ISSN 2063-5346

CLASSIFICATION OF RICE VARIETIES USING MACHINE LEARNING TECHNIQUES FOR AGRICULTURAL APPLICATIONS

[1] Sampath Kumar S. & [2] B.N. Manjunatha Reddy

Article History: Received: 10.05.2023 Revised: 29.05.2023 Accepted: 09.06.2023

Abstract

In this research paper, the classification of rice varieties using machine learning techniques for agricultural applications is presented along with the simulation results. Rice, the world's largely cultivated and consumed cereal crop, is also one of our country's most important sources of sustenance due to its economic and nutritional worth. Before reaching our tables, rice goes through multiple industrial stages, including washing, colour sorting, and categorising. Cleansing is the separation of extraneous elements from rice; classification is the separation of broken rice from sturdy rice; and colour extraction is the separation of discoloured and striped rice from whiteness on the rice surface. In this study, a computer vision system was developed to distinguish between two proprietary rice kinds. For the two species, 3810 rice grain pictures were captured, analysed, and feature inferences were produced. Each grain of rice was given seven morphological characteristics. SVM, RF, BB (Bagging Boost), and AB (Ada Boost) machine learning algorithms were used to generate models using these characteristics, Results of performance measurement were obtained. The success rates for classification were 90% (SVM), 99% (RF), 61% (AB), and 99% (AB) (BB). We may say that the study was a success based on the findings of the acquisition success rate. The simulation results shows the effectiveness of the methodology that is being developed & presented in this paper.

Keywords — Machine Learning algorithms, ML models and Classification of rice.

[1] Assistant Professor, Dept. of Artificial Intelligence & Data Science,

Don Bosco Inst. of Tech., Bengaluru, India & VTU Research Scholar.

[2] Professor, Dept. of Electronics & Communication Engg.,

Global Academy of Technology, Bengaluru, Karnataka

Emails : sampath.ksit@gmail.com manjunatha_reddy@gat.ac.in

DOI:10.48047/ecb/2023.12.9.19

1. Introduction

Agriculture is recognised as a critical component of the global economy since it meets one of humanity's most basic requirements, namely food. In the majority of countries, it is recognised as the key source of employment. Numerous countries, like India, continue to practise traditional farming practises. Farmers are unwilling to adopt modern agricultural technology owing to a lack of competence, high expenses, or ignorance of the technologies' benefits. A lack of awareness of soil types, yields, crops, weather and inadequate pesticide usage, irrigation challenges, wrong harvesting and a lack of market information resulted in the loss of farmers or increased the cost. A lack of knowledge at each level of agriculture causes new problems or exacerbates old ones, raising farming costs. As the world's population continues to rise, the agricultural business is under growing strain. Overall losses in agricultural operations range from crop selection through product sale and are extremely significant.

Keeping track of information about crops, the environment and the market, according to the adage "Information is Power" may help farmers make better decisions and address agricultural difficulties. IoT and machine learning are examples of cutting-edge technology, deep learning, cloud computing and edge computing may be used to collect and analyse information, therefore increasing productivity, improving quality, and eventually increasing the profitability of farmers and related areas. Precision learning in agriculture is critical for increasing total crop output. When the worldwide production values of grain products are considered, rice ranks third after wheat and maize. Rice is a carbohydrate and starch-rich grain. It is also crucial in human nutrition because of its economic and nutritional worth. At the same time, it is widely employed in a variety of industries.

There are several quality requirements for rice types grown all over the world. Rice characteristics such as cooking capabilities, physical appearance, aroma and taste traits, and production are examples of these aspects. The first consideration for rice kinds sold in packages on market shelves from the perspective of the eventual customer is physical. As a result, more technologically advanced and effective approaches are required. It is highly wasteful and time-consuming to calibrate rice and segregate it according to numerous quality standards throughout manufacturing, especially given the huge production volume. In the literature, recent investigations employing Approaches to image processing and machine learning have been studied. In one of these studies, the projection areas of many grain products such as wheat, barley, corn, chickpeas, lentils, beans, kidney beans, and soy were analysed in three different locations using an image processing technique.

