).

In the study of [3] the authors used a data set that consists of 75000 images of five classes of rice (Arborio, Basmati, Ipsala, Jasmine, and Karacadag). They used Vanilla CNN algorithms for the classification. The accuracy attained is 95.39%.

In the study of [21] the authors used two datasets (processed rice and paddy rice). The processed rice dataset contains 6 labels while the paddy rice dataset contain 4 labels. They used SVM and ResNet-B algorithms for the classification. The accuracy attained are SVM: 94.86%, ResNet-B: 97.54%. For the evaluation, theu used F1-score, Recall, precision and accuracy.

In the study [22], the authors fin-tuned, RiceNet, InceptionV3 and ResNetInceptionV2 for the classification of Rice. The accuracies achieved RiceNet, InceptionV3 and ResNetInceptionV2 models: 94%, 84% and 81.333% respectively.

In this paper [23], several hand-crafted descriptors and Convolutional Neural Networks (CNN) methods were evaluated and compared. The experiment was simulated on the VNRICE dataset with 6 labels (BC-15, Huong Thom-1, Nep-87, Q-5, ThienUu-8, and Xi-23) on which their methods showed significant result. The highest accuracy obtained was 99.04% by using DenNet21 framework.

In paper [24]-[25], they used the same dataset with 75000 images and 5 labels; but in [34] they adopted RF, DT with accuracy RF (99.85%) and DT (99.68%); while in [25] ANN was adopted with accuracy (99.87%.

In papers [26]-[32], they used different datasets with different labels and different models and their accuracies ranged from 86.85% to 99.50%

Table 1 summarizes the previous studies in terms of publishing year, dataset, number of labels, method used and accuracy attained.

Table 1. Summarizes the previous studies

Ref.	Year	Dataset	Classes	Methods used	Accuracy
3	2022	75000	5	Vanilla CNN	CNN: 95.39%
22	2022	4898	5	RiceNet, InceptionV3 and ResNetInceptionV2	RiceNet: 94.00%, InceptionV3: 84.00% and ResNetInceptionV2: 81.33%
23	2022	75000	5	RF, LR, DT, MLP, KNN, SVM	RF: 98.00%, LR: 97.80%, DT: 97.80%, MLP: 97.40%, KNN: 97.20%, SVM: 97.00%
25	2022	75,000	5	RF, DT	RF: 99.85%, DT: 99.68%
26	2021	75,000	5	ANN	ANN: 99.87%
29	2021	200	3	CNN	CNN: 88.07%
24	2020	VNRICE 10448	6	DenNet21	DenNet21: 99.04%
1	2019	2000	2	CNN, SVM, KNN	CNN: 93.85%, SVM: 85.06%, KN: 84.30%
2	2019	3810	-	LR, MLP, RF, SVM, DT	LR : 93.20%, MLP: 92.86%, RF 92.39%, SVM: 92.83%, DT: 92.49%
32	2019	3810	3	DCNN	DCNN: 95.50%
21	2018	processed rice and paddy rice datasets	6, 4	SVM, ResNet-B	SVM: 94.86%, ResNet-B:97.54%
28	2018	7399	3	CNN	CNN: 95.5%
31	2018	300	6	Multiclass-SVM	Multiclass-SVM: 92%

27	2014	17,00	2	SVM	SVM: 98.5%
30	2010	750	5	BP-ANN	BP-ANN: 86.85%

5. METHODOLOGY

- Data collection: Images of the five different types of rice (Arborio, Basmati, Ipsala, Jasmine, and Karacadag) were collected and annotated for use as a dataset for training and testing the deep learning system.
- Pre-processing: The images were pre-processed using image enhancement techniques such as histogram equalization and image segmentation methods such as thresholding.
- Deep learning model: The deep learning system was developed using CNNs. Different CNN architectures, such as ResNet, were implemented and evaluated for their performance.
- Model training and evaluation: The deep learning model was trained on the pre-processed dataset and evaluated on a separate test dataset. The classification performance was measured using metrics such as accuracy, precision, recall, and F1-score.
- Impact of pre-processing and CNN architectures: The impact of different pre-processing techniques and CNN architectures on the classification performance was investigated by comparing the results obtained with the different methods.
- Application and potential: The potential of the deep learning system for applications such as quality control in rice processing industries and research in crop science was also explored.

