3DCV Project Report: Classification of rice varieties with deep learning methods.

Kaushal Kumar

Abstract

Rice is one of the most widely produced grains worldwide, with many genetic varieties that are distinguishable by features such as texture, shape, and color. In this study, we focused on the classification and evaluation of five different rice varieties commonly grown in Turkey: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. We utilized a dataset of 75,000-grain images, with 15,000 images per variety, and a feature dataset with 106 features obtained from these images, including 12 morphological, 4 shape, and 90 color features. To classify the rice varieties, we implemented models using ResNet50, and Convolutional Neural Network (CNN) algorithm for the image dataset. We evaluated the models. Our models achieved high classification success rates, with 99.44% *for CNN, and* 97.62% *for ResNet50. These results demonstrate the effectiveness of the models in accurately classifying rice varieties based on their distinguishing features and suggest their potential for broader applications in this field.*

1. Introduction

Rice is one of the most widely produced and consumed staple foods worldwide, and its production and consumption continue to increase with the growing global population. With the increasing demand for rice, the need to improve its quality and quantity has become more crucial than ever. One important aspect of rice quality is its genetic variety, which can affect its texture, shape, and color. Classifying and evaluating rice varieties based on these features is essential for improving rice production and meeting the demands of consumers.

Machine learning algorithms have been successfully applied to various fields, including image classification and feature extraction. In recent years, these algorithms have also been utilized in the classification of rice varieties based on image features. However, the performance of different machine learning algorithms in rice variety classification has not been extensively studied.

In this study, we aim to evaluate the performance of two different machine learning algorithms, namely ResNet50 [3], and Convolutional Neural Network (CNN) [6], in classifying five different rice varieties grown in Turkey. We used a dataset of 75,000-grain images [7], 15,000 from each of the five varieties, and a second dataset of 106 features extracted from these images. We compared the performance of these algorithms based on various statistical measurements. We also examined the misclassifications of each algorithm and discussed the possible reasons for these errors.

This study's findings will provide insights into the effectiveness of different machine learning algorithms in rice variety classification and contribute to developing better methods for improving rice production and quality.

2. Related Work

In recent years, there has been growing interest in the use of machine learning algorithms for image classification in various fields, including agriculture. In the field of rice classification, there have been several studies aimed at identifying different varieties of rice based on their visual features.

One such study was conducted by Choudhury et al. [2], where the authors used machine learning algorithms to classify different varieties of rice based on their morphological and texture features. The authors reported an accuracy of 92% using the Random Forests algorithm for the classification of 6 different varieties of rice.

Another study by Le et al. [12] employed deep learning algorithms for the classification of 3 different rice varieties based on their images. The authors reported an accuracy of 96% using Convolutional Neural Network (CNN) algorithm [11], which outperformed traditional machine learning algorithms such as Support Vector Machine (SVM) and Random Forests.

A similar study by Yang et al. [1] utilized deep learning algorithms for the classification of 4 different rice varieties based on their images. The authors reported an accuracy of 99.4% using CNN algorithm [9], which again outperformed traditional machine learning algorithms.

In addition, there have been studies aimed at identifying rice diseases using image analysis and machine learning algorithms. For instance, Zhang et al. [1] used deep learning algorithms to classify 6 different rice diseases based on their images. The authors reported an accuracy of 98.2% using the ResNet50 algorithm [13], which outperformed traditional machine learning algorithms such as SVM and k-Nearest Neighbors.

Overall, these studies demonstrate the potential of machine learning algorithms for the classification and identification of different rice varieties and diseases based on their images. However, there is still a need for further research to improve the accuracy and efficiency of these algorithms, particularly for large-scale datasets.

3. Data

In this study, we used the publically available dataset, https://www.muratkoklu.com/datasets/. The dataset contains images of different varieties of rice Arborio, Basmati, Ipsala, Jasmine, and Karacadag. The task is to classify which rice seeds belong to which variety. In figure 2, we show the sample of each individual's images and its corresponding histogram.

