Rice Seed Varieties Classification at Various GPS Locations

Ohnmar Khin
Department of Multimedia Engineering
Sunchon National University
Jeollanam-do, Suncheon-si, Seokhyeon-dong, 255,
South Korea.

Sung Keun Lee
Department of Multimedia Engineering
Sunchon National University
Jeollanam-do, Suncheon-si, Seokhyeon-dong, 255,
South Korea.
Corresponding e-mail: sklee@scnu.ac.kr

***Abstract*—One of the most valuable subsistence crops that sustains people worldwide is rice. A strong system for analyzing and classifying rice grains can greatly increase efficiency in terms of time and accuracy. Because of its socio-economic significance, this field of study has attracted much attention in recent decades. This study reviewed the work on image-based rice classification and grading. This paper makes two contributions: first, it separates the historical literature on rice grains into different periods. Second, it splits the methods and algorithms in this field into two categories: deep learning and supervised learning. Among them, the Alex Net model, VGG16, and MLP classifier have yielded more encouraging outcomes and drawn interest for further study.**

Keywords—rice seeds, classification of several rice classes, supervised learning, deep learning.

I. Introduction

Rice seed variety classification is explanatory research. Exploratory research examines broad concepts and discovers why and how things happen. The main goal of this research is to comprehend the cause-and-effect relationship between rice seed varieties. When rice varieties are successfully classified, automated agricultural systems find use for them, particularly to improve the effectiveness of seed quality classification and assessment.

We utilized 1000 images for this investigation and gathered 5 different types of rice from Kaggle. Rice grains were collected from multiple sources to prepare this dataset. Five rice classes are Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Five classes of rice grains comprise the dataset, and each photo has been identified with the rice type to which it belongs. 5 different classes of different varieties of rice are included in a diverse dataset. The quality and flavor of these rice grains are well-known. Fig. 1 shows five rice classifications: the training, validation, and test sets.

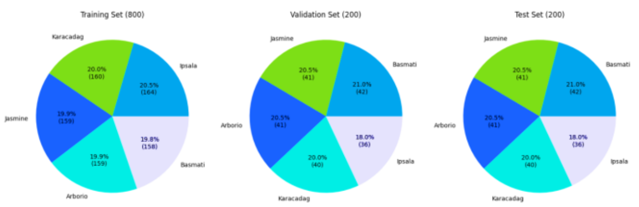

Fig. 1. Rice classes: Arborio, Basmati, Ipsala, Jasmine, and Karacadag.

The photos were preprocessed to standardize their dimensions and improve model performance before model training. One of the preprocessing methods is to resize each image to have uniformly sized pixels. Increasing the diversity of the dataset through augmentation and normalizing pixel values to the range [0,1].

In this project, we employ deep learning, supervised algorithms, and Jupyter Notebook to assist farmers in identifying the variety of rice they are cultivating. Initially, multiple photos of different types of rice are gathered. After that, we use these images to instruct our computer software on differentiating between these grains. We employ various architectures of CNNs and supervise learnings to accomplish this rice grain classification. The detailed workflow of our system's procedure is shown in Fig. 2. Farmers can use our systems to take images of their rice and receive the classified rice variety images after our system is implemented. We guarantee that our system will be usable and will keep getting better over time. This makes it possible for farmers to select the best seeds and achieve higher harvests.

The structure of this study is as follows. Section II reviews the earlier papers. Section III of the paper covers the dataset, cross-validation, performance measures, and study methodologies. The experimental findings from the investigation are explained in Section IV part. The conclusion and recommendations are discussed in the final section.

A. Motivation

Many studies that consider the deep learning of rice grain classification such as CNN, Resents, Mobile NetV2, Neural Network, and Transfer Learning have already been performed. In this paper, supervised learning algorithms of Logistic Regression, Linear SVC, Random Forest Classifier, and Multi-layer Perceptron (MLP Classifier) are used for rice grain classification in addition to Convolutional Neural Networks (CNN), Alex Net model, VGG16, MobileNetV2 are utilized to arrive at an improved accuracy comparison for classification. These proposed models perform better when dealing with unidentified rice, overlapping rice, and irregular lighting rice.