Regression analysis was used to analyse the correlations between the projection regions utilising the UTHSCSA (University of Texas Health Science Centre, San Antonio) image processing tool, which provided feature values such as length, width, and thickness of the products used in the study. According to the findings, the image processing method is competent for precisely defining the projection zones of fine grain objects. The UTHSCSA Imaging Tool Version 3.0 programme was used to examine the morphological properties of 13 different wheat varieties from bread and durum wheat kinds. As a result, the findings of manual measurements and image processing were found to be equivalent, and the image processing technique may be utilised to analyse some of the morphological aspects of wheat grains.

Another study employed dried tobacco leaves to build a system for automatic leaf evaluation using machine vision techniques. The system's purpose is to analyse tobacco leaves based on their colour, size, shape, and surface roughness. According to the findings of the trials, this technique is an excellent route for dried tobacco leaves. Additionally, it is claimed that these tobacco leaf properties may be used for automatic classification, which is the focus of the following studies. As a result, the authors suggested the machine learning (ML) system in this research by imposing ML models on the provided dataset where the accuracy of the algorithms is assessed and the rice is classified into five categories, viz., 1) Basmati rice 2) The arborio Jasmine rice 3) Ipsala rice 4) and 5) Karacadag rice.

2. Literature Survey

Visen *et.al.* created methods to assess barley, rye, oats, wheat, and durum wheat grain photos They used an artificial neural network with image processing techniques in their research. They achieved a success rate of more than 90% for all grain kinds using the generated photos, over 150 colour and textural criteria, and the classifier they built in grain identification [1]. In their research with wheat grains, Baykan et al. produced a grain grey level average with 9 morphological parameters for the photographs they received. They were successful in classifying 72.62% of the 5 distinct species utilised. Nevertheless, after they excluded a species that was difficult to classify, their success rate increased to 82.65% [2].

Dubey *et.al*. employed three different varieties of bread wheat in their research. They got around 88% accuracy by eliminating 45 morpho-logical characteristics for the artificial neural network categorization [3]. Zapoto Czny *et.al*. used an image analysis approach to infer 74 pieces of morphological traits for the categorization of five distinct barley species. As a classification approach, they employed Basic component analysis, linear and nonlinear discriminant analysis are all examples of discriminant analysis. As a result, they determined that LDA is the best classifier approach [4].

Chen and their *et.al*. collected 13 forms, 17 geometric and 28 colour characteristics from photos obtained from 5 different maize types in their investigation. They discovered indicates the average accuracy percentage across maize varieties was 90% as a consequence of their categorization [5]. Babalik *et.al*. employed In their work, they used multi-class support vector machines and binary particle swarm optimization approaches to categorise 9 geometric and 3 colour characteristics from 5 wheat species. They achieved 91.5% classification accuracy using the M-SVM approach and 92.02% using the BPSO methodology [6].

Yadav and Jindal used digital image analysis to determine milled rice quantifications by eliminating the perimeter, length, and shape properties of the rice grain [7]. Using image processing techniques, Ouyang et al. developed an automated photo capture band system to distinguish 5 unique rice seeds. The seeds were photographed as they went through the tape system using a CCD camera. Visual C++ 6.0 was used for picture analysis. They attained an average success rate of 86.65% for 5 unique rice kinds using back propagation classification [8]. Farahani used image processing technology to remove morphological information from 5 samples of Durum wheat variety to create five groups of attributes for linear discrimination analysis.

In the study, 11 morphological parameters were used, yielding the greatest classification accuracy of 67.66% [9]. Silva and Sonnadara used an artificial neural network to classify rice varietals in their study. They extracted 13 morphological characteristics, 6 colour features, and 15 texture features using algorithms data from photos of nine distinct rice kinds. They classified these qualities differently, both independently and together. When the categorization was done individually, it was discovered that tissue characteristics performed better than morphological and colour criteria. The categorization using all characteristics together resulted in a 92% success rate [10].

In their study, Kaur and Singh examined rice classification using multi-class support vector machines. They calculated their percentages for each one and defined their characteristics by isolating the rice from which their geometric features were removed. They had a higher success rate than 86% as a consequence of the classification [11]. Abirami *et.al*. classified basmati rice species using image processing and neural network pattern recognition techniques in their study. They employed filtering, thresholding, and edge detection approaches to remove unique morphological characteristics of rice grains before categorising them with neural network pattern recognition.

