Machine Vision Approach for Classification of Rice Varieties Using Texture Features

Salman Qadri\textsuperscript{a}, Tanveer Aslam\textsuperscript{b}, Syed Ali Nawaz\textsuperscript{b}, Najia Saher\textsuperscript{b}, Abdul-Razzaq\textsuperscript{a}, Muzammil Ur Rehman\textsuperscript{b}, Nazir Ahmad\textsuperscript{b}, Faisal Shahzad\textsuperscript{b}, and Syed Furqan Qadri\textsuperscript{c}

\textsuperscript{a}Department of Computer Science, Muhammad Nawaz Shareef University of Agriculture Multan (Mns-uam), Multan Punjab, Pakistan; \textsuperscript{b}Department of Information Technology, Islamia University of Bahawalpur, Bahawalpur Punjab, Pakistan; \textsuperscript{c}Computer Vision Institute, College of Computer Science & Software Engineering, Shenzhen University China

\begin{abstract}
The main objective of this study was to assess the machine vision (MV) techniques to classify six Asian rice varieties commonly named as Kachi-Kainat, Kachi-Toota, Kainat-Pakki, Super-Basmati-Kachi, Super-Basmati-Pakki, and Super-Maryam-Kainat (A1, A2, A3, A4, A5, and A6), mainly cultivated in Pakistan, China, India, Bangladesh, and neighboring countries. The sample of each selected rice variety contained 1800 grains, giving a total of 10800 (1800 \times 6) grain samples. A cell phone camera captured the actual field digital images dataset in an open climate. All the captured images were enhanced and converted into the standard 8-bit gray-scale format. Six radius-based non-overlapping regions of interest (ROI’s) were taken on each captured image inducing a total of 3600 (6 \times 600) ROI’s image dataset. We have extracted Binary (B), Histogram (H), and Texture (T) features from each image. We converted these forty-three features for each image into 154800 (43 \times 3600) feature vector (FV) space to discriminate rice varieties. After optimizing the FV, five MV classifiers, namely; LMT Tree (LMT-T), Meta Classifier via Regression (MCR), Meta Bagging (MB), Tree J48 (T-J48), and Meta Attribute Select Classifier (MAS-C), were deployed attaining the classification accuracies as 97.4\%, 97.0\%, 96.3\%, 95.74\%, and 95.2\%, respectively. The maximum overall accuracy (MOA) observed was 97.4\% by LMT-Tree.
\end{abstract}

\section*{Introduction}

Humans need food and nutrition for good health and food products fulfill nutrition needs. Rice is one of the best sources of food and nutrition. Nevertheless which are also providing nutrition supply for millions of Pakistani citizens \cite{1}. Nowadays, in sub-continental (Southern Asia), many researchers are working on rice varieties \cite{2} and their impacts on food chain supply. The agriculture sector is playing an important role in the economy of Pakistan. ‘Rabi’ and ‘Khari’ are two mainly sowing season from November to April and April to November. Rice is also contributing 3.1\% in the agriculture sector and 0.6\% GDP of Pakistan (FY-2020) \cite{3}. Rice has been cultivating in 33,304 (Hectares) of different agricultural land areas in Pakistan location or zones such as Azad Kashmir, Baluchistan, NWFP/North Area, Sindh, and Punjab \cite{4}. The growths of crops were most important for the increase in food production \cite{5}. The Punjab (Pakistan) produces 90.5\% of overall Basmati (Oryza Sativa) rice in Pakistan, and it is very helpful in economic growth \cite{6}. Many varieties of rice are cultivating in the different agricultural districts of Punjab, and Punjab province produced seventy percent of rice
crop [7]. Pakistani rice has always been the 3rd-largest cultivation crop after cotton and wheat. Different rice varieties are cultivated locally namely; Kachi-Kainat, Kachi-Toota, Kainat-Pakki, Super-Basmati-Kachi, Super-Basmati-Pakki, and Super-Maryam-Kainat [3]. Pakistan is the largest rice exporter country in the international market according to the Economic Survey of Pakistan (FY-2020) and has become a billion dollars in agricultural sectors of Pakistan. Gross Domestic Product (GDP) increment is the most important for economic growth and country development. Many food products played a very beneficial role in economic growth and rice is one of them, according to the Economic Survey of Pakistan (FY-2020) [8].