Overall, this study used a combination of image pre-processing, deep learning and performance evaluation methods to classify five different types of rice using CNNs. This methodology allowed to provide valuable insights and understanding of the performance of the deep learning system for rice classification and its potential for various applications.

4.1. Dataset:

For classifying Rice variety, a comprehensive dataset of 75,000 images was amassed. These images were obtained from the Kaggle website and were utilized to construct a convolutional neural network (CNN). To ensure consistency in image size, each picture was cropped to 64x64 pixels. Figure 6 illustrates the different variety of rice images utilized in the current study. To ensure a robust training process, a 70-15-15 split was employed, with 70% of the images allocated for training, 15% for validation, and 15% for testing. The proposed model was trained using the training dataset, and subsequently assessed using the testing dataset set.

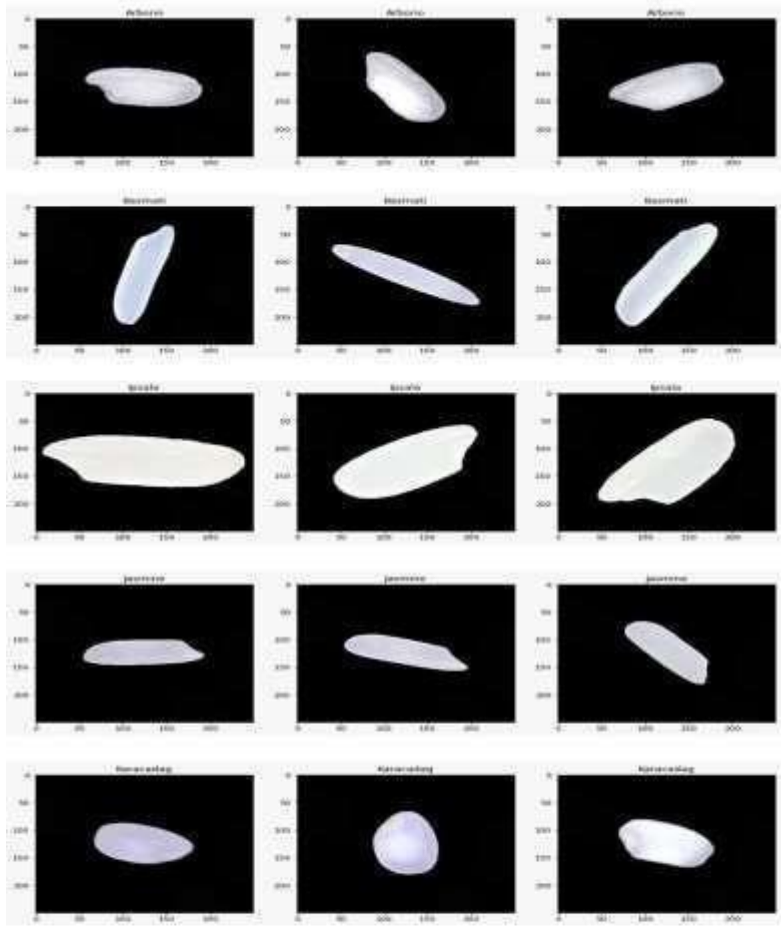

Fig. 6. Samples of the Rice Dataset

4.2. Evaluation:

For the assessment of the suggested ResNetV2 model using the Rice Dataset, we used Accuracy as in Eq. 4, Precision (Eq. 1), Recall (Eq. 2) and F1-Score (Eq. 3).

$$\text{"Precision} = \frac{\text{TP}}{\text{TP + FP}} \text{"} \tag{1}$$

$$\text{"Recall} = \frac{\text{TP}}{\text{TP + FN}} \text{"} \tag{2}$$

$$\text{"}F1 - score = 2 * \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \text{"} \tag{3}$$

$$\text{"Accuracy} = \frac{\text{TN + TP}}{\text{TN+FP+ TP + FN}} \text{"} \tag{4}$$

Where: FP=False Positive; "FN=False Negative; TP =True Positive; TN=True Negative "

4.3. Validation Method:

To evaluate our model, it was necessary to divide the available dataset into training, validation and test sets. Therefore, an experiment was conducted to determine the optimal ratio of splitting the dataset. The experiment involved splitting with a specific ration, training and evaluating the model for a few different ratio, each time selecting a different ratio of splitting the dataset. The results revealed a significant variation in accuracy and loss, but the finally we adopted the 70x15x15 ratio leading to the conclusion as illustrated in Figure 7, the training and validation accuracy history is plotted for the last 20 epochs while Figure 8 illustrates the training and validation loss for the same last 20 epochs.