As a result, they enjoyed a 98.7% success rate [12]. Pazoki *et.al*. successfully classified five unique rice varieties utilising 24 colour, 11 morphological, and four form factor factors using ANN and Neuro-Fuzzy machine learning algorithms [13]. In their research, Szczypiski *et.al*. studied the efficacy of cultivar identification on barley varieties based on shape, colour, and texture features. They employed linear distinctive analysis and artificial neural networks for classification. As a consequence, their success rates ranged from 67% to 86% [14].

3. Proposed Architecture

The goal of this project is to identify distinct rice kinds using the dataset's supplied attributes, namely, 1) Basmati rice. 2) The arborio Jasmine 3) Ipsala (4^{th}) and Karacadag (5^{th}). The primary characteristics are listed in the table 1.

Table 1. Offers the morphological characteristics from the dataset used for rice categorization

No.	Name	Explanation
1	Area	Returns the no. of pixels within the boundaries of the rice grain.
2	Perimeter	Calculated the circumference by calculating the distance between the pixels around the boundaries of the rice grain
3	Major Axis Length	The longest line that can be drawn on the rice grain, i.e., the main axis distance gives
4	Minor Axis Length	The shortest line that can be drawn on the rice grain, i.e., the main axis distance, gives
5	Eccentricity	It measures how round the ellipse, which has the same moments as the rice grain is.
6	Convex Area	Returns the pixel count of the smallest convex shell of the region formed by the rice grain.
7	Extent	Returns the ratio of the region formed by the rice grain to the bounding box pixels

A. Rice Dataset

In the study, five different types of rice were employed. Arborio, Basmati, Ipsala, Jasmine, and Karacadag rice types were chosen and sold under the Metro Chef Brand. Figure 1 depicts the rice types employed in the study, while Table 1 provides technical information on the cultivars.

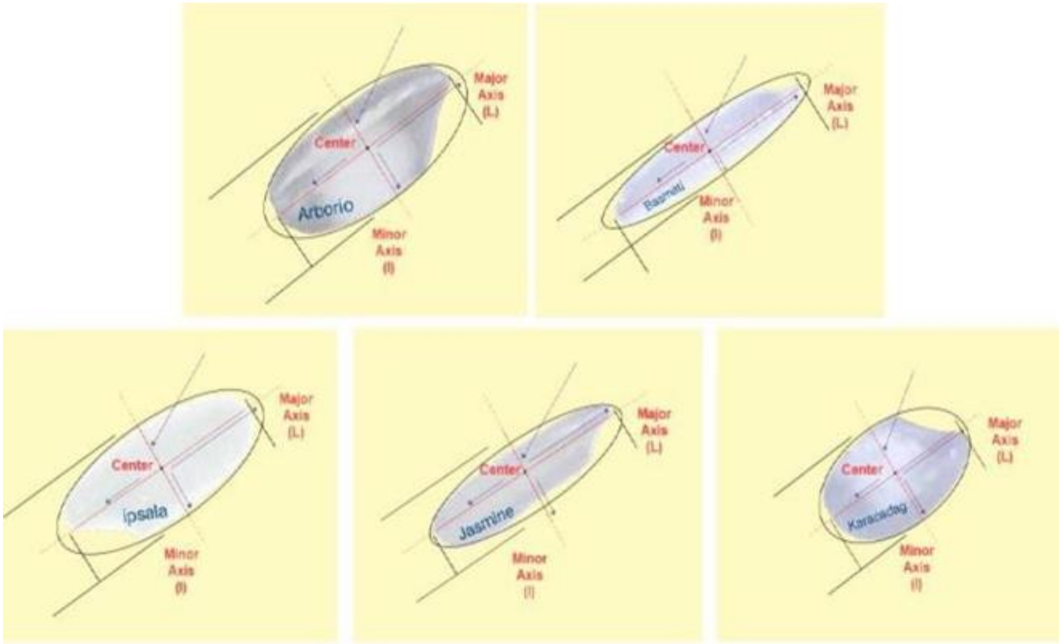

Figure 1. demonstrates the many classifications of rice depending on its shapes.