The identification of different varieties of rice was discussed at a huge level during the processing of food elements. The traditional and conventional techniques were used for the identification of different rice varieties, which were collected from different real field sources. This technique was more expensive, a slow process, and time-consuming. Furthermore, there are many varieties of rice and it is very difficult to identify the actual verity of rice with the help of rice grain. So, we need an automated system, which helps us in the identification of rice varieties with better accuracy. The main purpose of this study was describing the identification of rice varieties using different rice grains, which discuss at the start. Human vision approach was faced many hurdles due to some close resemblance of rice grain features namely; color, shape, texture, and size between different rice varieties. The confusing arguments about rice varieties may raise two major problems; (1) water needs of rice crop; (2) the use of different Pesticides and Urea without proper study or knowledge. Due to this reason, rice production affected. Therefore, in this study, we introduced three main directions to solve these problems; (1) implementation of the preprocessing techniques for feature optimization on available raw image dataset; (1) embedding (or arranged) these optimized features by using the sequential method; (3) finally employing different Machine Vision (MV) classifiers on optimized features.

**Literature review**

Segmentation played an important role in image processing [9]. Rice is the cash crop, and it also holds a major place in the agriculture sector of Pakistan. The demand for better quality rice was an increase due to manage food production for humans, and it is also innovation new seeds Apo and IR64 in paddy crops research institute [10]. Rice production and export is a backbone in Pakistan’s economy for earning foreign income. The statically time series (1972–2011 years data) method was applying for the analysis of rice growth, production, cultivation area, and export of rice [11]. Food is the most important element in human life. In Pakistan, the insecurity of rice food has been an impact on the quality and quantity of rice. The availability of rice was changed in different areas of Pakistan [1]. Humans need food and nutrition for good health and living purpose. Metabolism development of rice seeds is most important for increase food yield [12], and in [13], Leaf and roots growth are shown better performance for paddy crop yield in agriculture. Acyl-Coenzyme-A (CoA) protein and transport acyl-CoAs bind was performed better metabolism properties in leaf and roots activities. In [2], Rice & Wheat (R & W) cultivation is the most beneficial cropping system in Pakistan, and Punjab province produced almost 70% agricultural area of rice crops. The crop rotational period was to diagnose the variant problems during ‘Kharif’ and ‘Rabi’ crop farming time duration. The total dataset is collecting from 10 tehsils based on 4 districts in Punjab, Pakistan. Food demand increases day by day due to population growth around the world. R & W system needs most efficiency by implementing together experimental process. R & W cropping system improved the quality and use of resource efficiency in agriculture, and it is also needed for improvement in water quality, use of nutrient accuracy, increase crop production, and land management [14]. The use of a chemical in the agricultural environment for improved crop yield, but humans ignored the climate effect and clean environment. The environment efficiency and cleanliness are necessary for rice production by using five different varieties of rice from Gujranwala, Sheikhupura, Sialkot, Hafizabad, and Jhang districts in 2006. It saves Rs. 297 per acre and 1307.3 million save from total production of rice crops based on all districts cropland of Punjab,
Pakistan [15]. In [16], Pakistan is the largest rice exporter country in the international economic market. Poisson Pseudo Maximum Likelihood (PPML) technique was found out Pakistan is a high level competitive of rice export to over 144 countries in the international market (2003–2016). Pakistan was got a high level of competitive profit by using Revealed Comparative Advantage (RCA) and RCA# (Latest Version) method, and it describes the better results are 66.4% (RCA) and 4.27% (RCA#) between 2003 to 2016 years. In Pakistan, two different approaches were adopted for the rice transplanting based on Direct Seeds (DRY Procedure) and Paddle (Conventional Procedure) in Punjab. Kharif Cycle (2013–14) were collected 300 sample datasets of rice farmers by using management practice techniques at five different districts of Punjab namely; Jhang, Gujranwala, Sheikhpura, Hafizabad, and Vehari. The overall accuracy is 14% (DRY Procedure), and it is also increases the crop yield and efficiency of farmers [17]. The author [18], Randomized Complete Block Design (RCBD) technique was developing on the directed rice seeding process from Rice Research Institute, Pakistan (Kala Shah Kaku, 2003–2004). Fifteen fine and medium varieties of grain rice were used for the experimental process dated 23–06-2003 and 11–06-2004. The result is shown better paddy yield under RCBD by using KS-282, IR-6, NIAB-IR9, Supper Basmati, PK-5261-1-2-1, Basmati-2000, and 99512 rice varieties. Furthermore [5], RCBD layered approaches have been implemented on different paddy crop areas such as biological, grains, parameter, straw yield, and harvest index. Data is recorded after 30, 45, 60, and 75 days of the transplanting procedure. ACI Hybrid2 rice variety was achieved the highest 42.07 (Harvest Index) growth in the plant as compared to all other six paddy crop varieties namely; BRRI Dhan48, BRRI Hybrid Dhan1, BRRI Hybrid Dhan2, Jagoron, and Panna1.