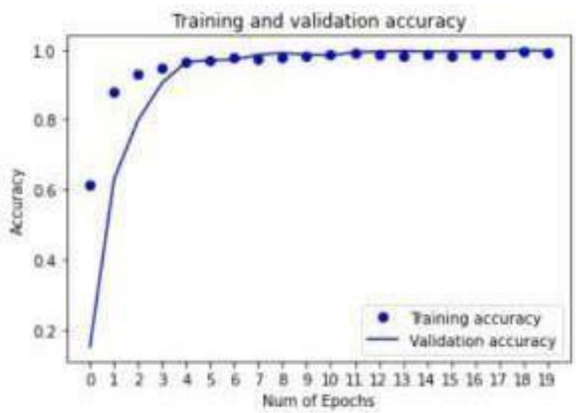

Fig. 7. Training and validation accuracy of proposed Model

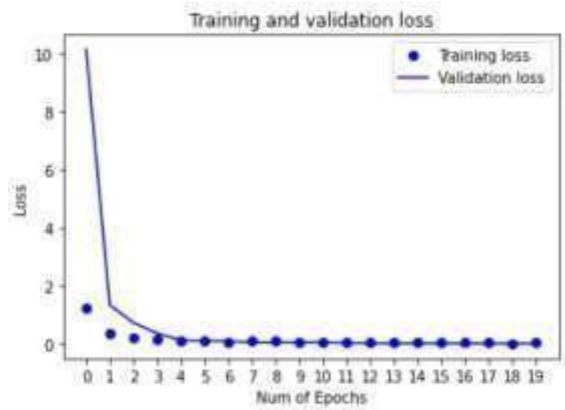

Fig. 8. Training and validation Loss of proposed Model

4.4. Loss and Accuracy rate:

Figure 9 shows the time taken in training and validation of the proposed model, the accuracy of training and validation, the loss of training and validation for the last 10 epochs. The accuracy of training (99.81%), accuracy of validation (99.90%), loss of training (0.0077), validation loss (0.0045) and time required for training and validation (272 seconds).

```
Epoch 11/20
147/147 [==============================] - 13s 90ms/step - loss: 0.0055 - accuracy: 0.9986 - val_loss: 0.0136 - val_accuracy: 0.9963
Epoch 12/20
147/147 [==============================] - 13s 89ms/step - loss: 0.0109 - accuracy: 0.9969 - val_loss: 0.0110 - val_accuracy: 0.9957
Epoch 13/20
147/147 [==============================] - 18s 123ms/step - loss: 0.0101 - accuracy: 0.9976 - val_loss: 0.0040 - val_accuracy: 0.9990
Epoch 14/20
147/147 [==============================] - 13s 85ms/step - loss: 0.0066 - accuracy: 0.9981 - val_loss: 0.0068 - val_accuracy: 0.9980
Epoch 15/20
147/147 [==============================] - 13s 86ms/step - loss: 0.0086 - accuracy: 0.9976 - val_loss: 0.0053 - val_accuracy: 0.9987
Epoch 16/20
147/147 [==============================] - 12s 85ms/step - loss: 0.0172 - accuracy: 0.9952 - val_loss: 0.0473 - val_accuracy: 0.9897
Epoch 17/20
147/147 [==============================] - 13s 86ms/step - loss: 0.0123 - accuracy: 0.9958 - val_loss: 0.0071 - val_accuracy: 0.9977
Epoch 18/20
147/147 [==============================] - 12s 84ms/step - loss: 0.0045 - accuracy: 0.9987 - val_loss: 0.0064 - val_accuracy: 0.9980
Epoch 19/20
147/147 [==============================] - 12s 84ms/step - loss: 0.0170 - accuracy: 0.9964 - val_loss: 0.0094 - val_accuracy: 0.9977
Epoch 20/20
147/147 [==============================] - 13s 86ms/step - loss: 0.0077 - accuracy: 0.9981 - val_loss: 0.0045 - val_accuracy: 0.9990
CPU times: user 3min 39s, sys: 6.67 s, total: 3min 46s
Wall time: 4min 32s
```