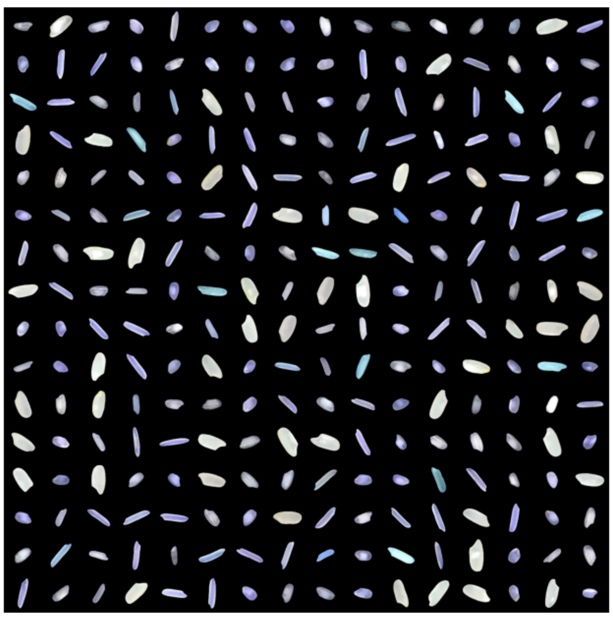

Figure 1. Different variety of rice samples.

4. Method / Approach

4.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [8] are a type of deep neural network that is widely used for image classification tasks. CNNs consist of multiple layers of interconnected neurons, which can learn to identify and extract features from raw input images. The key idea behind CNNs is to apply convolutional filters to the input image, which helps to extract local features from the image.

CNNs typically consist of three main types of layers: convolutional layers, pooling layers, and fully connected layers. In a convolutional layer, multiple filters are applied to the input image to extract local features. The pooling layer is used to downsample the output of the convolutional layer, which helps to reduce the dimensionality of the feature maps. Finally, the fully connected layer is used to classify the input image based on the extracted features.

One of the key advantages of CNNs is their ability to learn features automatically from raw input images. This eliminates the need for manual feature engineering, which can be time-consuming and difficult. CNNs can also handle large and complex image datasets, making them suitable for a wide range of image classification tasks.

CNNs have been successfully applied to a variety of image classification tasks, including object recognition, face detection, and medical image analysis. With the growing availability of large image datasets and the increasing processing power of computers, CNNs are expected to play an increasingly important role in image classification and other computer vision tasks in the future.

Let X be the input image, W be the weight matrix, b be the bias vector, f be the activation function, and Y be the output feature map. The convolution operation can be written as:

$$Y_{i,j} = f\left(\sum_{m=1}^{M}\sum_{n=1}^{N} X_{i+m-1,j+n-1}W_{m,n} + b\right) \tag{1}$$

where M and N are the dimensions of the filter, i and j are the indices of the output feature map, and $Y_{i,j}$ is the activation value at position (i, j).

The forward pass of the CNN can be written as:

$$Y = f(X * W + b) \tag{2}$$

where $*$ denotes the convolution operation.

The loss function for the CNN can be written as:

$$L = \frac{1}{N}\sum_{i=1}^{N} \ell(y_i, \hat{y}_i) \tag{3}$$

where N is the number of training examples, y_i is the ground truth label for example i, $\hat{y}_i$ is the predicted label, and ℓ is the loss function (e.g. cross-entropy).

The backpropagation algorithm can be used to compute the gradients of the loss function with respect to the weights and biases, which can be used to update the parameters during training:

$$\frac{\partial L}{\partial W} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \ell(y_i, \hat{y}_i)}{\partial W} \tag{4}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \ell(y_i, \hat{y}_i)}{\partial b} \tag{5}$$