Rice grain quality was playing an important role in the agricultural cropland of Pakistan. The total 475 (Accessions) sample datasets were collected from local rice genetic areas based on four cultivation locations or zones namely; Azad Kashmir, Baluchistan, NWFP/North Area, and Punjab. Grains seeds variation are distributed by following characters are length (6.0–10.66 mm), width (1.6–3.7 mm), breadth (1.14–2.36 mm), length to width (2.04–5.24 mm) ratio and weight (0.66–3.02 gm). The medium grain (0.4%) type is found from the NWFP/N zone [4]. Cropping System Model – Crop Environment Resource Synthesis – Rice (CSM-CERES-Rice) model was performing the better opportunity for increase crop yield under irrigation area of Pakistan. A total of 68 variant datasets were collecting from the past 35 years of weather in the city of Faisalabad Punjab, Pakistan. CSM-CERSE-Rice model was improved crop yield and increase rice production under the poor resources of farmers [19]. Deep Learning (DL) techniques were successfully present the framework for automatically classifying paddy crops stresses with field images dataset. This approach has been deployed over 30,000 field images dataset from 5 different types of paddy crops for the experimental process. DL VGG-16 model has obtained accuracy 95.08% over 6000 field images as compared trained dataset model is 92.89% over 30000 field images dataset [20]. A molecular marker is one of the most important tools, which was used for improved rice yield in Pakistan. Information on Genetic Diversity (GD) and the relationship between rice genotype researches is limited in Pakistan. Random Amplified Polymorphic DNA (RAPD) markers techniques were used for the research process on the available dataset namely; 10 traditional, 25 Japanese’s and 28 improved cultivation plants of rice crop. The overall similarity coefficient result is 0.60 to 0.74 (Jhona-349) and 0.60 to 0.76 (Swat-1) [21]. Researcher [22] explained the variation of water quality in South-Eastern Asia agricultural land. Seven organic and seven conventional lands were used for rice framing based on water purity indices during the (July to October) rainy period. The result of water quality indices affected due to long term (15 and 20 years) agriculture framing in south-eastern Asia and water NO3 need more attention in feature for rice crop framing. Different varieties of rice crops are cultivated in different countries. Rice automated classification system is a time and cost-saving procedure, which was deployed with the Artificial Intelligence System (AIS). The Least Significant Difference (LSD) method was evaluating the better accuracy by using the rice variety comparison process. The overall accuracy of the result is 89.2% (Paddy), 87.7% (Brown Rice), and 83.1% (White Rice) by using discriminant analysis (DA) [23]. Furthermore, in [24], rice is a major crop that’s cultivating in agricultural areas or countries, and it is also very helpful in economic growth. Rice is the best alternative source of food and nutrition. Seven
different morphological features were employing on the available dataset. Machine learning technique obtained best results are 92.49% (LR), 92.86% (SVM), 92.49% (DT), 92.39% (RF), 91.71% (NB) and 88.58% (K-NN) respectively. Researcher [25], A Neural Network (NN) approach was deployed for cultivation process based on individual and group feature dataset. Individual dataset classifier achieved different accuracy results are 94% (AT307 Rice), 98% (BG250 Rice), 84% (BG358 Rice), 100% (BG450 Rice), 94% (BW262 Rice), 68% (BW267 Rice), 98% (BW361 Rice), 94% (BW363 Rice) and 94% (BW364 Rice) respectively, but overall combine feature classification accuracy is 92%. In [26], Rice is export at a huge level around the world, and it is also a water-consuming crop. The experimental procedure has been conducting continuously for two years in china by using four irrigation methods are shallow water irrigation (FSI), wet shallow irrigation (WSI), controlled irrigation (CI), and rain-catching and controlled irrigation (RCCI). RCCI irrigation technique is more beneficial in Lianshui and similar areas by using a comparative analysis report. The rice roots are the major part of crops in the cultivation process, which play an important role in increasing the rice yield due to under water acquisition [27]. The author [28], Rice quality measurement research institute has been increasing in his study area worldwide. Rice chalkiness is one of the oldest methods adopt during traditional crop cultivation with naked eyes. Microcomputer Tomography (Micro-CT) X-Ray represents a high-resolution scan for analyzing the 60 gains laid in cubic shaped areas. Micro-CT achieved the best result in rice quality identification. In [29], the System of Rice Intensification (SRI) was identifying the impact of crop yield by using irrigation water. Tainan11 (TN11) and Tidung30 (TD30) rice varieties were used for the experimental process with different time intervals and showed. The overall accuracy results were based on three or seven days’ intervals are 55% and 74% of water-saving, The TD30 variety was performed better result and reduced 30.29% of water. Furthermore, described in [30], SRI provided a better crop cultivation process as compared to standard management procedures. Early maturity of rice variety (V1 to V16) was taking for the experimental process. SRI describes the maximum grains found biologically from V12 (13.2 t/ha). In this research [31] the varieties have been arranged into two main groups such as groups A and B (Cluster I and II). It is observing that the maximum dissimilarity 133.0 and the minimum 43.1. The production of rice is most affected due to impurity or shortage of water in Asia. Rice cultivation and novel technology have introduced a system to increase water availability. The integrated system introduced a technique for improved water quality and decrease water use in the crop field. These technology systems provided a facility that accesses scale levels from field-to-plant and agro-logical areas. The purity of water is most important for increasing food production on the available field level and grain quality [32]. Synthetic Aperture Radar (SAR) time series methodology was used as the best census survey of rice as a compared field study. C-band radar data study has been evaluated for rice monitoring. Long-short term memory (LSTM) and Bidirectional LSTM (Bi-LSTM) model results are compared with four different Machine Learning classifier such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (K-NN), and Normal Bayes (NB) with the help of sentiment-1-time series method. The overall accuracy is 0.98 (ML) and significant of 0.05 (LSTM) [33]. In [34], Genome-Wide Association Study (GWAS) was selecting the population by dividing the result into ten partitions accordingly; two different approaches adopted are Fixation Index (FST) and Quantitative Trait Nucleotides (QTNS) polymorphism.