Fig. 9. The last 10 epochs of the training and validating the proposed model

4.5. Confusion Matrix:

In the current study, the proposed CNN ResNet50V2 model for the classification of Rice variety using the rice dataset was tested using the test dataset. Figure 10 illustrates good performance in the confusion matrix.

```
[508   0   0   0   0]
[  0 547   0   1   0]
[  0   0 542   0   0]
[  0   0   0 510   0]
[  0   0   0   0 532]
```

Fig. 10: Confusion matrix

6. RESULTS AND DISCUSSION

To evaluate the efficiency of the proposed ResNet50V2 model, we got the precision, recall, F1-score and supporting number of images for each class of the Rice dataset as shown in Figure 11. Furthermore, we got the overall accuracy, Precision, Recall, F1-score of the proposed mode. The overall F1-score (99.96%), accuracy attained (99.96%), Recall (99.96%) and Precision (99.96%). Thus, the proposed model proved that it leaned the five categories of rice and it can generalize these categories and it is effective.

	precision	recall	f1-score	support
Arborio	1.0000	1.0000	1.0000	508
Basmati	1.0000	0.9982	0.9991	548
Ipsala	1.0000	1.0000	1.0000	542
Jasmine	0.9980	1.0000	0.9990	510
Karacadag	1.0000	1.0000	1.0000	532
accuracy			0.9996	2640
macro avg	0.9996	0.9996	0.9996	2640
weighted avg	0.9996	0.9996	0.9996	2640

Fig. 11. Classification Report of the proposed model

The fine-tuned ResNet50V2 outperform the models in the previous studies. Table 2 summarize the comparisons of the previous studies with the current study.

Table 2. A comparisons of the previous studies with the current study.

Ref.	Year	Dataset	Classes	Methods used	Accuracy
3	2022	75000	5	Vanilla CNN	CNN: 95.39%
22	2022	4898	5	RiceNet, InceptionV3 and ResNetInceptionV2	RiceNet: 94.00%, InceptionV3: 84.00% and ResNetInceptionV2: 81.33%
23	2022	75000	5	RF, LR, DT, MLP, KNN, SVM	RF: 98.00%, LR: 97.80%, DT: 97.80%, MLP: 97.40%, KNN: 97.20%, SVM: 97.00%
25	2022	75,000	5	RF, DT	RF: 99.85%, DT: 99.68%
26	2021	75,000	5	ANN	ANN: 99.87%
29	2021	200	3	CNN	CNN: 88.07%
24	2020	VNRICE 10448	6	DenNet21	DenNet21: 99.04%
1	2019	2000	2	CNN, SVM, KNN	CNN: 93.85%, SVM: 85.06%, KN: 84.30%
2	2019	3810	-	LR, MLP, RF, SVM, DT	LR : 93.20%, MLP: 92.86%, RF 92.39%, SVM: 92.83%, DT: 92.49%
32	2019	3810	3	DCNN	DCNN: 95.50%
21	2018	processed rice and paddy rice datasets	6, 4	SVM, ResNet-B	SVM: 94.86%, ResNet-B:97.54%
28	2018	7399	3	CNN	CNN: 95.5%
31	2018	300	6	Multiclass-SVM	Multiclass-SVM: 92%

27	2014	17,00	2	SVM	SVM: 98.5%
30	2010	750	5	BP-ANN	BP-ANN: 86.85%
Current study	**2023**	**75000**	**5**	**ResNetV2**	**ResNetV2: 99.96%**

7. CONCLUSION

Rice is considered one of the most vital staple crops worldwide and used as a primary food for more than half of the global inhabitants. It is a critical source of nutrition, providing carbohydrates, vitamins, and minerals to millions of people, particularly in Asia and Africa. The purpose of current study is the proposition an effective CNN model for the classification of Rice variety. ResNet50V2 model was fine-tuned of this task. The model was trained, validated and tested using the rice dataset that was collected for Kaggle depository.

The attained overall accuracy of the proposed model (99.96%), Precision (99.96%), Recall (99.96%) and F1-score (99.96%). Thus, the proposed model proved that it leaned the five categories of rice and it can generalize these categories and it is effective.