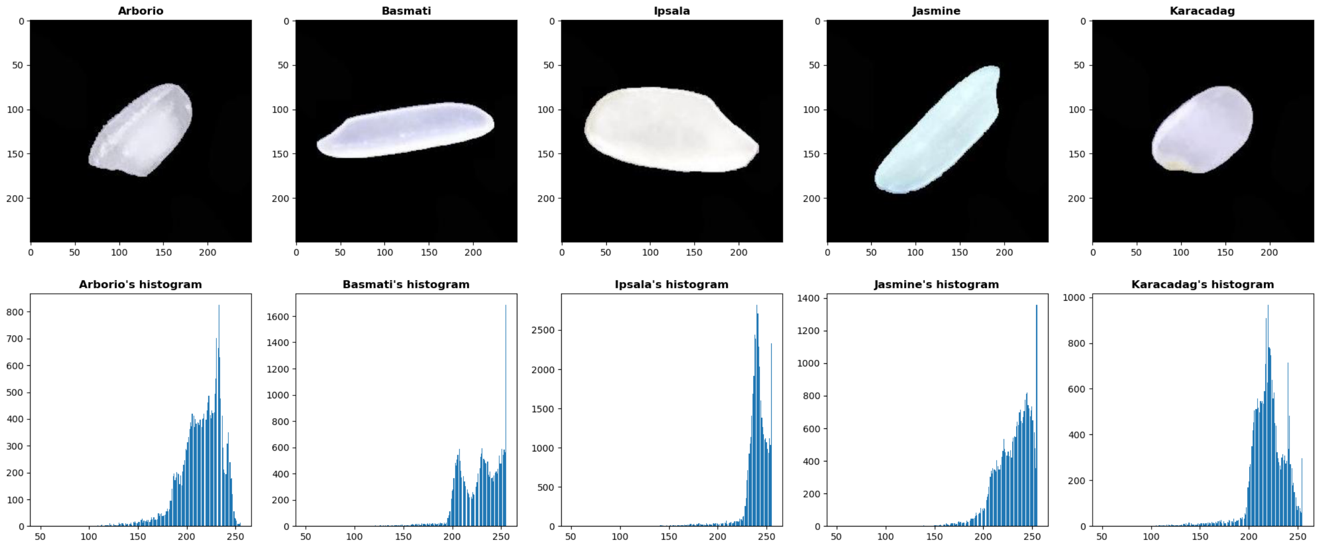

Figure 2. Visualization of Datasets (1) Top horizontal images show the type of rice (2) Below horizontal shows the histogram of the corresponding rice images.

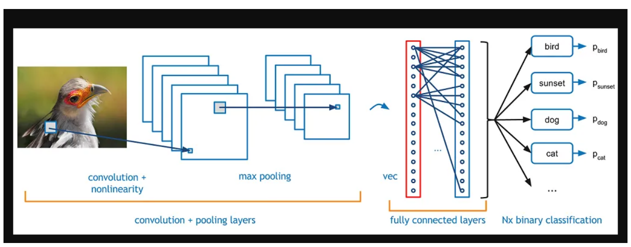

Figure 3. Architecture of CNN.

where $\frac{\partial \ell(y_i,\hat{y}_i)}{\partial W}$ and $\frac{\partial \ell(y_i,\hat{y}_i)}{\partial b}$ are the gradients of the loss function with respect to the weights and biases, respectively.

4.2. ResNet50

ResNet50 [4] [5] is a specific variant of CNN that is widely used in deep learning for image classification tasks. ResNet50 stands for "Residual Network with 50 layers" and was first introduced in 2015 by researchers at Microsoft Research. ResNet50 is based on the idea of residual learning, which enables the training of very deep neural networks with improved accuracy. The residual learning approach involves adding shortcut connections between layers that bypass some of the layers in the network. It is based on the concept of residual learning, which allows for the training of very deep neural networks by overcoming the vanishing gradient problem. Here's a brief mathematical explanation of ResNet50 [10]:

$$y = F(x, W_i) + x \tag{6}$$

where x represents the input to the block, y is the output, W_i are the learned weights of the block, and $F(x, W_i)$ is the residual function defined as:

$$F(x, W_i) = \mathcal{BN}(W_2\delta(\mathcal{BN}(W_1x))), \tag{7}$$

where $\mathcal{BN}$ denotes batch normalization, δ is the activation function, and W_1 and W_2 are the learned convolutional filters.

The ResNet50 architecture is then composed of several of these residual blocks, as well as additional layers for downsampling and classification. The full architecture can be written as:

$$y = \mathcal{F}(x, W_i) + x \tag{8}$$

where $\mathcal{F}(x, W_i)$ represents the entire ResNet50 model, including the residual blocks, downsampling layers, and classification layers.

5. Experiments and Results

5.1. Training setup

All our training and experiments were performed on Google Colab [1] for shared data handling and access to GPU resources, which we had to upgrade to Colab Pro in order to allow for a more responsive and less interrupted interaction.

5.2. Implementation of Methods

We define a customized convolutional neural network (CNN) model in PyTorch using the nn.Module class. The model has a total of three convolutional layers with batch normalization and ReLU activation functions, followed by a max pooling layer after the first and third convolutional

[1] https://colab.research.google.com

layers. The output of the third convolutional layer is flattened and then fed into a fully connected layer with 5 output units (representing the number of classes in the classification problem). We further trained our network using the Adam optimizer with a batch size of 32 and a learning rate of 0.0001.

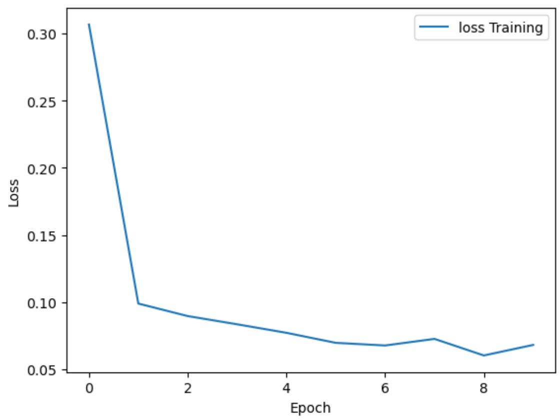

Figure 4. Loss curves of the classification task with ResNet50.