B. Support-Vector-Machine Classifier

Support vector machine (SVM) is a hyperplane-creating core-based technique for classification and regression. Using separation approaches, SVM can classify data as linear in two-dimensional space, planar in three-dimensional space, and hyperplane in multidimensional space. SVM does classification by selecting the best hyperplane that separates the data. As shown in Figure 2, the best hyperplane for an SVM is the one with the largest margin between the two classes. SVMs have similarities to other machine learning approaches. It is quite similar to neural networks, although it is more comparable to the k-NN approach. SVM, like the k- NN algorithm, computes its neighbours based on the sample data provided to the algorithm and anticipates making predictions for more data [16] for comprehensive computations.

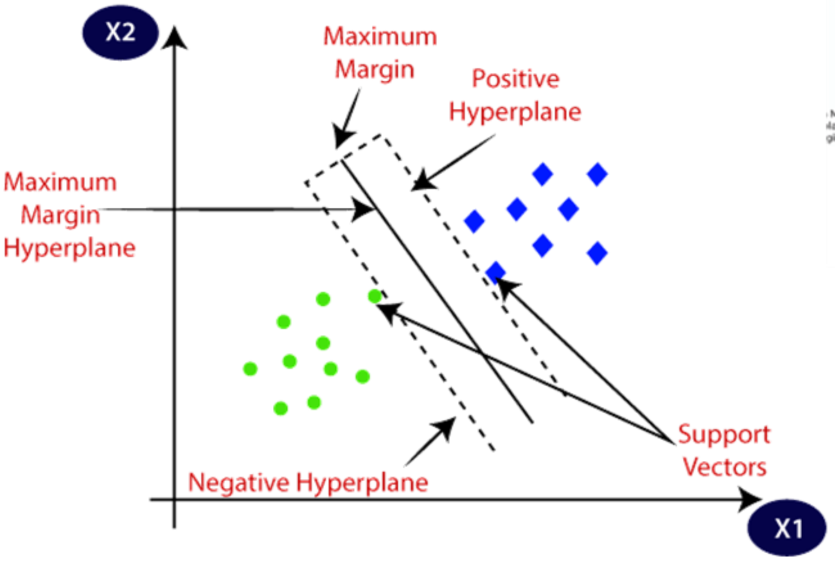

Figure 2 : Support Vector Machine Classifier

C. Random-Forest Classifier

The Random-Forest (RF) classifier is made up of several DTs. Each DT gives a classification for the inputs in order to create a new classification. The classifications are then evaluated by RF, and the estimate with the highest votes is chosen [17]. In a data collection, RF can handle a large number of variables. It is also quite good at estimating missing data. The most serious disadvantage of RF is its lack of reproducibility. Therefore, interpreting the finished model and its following findings is tricky. This is because it has a large number of separate decision trees [18]. Figure 3 depicts the overall structure of RF.