B. Workflow

An outline of our research workflow is provided here:

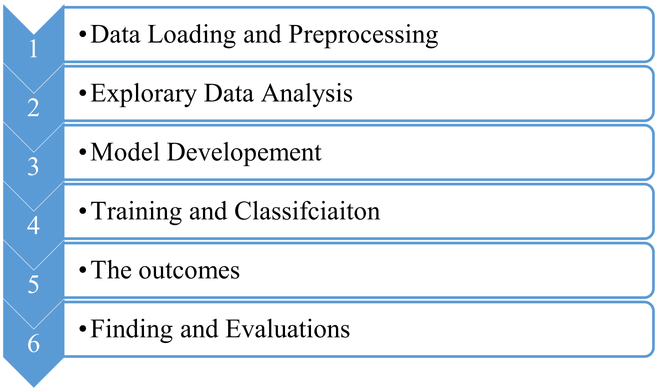

Fig. 2. Workflow.

II. RELATED WORK

The rice types were categorized, and their quality was assessed using two sophisticated deep-learning models, ResNet50 and Xception. The first preprocessing step was to apply Canny Edge Detection. The research divides rice quality into three categories: excellent, good, and fine. This summary illustrates the study's contributions to rice quality analysis and classification [1].

The second research aims to detect rice seeds by applying machine learning and deep learning techniques. A novel 34-layer architecture of convolutional neural networks (CNNs) and machine learning is used to maximize computing efficiency. Several CNN architectures are compared in this study. The results show where the accuracy of the model needs to be improved, with a confusion matrix revealing several misclassifications [2].

In the third portion of the literature, systems for rice variety identification and quality assessment have been successfully implemented, as demonstrated by the research, pointing to real-world uses for the conclusions. All these factors taken together support the study's assertions on the efficiency of deep neural networks for rice quality analysis and classification [3].

As the fourth paper demonstrates, this research compares the classification of rice seeds using machine learning and deep learning approaches. It achieves a 98% accuracy rate when evaluating the performance of random forests and extreme learning machines. This is crucial in fulfilling Indonesia's growing need for imported rice. A confusion matrix showing misclassifications is one of the outcomes, pointing out areas where the model's accuracy has to be improved [4].

In the fifth paper [5], for rice seed variety classification, the Random Forest method outperformed the Support Vector Machine with an accuracy of 95.12%. The High-precision Residual Network (HResNet) achieved a classification accuracy of 95.13% for distinguishing rice grown in various geographic settings. Overcoming the difficulties posed by rice seed variations' visual similarity. Effects on the agriculture sector's marketing and production.

The comparison for the last paper on deep learning-based methods has been widely used for decades to identify the impacted areas of the photos. Of all the deep learning methods, convolutional neural networks (CNN) have shown the highest level of effectiveness in this area. This research proposes an enhanced CNN-based method for classifying photos. Comparative results point to significant gains over current models in terms of classification accuracy and computational time of validation loss for both training and testing data [6].

III. METHODS

The steps included to contribute to this research are as follows:

- Analysis of papers related to the classification of rice seed varieties using deep learning and supervised algorithms.
- Establishment of seven model architectures development environments.
- Utilize these structures to design rice seed classifications.
- Implementation of deep learning and supervised learning-based rice seed classification algorithms and performance analysis.
- Comparative analysis of results is carried out by several earlier algorithms.

A. Pointers to the dataset

Rice is a grain commodity of worldwide significance, with a wide range of genetic variations. The Kaggle database is used to download 1000 rice grain photos in total, divided into 5 classes: Arborio, Basmati, Ipsala, Jasmine, and Karacadag.

- Arborio: This rice, well-known for its high starch content and small grain texture, is usually used in risotto because it absorbs liquids while maintaining a firm texture.
- Basmati: This long-grain rice is renowned for its long, fragrant grains and for being frequently used in many different recipes, particularly in South Asian cooking.
- Ipsala: Grown throughout Turkey, Uppsala rice is prized for its medium-grain shape and ability to cook.
- Jasmine: Named because of its aroma, jasmine rice has a slightly sticky texture and is frequently utilized in Southeast Asian cooking.
- Karacadag: This rice type is unique for its growing environment and may have special flavor, texture, or cooking qualities that appeal to regional tastes.

B. Pre-processing

The preprocessing procedure consists of four phases. These include resizing, tensor conversion, normalization, and folder. The preprocessing steps are displayed in Fig. 3.