International Rice Research Institute (IRRI) was introduced high-quality rice seed between 1960 to 1970. The farmers achieve a huge advantage during to rice cultivation process, which increases the crop yield, and minimize the production cost. In 1985, IRRI has introduced a new verity of rice seed is IR64. IR64 were achieved better accuracy results during the past ten years of rice cultivation, and it also increases the food material for the new generation [35]. Water purity and quality were the main elements for crops and plant yield. Agricultural Policy/Environmental Extender (APEX) model is deployed around all areas, where minimizes the issue of rice cultivation and water impurity [36]. In [37], every day the agriculture field improved its progress and plant growth by using virtual breeding. The breeding level was to increase rice production by employed less manpower. Supergene software used for the plant breeding process in Delphi version 5 and this methodology is useful for plant breeding
management. Furthermore, in [38], Rice production faced different hurdles due to water use efficiency (WUE) and produce grain yield. Harvest Index (HI) builds a change in the crop cultivation process. Every grain increases its growth production through WUE and crop management in the harvest index of rice. The author [39] describe everyday human was facing different food cultivation challenges. Molecular-level is one of the most beneficial characteristics in plant framing for recognized the food challenges. Single Sequence Repeat (SSR) markers (types are Common, Rare, and Unique alleles) methodology was adopted for the molecular diversity of Malaysian rice varieties. Bio-geo-chemical cycling was needed good information for cropland by using a field-based agricultural survey dataset. Country-level agricultural survey statistics is to examine sown and cropland area total 17 main crops in 1990, which based on optical remote sensing (Land Sat) 0.50 resolution map (1995–1996) datasets. The overall accuracy results are 75% (Paddy Land) and 56% (Two-Piece Cultivation) every year. An estimated 0.30 million (KM2) area was cultivated paddy crop in china [40]. Biodiversity variation is recognizing the following four organization levels such as genetic, ecosystem, species, and land space. Biodiversity secures our ecosystem and cleans our climate [41]. In [42], Food requirements increase day-by-day for the causes of world population growth. An automated system is a more effective and time-consuming process for rice classification. The overall accuracy is 90.61%, 82.71% (Sub-group 1 and 2), and 83.9% (collective data) on based SVM methodology, but the Inspection ResNetV2 model achieved the best result with deep learning technique is 95.15%. As above discussed literature, it is obvious that [43] computer vision is playing vital role in agriculture for early detection of fruit and leaf diseases [44]. In [45] apple disease namely apple scab, brown spot and apple cedar were correctly recognized using the deep learning inception V3 model and overall 97% accuracy was achieved. For this study, we did not employ deep learning models due to usage of small and trivial dataset of rice varieties and conventional machine learning model provided quite satisfactory results in term of accuracy, time and space as compared to deep learning models [46].