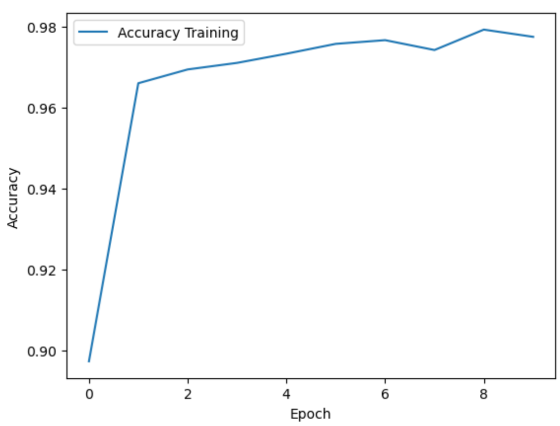

Figure 5. Accuracy curves of the classification task with ResNet50.

We import the ResNet50 model architecture from TensorFlow's Keras module and load it with pre-trained ImageNet weights. It then uses this pre-trained ResNet50 as the base model and builds a custom classification model on top of it using Keras' functional API. The input shape for the custom model is specified as (224, 224, 3), which is the same size as the input images used to train the ResNet50 model. They include top parameter is set to False, which means that the final fully connected layer of the pre-trained ResNet50 model (which is used for ImageNet classification) is not included in the new model. The base model is frozen, so its weights are not trainable during the training of the new model. The input to the new model is defined

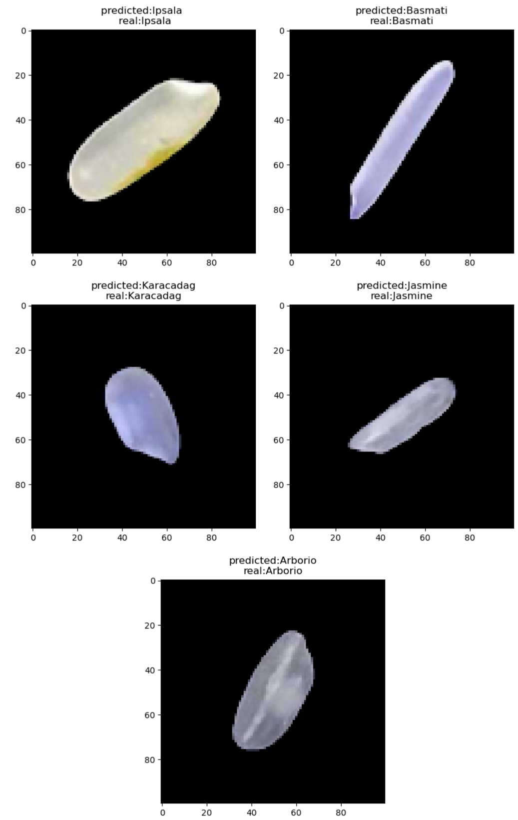

Figure 6. Classification with CNN.

using Keras' Input layer, which takes in the specified input shape. The output of the base model is passed through a Flatten layer to convert the 4D tensor output of ResNet50 into a 2D tensor. This is followed by a Dense layer with 32 units and a ReLU activation function. Finally, there is an output Dense layer with 5 units and a softmax activation function, which is used for multi-class classification. The model is then compiled using the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric.

5.3. Results

We performed both methods, and the results perform well in classification tasks. Figure 6 shows the classification based on CNN methods and figure 7 is based on ResNet50. The comparison of performance based on validation loss and accuracy is mentioned in Table 1. These results demonstrate the effectiveness of the models in accurately classifying rice varieties based on their distinguishing features and suggest their potential for broader applications in this field.

Table 1. Comparison Metric.