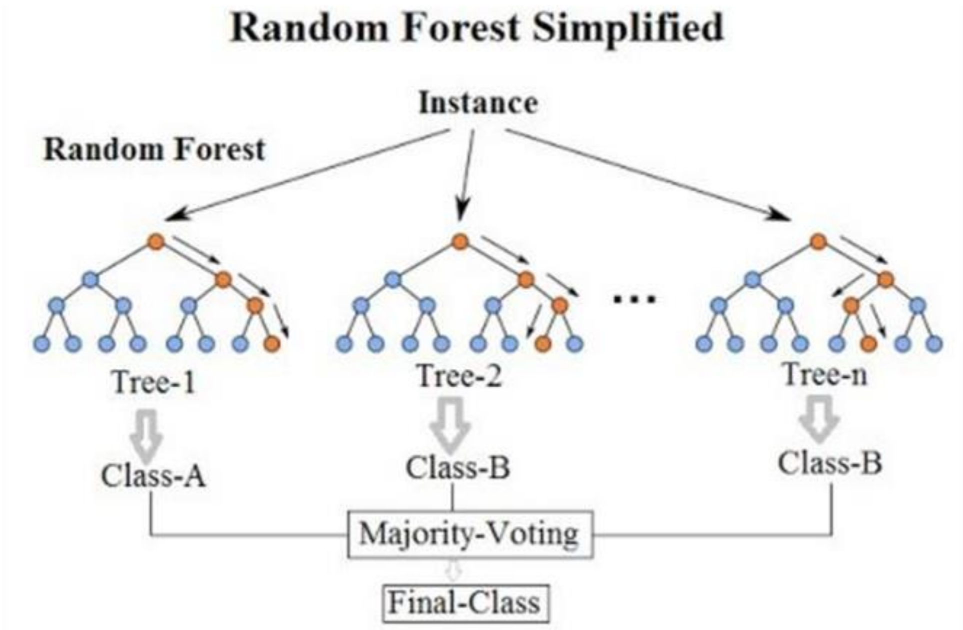

Figure 3: Random Forest Classifier

In 1996, Yoav Freund and Robert Schapire introduced Ada- boost, also known as Adaptive Boosting, as an ensemble boosting classifier. It combines many classifiers in order to increase classifier accuracy. AdaBoost is an iterative ensemble formation approach. The AdaBoost classifier builds a strong classifier by combining many low- performing classifiers, yielding a high-accuracy strong classifier. The primary premise of Adaboost is to construct classifier weights and train the data sample in each iteration to ensure accurate predictions of unusual occurrences. As the fundamental classifier, any machine learning algorithm that uses weights on the training set can be utilised. Adaboost must meet two criteria: The classifier should be trained interactively on a large number of weighted training cases. In each iteration, it strives to provide a good match for these circumstances with error reducing using training.

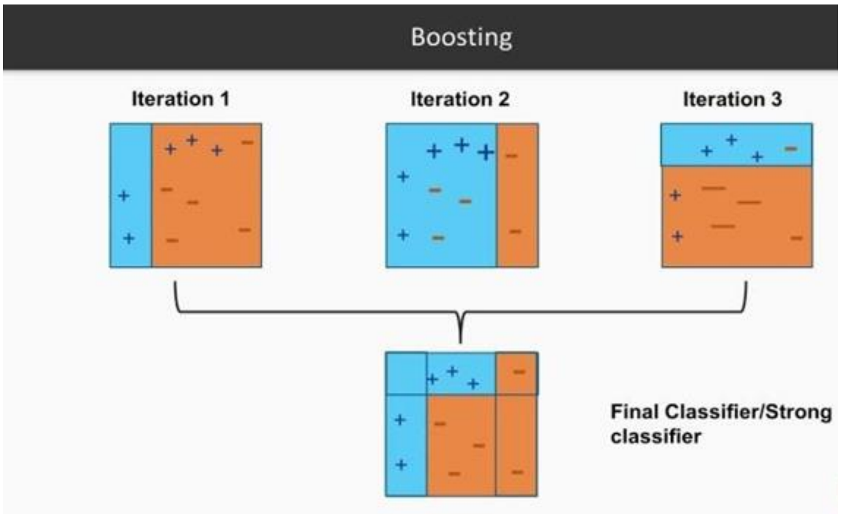

Figure 4: Ada Boosting Classifier

E. Bagging Boosting Classifier

Ensemble Learning models are founded on the premise that by merging numerous models, strong models may be produced. The Ensemble Learning approach combines numerous models, often known as 'weak learners,' to get better outcomes, stability, and prediction performance than any of the constituent models alone. Noise, bias, and variation are the primary causes of machine learning predictions and classification mistakes. Ensemble learning approaches aid in reducing these effects and increasing model stability. Bagging, also known as Bootstrap aggregation, is an ensemble learning strategy that varies the training dataset to find various ensemble learners. Bagging, as opposed to training a single model on the complete dataset, produces numerous weak learners or base models trained on a fraction of the original dataset. The data scientist creating the model determines the number of models to employ and the size of the subsets. Random sampling is employed to produce the data subsets needed to train the weak models, with replacement in each successive model trained. Replacement indicates that the data subset sample used to train each model may contain duplicate data.