Materials and methods

In the earlier discussion, this experimental process adopted the MV classifier of different varieties of rice namely; Kachi_Kainat, Kachi_Toota, Kainat_Pakki, Super_Basmati_Kachi, Super_Basmati_Pakki, and Super_Maryam_Kainat (A1, A2, A3, A4, A5, A6). The complete dataset process has been proceeding in openly climate at the department of computer science the Muhammad Nawaz Shareef University of Agriculture Multan (MNS-UAM), Pakistan is locating at 71° 27′ 21″ (East) longitude and 30° 10′ 51″ (North) latitude [47] and Tehsil Ahmad Pur East is locating at 71° 16′ 61″ (East) longitude and 29° 5′ 13″ (North) latitude, in District Bahawalpur (Area is 6,857 km2). Rice grain image was capture with 13 megapixels (MP) camera by using OPPO F3 cell phone between 11:00 (A.M) to 12:00 (P.M) according to Pakistani time zone. The rice image resolution is a variation in pixel size with Joint Photographic Expert Group (JPEG) frames [48]. In this study process, about 1800 healthy grains were arranged for each rice variety. All images of rice grain were captured at 1.5 feet (Height Distance) by using still based cell phone camera, and it also eliminates all sun shadow images. All images dataset was captured between noon (11:00 A.M to 12:00 P.M) times.

Finlay, a colored image-based dataset of 100 × 6 = 600 were collecting with high-quality rice of variant image in pixel size was developed with 32-bit Joint Photographic Expert Group (JPEG) image format. These available datasets were used for our experimental process. All grain samples of six different categories of rice image, which is shown in Figure 1. A white color paper was used for the high quality of image.

The image segmentation technique used for the purity of the image dataset on all JPEG images is also eliminating irrelevant and noisy information. Initially convert all image data into gray-scale, it is also implanting Median filter for more appropriate rice image classification [49]. Furthermore, all the image datasets have been pre-processed by using free image convert and resize software. All the 600 (100×6) colored image resolution was converted into (512 × 512) pixel size and 8-bit gray-scale level; it
is also saved in JPEG format. There are six different non-overlapping Region of Interest (ROI’s) was created on each image of rice. The total evaluated 154,800 (3600 × 43) multi-feature datasets have been captured on each ROI, which is shown in Figure 2.

**Proposed technique and methodology**

A complete discussion about this experimental methodology describes in the following steps. At the start, images were captured by employing preprocessing techniques. In this study, the Range Oriented Pixel based Resolution (R_O_P_S) Algorithm has been used for image segmentation, which is shown in Figure 3 [2]. This experimental process divides into three different steps; Initially, the R_O_P_S
methodology has been calculated the total pixels (T_P) range for some particular area of the threshold by describing a total arrangement of clustering based on the back end of the image. The next step, in this level of threshold T_P, was deployed on the based value of pixels by using the total area of an image. In the end, the calculate rice pixel (R_P) value of homogeneity was identifying is not equal than T_P. We evaluate this region of fore-ground pixels and grow up the total area of clustering on the base of ROI’s.

**Multi features extraction procedure**

Feature segmentation was depending on the researcher’s methodology and techniques [50]. The process of evaluating a multi-features dataset by using rice images, the 28 binary (B) feature based on the 10-pixel distance between width and height of an image, 5 first order of histogram (H) and 10-second order of texture (T) multi-features were describing on each image of rice [51]. Furthermore, the 43 (B, H, and T) features were calculated on each ROI’s (or Sub-Images), and a total calculated of 154,800 (43 × 3600) multi features on available datasets. In this experimental process were employing computer vision of image processing (CVIP) and WEKA (version 3.8.1) software tool on Intel (R) core i3, 2 GB-RAM, 2.4 Giga Hertz (GHz) processor with 64-bit window 7 operating system [49]. B, H, and T based multi-features were describing a detailed discussion below section.

**Binary**

This method also evaluated the shape or design and identified object placement in image segmentation by using the Axis of last-second moments, Area of the center object, Area, Number of Euler, and object projection. Binary features have the capacity to empower the computer vision approach because these features do not rely on pixel values or relative neighboring pixels. Binary features reduce the extra load of computation, complexity and provides actual information of the image for better analysis and classification. The k-th is an area of the object (P_k), which is defined in equation 1.
\[ P(k) = \sum_{a=0}^{\text{width}-1} \sum_{b=0}^{\text{height}-1} K_k(a, b) \]  

**Histogram or first order of statistics histogram feature**

Histogram or first-order features are calculated the individual value of a pixel-based on intensity. A histogram method is a set of adjacent momentum angles based on the x-axis with the proportional area and class mark of frequency. Histogram probability \( P(g) \) has been used for collecting data or information with available pixels and gray-scale in image processing. That is shown in equation 2.

\[ P(g) = \frac{K(g)}{M} \]  

Here \( K(g) \) calculates all gray-scale levels of \( g \), and \( M \) is the total calculated value of the pixel in an image. Histogram features also calculated are standard deviation (SD), Energy, Mean, Entropy, and skewness.