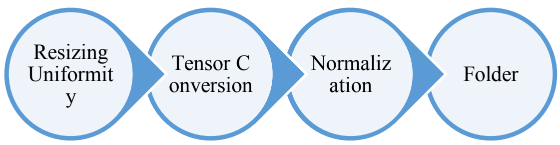

Fig. 3. Stages of preprocessing.

Below are the steps of the data preprocessing for this study:

- Resizing Uniformity: It's critical to standardize image sizes. To promote uniformity throughout the collection, we ensure every image is the same size.

Tensor Conversion: Images are converted into PyTorch tensors, a basic format for the input.

- Normalization: Use normalizes to adjust pixel values according to the standard deviation and mean. This crucial stage improves the model's generalizability to a variety of data.
- Folder: Based on the directory layout, the robust Image Folder class automatically groups images into classes, simplifying the dataset preparation process. Images unique to each class are stored in their folders. The changes are implemented smoothly during the dataset's loading, which is noteworthy because it creates a ready-made dataset for the subsequent training of models.

This section of the study article explains how to classify different types of rice using the suggested Alex Net model, VGG16, CNN model, Mobile Net, MLP Classifier, Logistic Regression, Random Forest Classifier, and Linear SVC.

A. CNN (Convolutional Neural Network)

- First, an input layer is defined to specify the shape of the input data. Set the input shape (50,50,3) for the model.
- Then, a sequential model is constructed to define the architecture of CNN.
- The output from the convolutional layer undergoes flattening, transforming the 2D matrix data into a 1D vector. This flattened representation is then fed into fully connected layers. The last layer serves as the output layer, containing 10 labels.
- Finally, the model is compiled and printed the the accuracy.

B. Mobile Nets

Depending on the level of the convolutional layers, a multiplier controls the size of the mobile nets. To manage inference speed, they are trained for different input image sizes. Tasks involving image classification are handled by MobileNets. MobileNets take certain actions. First, download the TensorFlow model's pre-trained, model weights and architectures. Next, use the TensorFlow Lite converter to convert the model to TensorFlow Lite format. Using a TensorFlow model as input, this converter creates a TensorFlow Lite model file that can be loaded and used on mobile devices.

C. Alex Net Model

Alex Net comprises:

- two normalized layers
- two fully linked layers
- three max-pooling levels
- five convolution layers
- one SoftMax layer

Each convolution layer is made up of a non-linear activation function known as "ReLU" and a convolution filter. The max-pooling function is carried out by the pooling layers. These are the main features: Utilized as an activation function is "ReLU." The batch size is 128 SGD. One learning algorithm is momentum. Data augmentation techniques include color normalization, flipping, jittering, and cropping.

D. VGG 16

A classification method called VGG16 has a 98% accuracy rate in classifying 1000 photos across 5 image classes. It is among the widely used algorithms for classifying images. The CNN-type VGG16 is regarded as one of the best models for computer vision available today. By employing an architecture with extremely tiny convolution filters, the model's developers assessed the networks and enhanced the depth, demonstrating a notable advancement over previous setups. Additionally, the depth was increased to 18 weight layers, resulting in roughly 138356392 trainable parameters.

E. Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the most widely used classifiers. SVM is mainly used for classification problems, but it can also be used for regression problems. For the classification of rice grains, the SVC of supervised learning systems is 94% accuracy.

F. Random Forest

A random forest is a meta-estimator that employs averaging to increase predictive accuracy and manage over-fitting while fitting certain decision tree classifiers on different dataset subsamples. The best-split technique is employed by the forest's trees.The max_samples parameter controls the sub-sample size if bootstrap=True (default); if not, the entire dataset is used.

G. Logistic Regression

Logistic regression's outcome variables are categorized. In logistic regression, outcome variables can be three types of categorical variables. The predictive variable is also categorical, and continuous. Logistic regression can be bivariate or multivariate.

In the supervised learning model comparison, the logisti c regression report has a precision of over 90.

H. MLP Classifier

hidden_layer_sizes=(5,2),max_iter=300,activation='relu', and solver='adam', then mlp_clf = MLPClassifier.

The number of layers and nodes in the neural network classifier is specified with this parameter. The total number of hidden layers in the neural network is thus shown by the length of the tuple. max_iter: many epochs there is activation: The hidden layers' activation function. The algorithm for weight optimization across the nodes is specified by this parameter.