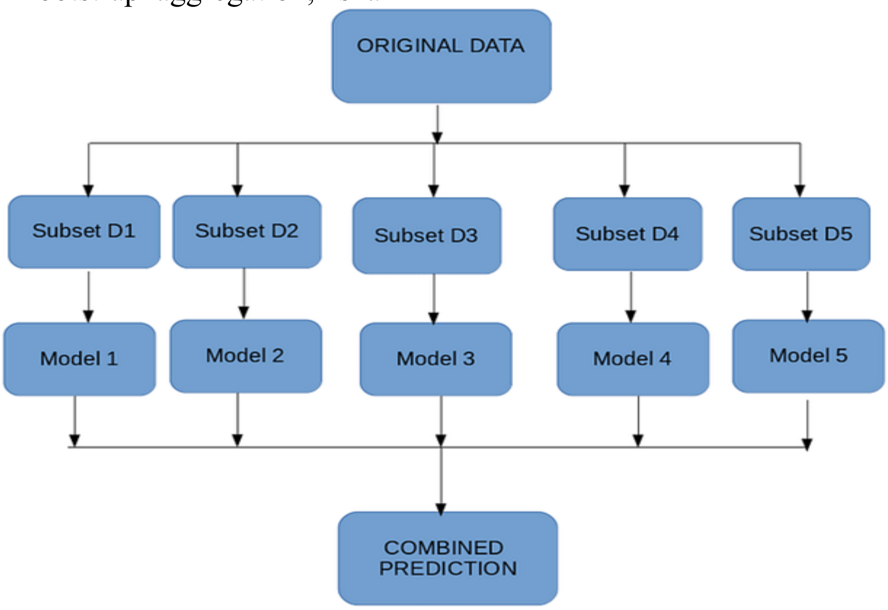

Figure 5: Bagging Boost Classifier

4. Results

Preliminary processing of the CVS pictures was performed in order to categorise the rice kinds used in our study, giving a total of 3810 rice grains. Additionally, each grain has been assigned seven morphological characteristics. For the acquired properties, a dataset has been built. SVM, RF, AdaBoost, and Bagging Boost machine learning techniques for classification were used to develop models and get results. On all models utilised, the cross validation k value is set at 10.

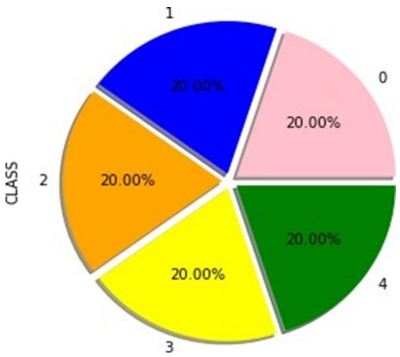

Figure 6: Different Categories of rice derived from the dataset

Here, the rice are numbered as

‘Basmati’ : ‘0’,

‘Arborio’ : ‘1’,

‘Jasmine’ : ‘2’,

’Ipsala’:’3’,

’Karacadag’: ‘4’.

The Accuracy of Support Vector Machine : 90%. The Accuracy of Random Forest: 99%.

The Accuracy of Ada Boost: 61%. The Accuracy of Bagging Boost: 99%.