**SD** identified the total contrast of the object in image processing. SD is shown in equation 3.

\[ \sigma_b = \sqrt{\sum_{b=0}^{M-1} (b - \bar{b})^2 M(b)} \]  

The image bright and dark area is measured and identifies through High and Low mean average values. **Mean** define in equation 4.

\[ \bar{s} = \sum_{c=0}^{M-1} c M(c) = \sum_{a} \sum_{b} J(a, b) M \]  

Here \( J \) is measure the all number of gray level from 0 to 255, is also \( a \) and \( b \) represent Rows and Colum pixel.

**Skewness** is a static measurement procedure based on central values (Mode, Mean, and Median). Two main parts which are implementation in skewness distribution are positive (tail-to-right) and negative (tail-to-left). Skewness has shown in equation 5.

\[ \text{Skew} = \frac{1}{\sigma^3} \sum_{c=0}^{M-1} (c - \bar{c})^3 M(c) \]  

In this method is identified the total calculated number of bits by using image code data. **Entropy** has shown in equation 6.

\[ \text{Entropy} = - \sum_{c=0}^{M-1} M(c) \log_2[M(c)] \]  

Gray level is identified with **energy** equation which shown in equation 7.

\[ \text{Energy} = \sum_{c=0}^{M-1} [M(c)]^2 \]
**Texture or second order of statistics histogram feature**

Histogram features are also called second order statistical features, and based on Gray Level Co-occurrence Matrix (GLCM) with four (0, 45, 90 and 135 degrees) dimensions. Rows and columns (XY-Coordinates) were played an important role in selecting an object of image texture processing. Five different statistical histogram features were evaluated such as inverse difference, energy, inertia, correlation, and entropy. Local homogeneity is also called inverse difference of image, which is shown in equation 8.

\[
\text{InverseDifference} = \sum_a \sum_b \frac{K_{ab}}{|a - b|}
\]  

(8)

The identification of homogeneity or smoothness between image gray levels is called energy defining in equation 9. Here \(K_{ab}\) are examining the distribution data values of pixels in matrix form.

\[
\text{Energy} = \sum_a \sum_b (K_{ab})^2
\]  

(9)

The image object contrast is measure by inertia, which defines in equation 10.

\[
\text{Inertia} = \sum_a \sum_b (a - b)^2 K_{ab}
\]  

(10)

**Correlation** is to identify the pixels value from specific pixels distance, which defines in equation 11. Here \(\mu_m\) and \(\mu_n\) are define the value of mean from a and b respectively.

\[
\text{Correlation} = \frac{1}{c d c^d} \sum_c \sum_d (c - \mu_c) (d - \mu_d) K_{cd}
\]  

(11)

The Second-order statically histogram technique is described for probability measurement based on the Gray Level Co-occurrence Matrix (GLCM).

**Entropy** is describing the content information about an image object. Entropy has shown in equation 12.

\[
\text{Entropy} = - \sum_c \sum_d K_{cd} \log_2 K_{cd}
\]  

(12)

**Feature optimization and selection procedure**

According to above explanation, the 43 (B, H and T) multi-features were extracted on each ROI’s by used different rice image, and managed 154,800 a huge image size of multi features. A huge multi-feature handling process was very critical. So, this problem has been re-solved by using feature optimization technique. The correlation based feature selection (WEKA) method with Best First Search algorithm was deployed on huge multi-feature dataset, and it’s evaluating 11 optimized features on each rice image, which is shown in Table 1. It was calculated total 154,800 (3600 × 43) features and its optimized multi-features are 39,600 (3600 × 11). This optimized features dataset was used for

| Sr. No | Features                        | Sr. No | Features                        |
|--------|--------------------------------|--------|--------------------------------|
| 1.     | Histo_Mean                      | 7.     | Texture_entropy_range           |
| 2.     | Histo_Standard_Deviation        | 8.     | Inertia_average                 |
| 3.     | Histo_Skew                      | 9.     | Correlation_average             |
| 4.     | Histo_Energy                    | 10.    | Inverse_diff_average            |
| 5.     | Texture_energy_average          | 11.    | Centroid (row_column)           |
| 6.     | Texture_energy_range            |        |                                |

*Table 1*. Best First Search Algorithm Based Feature Optimization.
employing different MV classifier. In this practical process, the 10 K-fold cross validation technique has been employing for eliminating the complexity of testing and training ratio. This experimental procedure has been repeated 15 times, and evaluates the result. This experimental or identification framework of rice varieties was described in Figure 4.