- To identify rice photos, we suggested using CNNs and compared their performance. We aim to select the model that performs best in predicting new data. We prepared our dataset and used the train_test_split function to split it into training and testing sets. 80%

was set up for training, while 20% was dedicated to pre-testing.

- This project aims to use deep learning and supervised models to classify five different types of rice. To monitor its performance, we compared the performance of our traditional model with the previously trained model. We created and trained CNN models on our dataset to achieve excellent accuracy in rice variety classification. Fig. 4 displays the accuracy of SVC, random forest classifier, logistic regression, and MLP classifier.
- In a comparison analysis, eight algorithms of models are assessed and their performances are compared according to earlier research. When comparing several architectures of CNNs (Alex Net, VGG Net, Mobile Net) to the task of rice grain variety classification, our comparative analysis evaluates the algorithms' accuracy, computational speed, use of memory, or additional important factors. Eight classifiers are the models that will be evaluated.

IV. Results and Discussion

By hand, laborers examine the rice grains. Identifying which rice variety is particular as many types are cultivated becomes increasingly challenging. It takes a lot of time and is prone to human mistakes. Accordingly, we develop the capacity to recognize the genus of rice grain with speed and accuracy. More specifically, the solution for the rice species is a mixed strategy based on CNN techniques.

Several different machinelearning approaches have been tested using rice photos and the features that are produced fr om each of these images.The success of algorithms in testin g and training has been estimated using these metrics. Fig. 5 illustrates the results of the classification of rice seed varieties that we expected. Table 1 shows the precision, recall, and f1-score of the supervised learning models that classify rice grain varieties in the five rice classes of Arborio, Basmati, Ipsala, Jasmine, and Karacadag.

Statistical measurements of the confusion matrix resultin g from categorization are used as evaluation metrics.Accura cy, recall, precision, and F1_score statistics are collected an d aggregated for each technique and category.The maximum rate of 98 percent for total classification success is achieved by the AlexNet and VGG16 approaches.Classification succ ess is determined to be good since the learning process incor porates a lot of hidden information, including size, color, etc ., and analyzes photos. The MLP approach has also had great success using this tactic. The MLP strategy had a 97 percent classification accuracy rate, which was good when compared to another study methodology.

AlexNet and VGG16 are the top classifiers when the accuracy, recall, precision, and F1-score values are examined.

The accuracy of the AlexNet model, VGG16, CNN model, Mobile Net, MLP classifier, Logistic Regression, Random Forest classifier, and Linear SVC are displayed in Table 2. An intelligent rice classification system is needed to evaluate the rice sample and guarantee its quality. The rice species can then be promptly and precisely recognized.

TABLE I. Classification Report for Supervised Learning Models

Classification Report for Supervised Learning Models

	precision	recall	f1-score
Linear SVC	0.94	0.94	0.94
Random Forest Classifier	0.94	0.93	0.94
Logistic Regression	0.95	0.95	0.95
MLP Classifier	0.97	0.96	0.97

Our findings show that the algorithms performed better than other algorithms in terms of accurate grain classification. Our experimental results are compared with the previous research investigations. The accuracy of the CNN approach, which is cited in [6], is 89.50%. Table 3 displays a comparison between the current study and previous research. Our investigated study improves crop forecasts, improves crop management, and boosts agricultural output. Future work can investigate how to enhance and improve our model, either by adding new sources of data or advanced techniques for deep learning.

TABLE II. Comparison of the accuracy of Seven models

	Accuracy	
1	AlexNet Model	98%
2	VGG16	98%
3	CNN Model	92%
4	Mobile Net	56%
5	MLP Classifier	97%
6	Logistic Regression	95%
7	RandomForest Classifier	94%
8	Linear SVC	94%

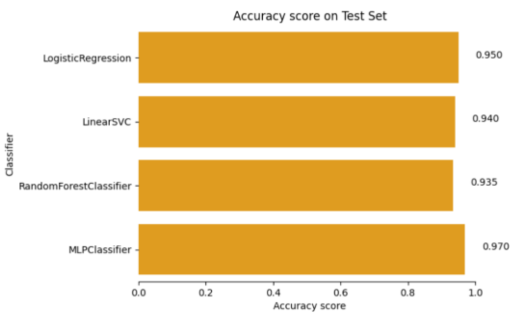

Fig. 4. Accuracy results from supervised learning models.