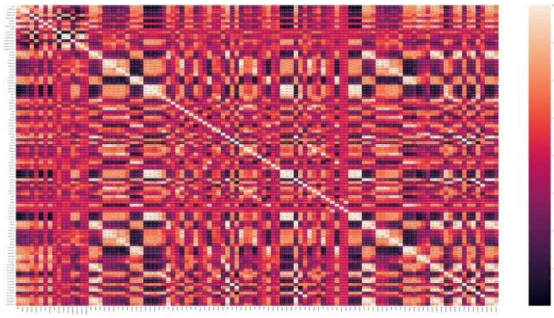

Figure 7: The EDA graph

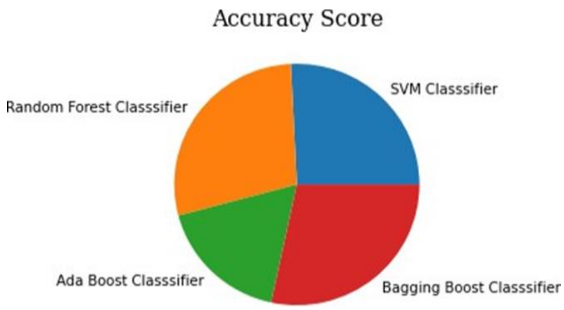

Figure 8: Overall Accuracy

5. Conclusion

In this study, 3810 rice grains were collected from 5 different rice varieties for rice grain classification. These photos were pre- processed using PYTHON software to remove any undesirable components from the image and ready it for the feature extraction step. Previous work of several methodologies was described in this paper, together with its advantages and disadvantages. The classification results produced by using feature normalisation and the Machine Learning model are pretty encouraging. SVM, Adaboost, Bagging Boost, and RF algorithms were employed for classification, which are the most extensively used artificial intelligence approaches. Self-collected datasets-maintained accuracy and recall scores over 99% and 99%, respectively. The results indicate that this method can give an accurate answer to the rice varieties' classification and/or identification problem as an alternative to complex picture segmentation techniques, as previously described. The suggested technique may be utilised for Farmers, where an occupational worker on the field can measure the properties of rice varieties to determine the exact class to which the rice belongs in order to avoid admixture. Confusion Matrix algorithm conclusions were made, and performance evaluations were carried out. Ten cross-validation iteration folds were employed to regulate algorithm generalisation. Our future study will concentrate on expanding our self-collected dataset to include more rice varieties in order to address automated identification of rice varieties in particular and other field crop types in general.

References

[1]. Visen, N.S., et al. Image analysis of bulk grain samples using neural networks. in 2003 ASAE Annual Meeting. 2003. American Society of Agricultural and Biological Engineers.

[2]. Baykan, O., A. Babalik, and F. Botsalı, Recognition of wheat species using artificial neural network.4. International Symposium On Advanced Technologies, Konya, September, 2005: p. 28-30.

[3]. Dubey, B., et al., Potential of artificial neural networks in varietal identification using morphometry of wheat grains.Biosystems Engineering, 2006. 95(1): p. 61-67.

[4]. Zapotoczny, P., M. Zielinska, and Z. Nita, Application of image analysis for the varietal classification of barley:: Morphological features.Journal of Cereal Science, 2008. 48(1): p. 104-110.

[5]. Chen, X., et al., Combining discriminant analysis and neural networks for corn variety identification.Computers and electronics in agriculture, 2010. 71: p. S48-S53.

[6]. Babalik, A., et al. Effects of Feature Selection Using Binary Particle Swarm Optimization on Wheat Variety Classification. in International Conference on Advances in Information Technology. 2010. Springer. Algorithms Confusion

[7]. Yadav, B. and V. Jindal, Monitoring milling quality of rice by image analysis.Computers and Electronics in Agriculture, 2001. 33(1): p. 19-33

[8]. OuYang, A.-G., et al. An automatic method for identifying different variety of rice seeds using machine vision technology. in Natural Computation (ICNC), 2010 Sixth International Conference on. 2010. IEEE

[9]. Farahani, L., Discrimination of some cultivars of durum wheat (Triticum durum Desf.) using image analysis.International Research Journal of Applied and Basic Sciences, 2012. 3(7): p. 1375-1380.

[10]. Silva, C.S. and U. Sonnadara, Classification of rice grains using neural networks.2013.

[11]. Kaur, H. and B. Singh, Classification and grading rice using multi-class SVM.International Journal of Scientific and Research Publications, 2013. 3(4): p. 1- 5.

[12]. Abirami, S., P. Neelamegam, and K.H. Thanjavur India, Analysis of rice granules using image processing and neural network pattern recognition tool.2014.

[13]. Pazoki, A., F. Farokhi, and Z. Pazoki, Classification of rice grain varieties using two Artificial Neural Networks (MLP and Neuro-Fuzzy).The Journal of Animal & Plant Sciences, 2014. 24(1): p. 336-343.

[14]. Szczypiński, P.M., A. Klepaczko, and P. Zapotoczny, Identifying barley varieties by computer vision.Computers and Electronics in Agriculture, 2015. 110: p. 1-8