**Results and discussion**

The summary of different features classification for rice varieties have been employing variant MV techniques that are shown in Tables 2 and 4. MV classifier results discuss below:

**Meta based classifier**

At the start, Meta based classifier has been employing for the experimental process name Meta Classifier via Regression (MCR), Meta Bagging (MB) and Meta Attribute Select Classifier (MAS-C). The overall accuracy of MCR gives a better result of 97.0278% as compared MB 96.25% and MAS-C 95.1667% that is shown in Table 2.

![Figure 4. Multi feature classification framework of rice varieties.](image)
Confusion matrix

Meta classification via regression (MCR) is shown the highest output values of the confusion matrix diagonally in Table 3.

Tree based classifier

A tree-based MV classifier has been deploying for further better accuracy results. It is also observed that the tree classifier has shown better output accuracy results as compared to Meta classifiers. The LMT-T has shown better result accuracy is 97.3889% as compared to T-J48 95.6944%, which result shown in Table 4.

Confusion matrix

The confusion matrix of the LMT-T classifier has shown the highest values in the diagonal arrangement in Table 5.

Table 3. Confusion Matrix Table for Meta Classification via Regression.

| Classes | A1  | A2  | A3  | A4  | A5  | A6  |
|---------|-----|-----|-----|-----|-----|-----|
| A1      | 582 | 7   | 4   | 4   | 1   | 2   |
| A2      | 5   | 592 | 1   | 0   | 1   | 1   |
| A3      | 3   | 2   | 576 | 2   | 1   | 16  |
| A4      | 2   | 0   | 2   | 591 | 1   | 4   |
| A5      | 0   | 0   | 2   | 1   | 595 | 2   |
| A6      | 8   | 1   | 29  | 4   | 1   | 557 |

Table 4. TREE Based Classification OA Dataset.

| TREE Classifiers | TP Rate | Kappa Statistics | MAE | RMSE | FP Rate | TNI | ROC | Time (sec) | OA    |
|------------------|---------|------------------|-----|------|---------|-----|-----|------------|-------|
| LMT Tree         | 0.974   | 0.9687           | 0.0115 | 0.0879 | 0.005 | 3600 | 0.995 | 15.37     | 97.3889% |
| J48 Tree         | 0.957   | 0.9483           | 0.0173 | 0.118 | 0.009 | 3600 | 0.979 | 0.13      | 95.6944% |

Table 5. Confusion Matrix Table for LMT-T Classification.

| Classes | A1  | A2  | A3  | A4  | A5  | A6  |
|---------|-----|-----|-----|-----|-----|-----|
| A1      | 590 | 5   | 1   | 2   | 0   | 2   |
| A2      | 3   | 596 | 0   | 0   | 0   | 1   |
| A3      | 4   | 0   | 572 | 3   | 2   | 19  |
| A4      | 0   | 0   | 5   | 592 | 0   | 3   |
| A5      | 0   | 0   | 2   | 0   | 594 | 4   |
| A6      | 4   | 0   | 31  | 3   | 0   | 562 |

Table 6. Comparison Table of all Machine Vision Classifiers Overall Accuracy Result.

| MV Classifiers             | TP Rate | Kappa Statistics | MAE | RMSE | FP Rate | TNI | ROC | Time (sec) | OA    |
|----------------------------|---------|------------------|-----|------|---------|-----|-----|------------|-------|
| LMT Tree                   | 0.974   | 0.9687           | 0.0115 | 0.0879 | 0.005 | 3600 | 0.995 | 15.37     | 97.3889% |
| Classification via Regression | 0.970   | 0.9643           | 0.0349 | 0.1037 | 0.006 | 3600 | 0.996 | 0.73      | 97.0278% |
| Bagging                    | 0.963   | 0.955            | 0.0318 | 0.1052 | 0.008 | 3600 | 0.997 | 0.27      | 96.25%  |
| J48 Tree                   | 0.957   | 0.9483           | 0.0173 | 0.118 | 0.009 | 3600 | 0.979 | 0.13      | 95.6944% |
| Attribute selected Classifier | 0.952   | 0.942            | 0.0184 | 0.1221 | 0.010 | 3600 | 0.979 | 0.17      | 95.1667% |
Comparison of all machine vision classifiers