TABLE III. A COMPARISON BETWEEN THE PRESENT WORK AND EARLIER RESEARCH

Ref.	Year	Dataset	Classes	Methods Used	Accuracy
6	2017			CNN	CNN: 89.50%
Current Research	2025	1000	5	Deep Learning Supervised Learning	AlexNet: 98% VGG16: 98% CNN: 92% MobileNet: 56% MLP: 97% Logistic: 95% Random Forest: 94% SVC: 94%

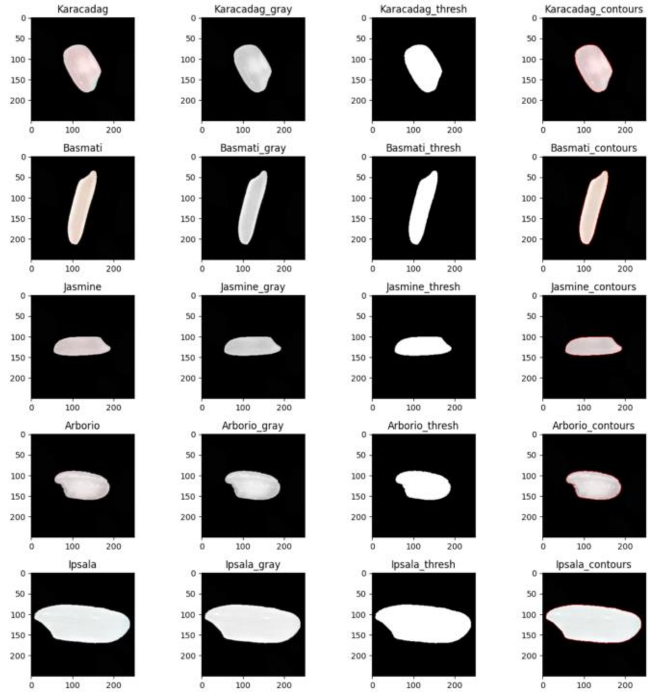

Fig. 5. Findings from the CNN model's classification of several types of rice seeds.

V. CONCLUSION

This article examines current deep learning and rice grain classification methods. The history of algorithms for rice grains is reviewed in this study. The current study's goal is to suggest an efficient Alex Net model, VGG16, and MLP classifier for rice variety classifying. For this challenge, the models Alex Net, VGG16, CNN, Mobile Net, MLP classifier, logistic regression, random forest, and liner SVC are used. The rice dataset that was gathered for the Kaggle repository was used to train, validate, and test the model. The table displays the suggested model's achieved accuracy in terms of precision, recall, and F1-score. As a result, the suggested model is efficient and has learned the many types of rice.

Additionally, they are in many accurate running detection models. This study will face significant challenges due to data sets with irregular lighting, closing between rice grains, and unidentifiable hidden rice grains. We conclude that deep learning and supervised approaches produce encouraging outcomes.

ACKNOWLEDGMENT

I sincerely thank the Grand-ICT. This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2025-2020-0-01489) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] V. S. Lakshmi, K. Seetharaman, “Rice Classification and Quality Analysis using Deep Neural Network,” pp. 307-314, September 2022.

[2] B. D. Satoto, D. R. Anamisa, M. Yusuf, M. K. Sophan, S. O. Khairunnisa, B. Irmawati, “Rice seed classification using machine learning and deep learning,” IEEE, Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Bali, Indonesia, January 2023.

[3] V. Lakshmi, K. Seetharaman, “Rice Classification and Quality Analysis using Deep Neural Network,” International Conference on Intelligent Innovations in Engineering and Technology (ICIIET), IEEE, December 2022.

[4] B. Satoto, D. Anamisa, M. Yusuf, M. K. SophanSiti, O. Khairunnisa, B. Irmawati, “Rice seed classification using machine learning and deep learning,” Computer Science, Agricultural and Food Sciences, International Conference on Intelligent Computing, 2022.

[5] M. Sandhiya, Dr. S. Christy, “Enhancing the Accuracy in Classifying Rice Seed Varieties using Novel Random Forest Algorithm Compared with Support Vector Machine,” pp. 410-414, December 2023.

[6] F. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks", 2017.