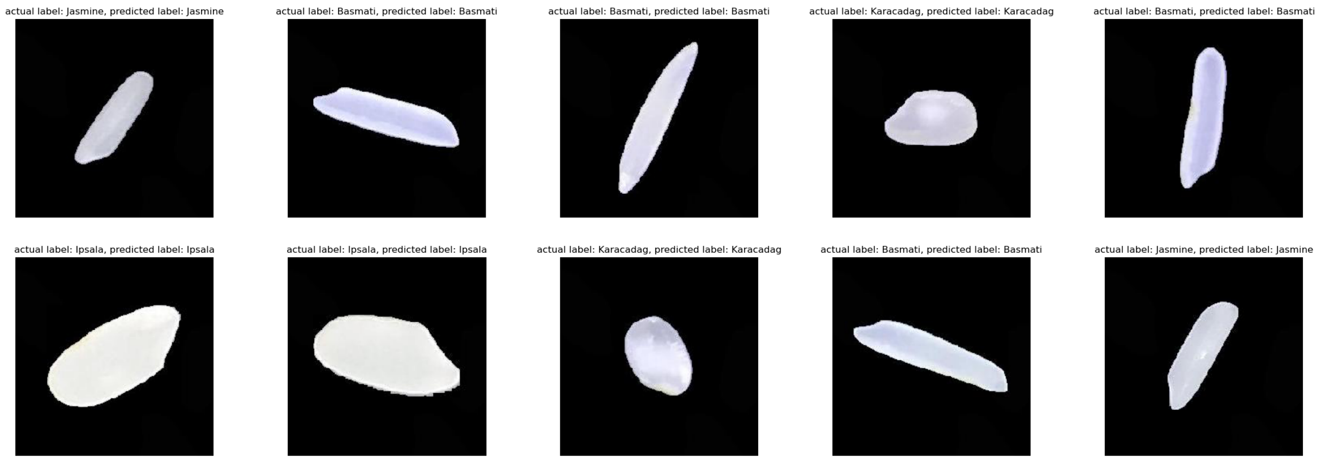

Figure 7. Classification with ResNet50.

Methods	Loss	Accuracy
CNN	0.0270	0.9944
ResNet50	0.0759	0.9762

6. Conclusion

This study aimed to evaluate the performance of two different machine learning algorithms in classifying rice varieties using image features. Performance metrics based on the validation Loss and accuracy were used to compare the algorithms for each method. Results showed that the CNN method achieved the highest average classification success rate of 99.44%, likely due to its ability to directly process images and use hidden features such as size and color. The ResNet50 method had the second-highest success rate at 97.62%, thanks to its ability to perform a wide range of learning in large data sets.

References

[1] Ki Wook Bae, Jeong Yun Hwang, Sang-il Kim, Hyung Mo Jeong, Sunuk Kim, Jae-Hong Lim, Hyun-Sik Kim, and Kyu Hyoung Lee. Thermoelectric transport properties of n-type sb-doped (hf,zr,ti)nisn half-heusler alloys prepared by temperature-regulated melt spinning and spark plasma sintering. *Applied Sciences*, 10(14), 2020. 1

[2] Varsha Bapat, Prasad Kale, Vijaykumar Shinde, Neha Deshpande, and Arvind Shaligram. Wsn application for crop protection to divert animal intrusions in the agricultural land. *Computers and Electronics in Agriculture*, 133:88–96, 2017. 1

[3] Celano Giuseppe. A resnet-50-based convolutional neural network model for language id identification from speech recordings. *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP*, 2021. 1

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 3

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3

[6] Nadia Jmour, Sehla Zayen, and Afef Abdelkrim. Convolutional neural networks for image classification. In *2018 International Conference on Advanced Systems and Electric Technologies (IC*$_ASET), pages 397 - -402$, 2018. 1

[7] Murat Koklu, Ilkay Cinar, and Yavuz Selim Taspinar. Classification of rice varieties with deep learning methods. *Computers and Electronics in Agriculture*, 187:106285, 2021. 1

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 2

[9] Sachin Padmanabhan. Convolutional neural networks for image classification and captioning. 2016. 1

[10] Devvi Sarwinda, Radifa Hilya Paradisa, Alhadi Bustamam, and Pinkie Anggia. Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer. *Procedia Computer Science*, 179:423–431, 2021. 5th International Conference on Computer Science and Computational Intelligence 2020. 3

[11] Farhana Sultana, Abu Sufian, and Paramartha Dutta. Advancements in image classification using convolutional neural network. In *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, nov 2018. 1

[12] Xi Tian, Jiangbo Li, Qingyan Wang, Shuxiang Fan, Wenqian Huang, and Chunjiang Zhao. A multi-region combined model for non-destructive prediction of soluble solids content in apple, based on brightness grade segmentation of hyperspectral imaging. *Biosystems Engineering*, 183:110–120, 2019. 1

[13] Ross Wightman, Hugo Touvron, and Hervé Jégou. Resnet strikes back: An improved training procedure in timm, 2021. 1