As the above discussion, the overall result accuracy sequence of available MV classifiers such as LMT Tree (LMT-T), Meta Classifier via Regression (MCR), Meta Bagging (MB), Tree J48 (T-J48) and Meta Attribute Select Classifier (MAS-C). LMT-T and Classification-via-Regression achieved better accuracy results is 97.3889% and 97.0278% as compared to all other classifiers that are shown in Table 6 and Figure 5 from high to low accuracy result sequence.

Confusion matrix of overall result accuracy MV classifier

A confusion matrix graph between all available machine vision classifiers such as LMT Tree (LMT-T), Meta Classifier via Regression (MCR), Meta Bagging (MB), Tree J48 (T-J48) and Meta Attribute Select Classifier (MAS-C) which shown in Figure 6. In Figure 6 describe the confusion matrix of available all six different rice verity is Kachi_Kainat, Kachi_Toota, Kainat_Pakki, Super_Basmati_Kachi, Super_Basmati_Pakki, and Super_Maryam_Kainat that is representing with A1, A2, A3, A4, A5, and A6. A comparison between the existing and proposed approaches has been showing in Table 7.

![Figure 5](image1.png)

**Figure 5.** Comparison Graph of all Machine Vision Classifiers Overall Accuracy Result.

![Figure 6](image2.png)

**Figure 6.** A Comparison Graph of Confusion Matrix on Six Different Rice Verities (A1 to A6)
### Table 7. Comparison between Existing and Proposed Approach.

| Ref. No | Dataset                      | Features/Models          | Algorithm/Classifier                | Overall Accuracy |
|---------|------------------------------|--------------------------|------------------------------------|-----------------|
| [33]    | Sentinel-1 Radar Data        | LSTM and Bi-LSTM         | SVM, Random Forest, KNN, and Normal Bayes | 0.98            |
| [42]    | Field Based Image Data       | InspectionResNetv2, VGG16 and VGG19 | SVM, KNN, LR, LDA                   | 95.15%          |
| [20]    | Field Based Image Data       | VGG16, ResNet, LetNet    | DLNN                               | 95.08% & 92.89% |
| [23]    | Rice Cultivars Data          | ANN, LSD                 | PCA, Discriminant Analysis         | 89.2%           |
| [24]    | Field Based Image Data       | Morphological            | LR, MLP, SVM, RF, KNN, NB          | 92.86%          |
| [21]    | Pakistani Rice Dataset       | Cluster Analysis DNA     | Molecular RAPD                      | 0.74 and 0.76   |
| [25]    | Pakistani Rice Dataset       | Texture                  | AT307, BG250, BG358 etc.           | 92%             |
| Proposed Approach | Pakistani Rice Dataset | Binary, Histogram and Texture | LMT Tree and Meta Classifier-via-Regression | 97.3889% and 97.0278% |

### Conclusion

Pakistan is an agricultural country. Many varieties of rice crops were cultivated in Pakistan. In this study, the experimental process was an implementation for the classification of six rice varieties (A1 to A6) by using a multi-features dataset. For the experimental process, five different machine vision (MV) classifiers such as LMT Tree (LMT-T), Meta Classifier via Regression (MCR), Meta Bagging (MB), Tree J48 (T-J48), and Meta Attribute Select Classifier (MAS-C) were employing successfully. Different techniques were deployed for feature optimization on the available dataset. All the MV classifiers describe the efficiency of the result but the LMT-T classifier gives better maximum overall accuracy (MOA) result that is 97.3889%. The achieved MOA result showed that the presented methodology is beneficial, and it is an implementation in real-life environment application. In the feature, we employed more texture and multi-features with additional instances of an attribute on the same available rice image dataset. Furthermore, in feature, we implement the same experiment process on rice field crops datasets for identification of rice crop diseases by using easily and cost-saving method, without using costly and difficult equipment system.

### Acknowledgments

The authors would like to especially thank for Department of Computer Science; MNS-UAM, Multan Punjab, Pakistan), & Department of Information Technology; The Islamia University of Bahawalpur Pakistan) for their technical support during this experimental process. Especial thanks for co-operation in research College of Computer Science & Software Engineering, Shenzhen University China. Dr. Salman Qadri supervised this study process, provides all technical support, and analyzed all datasets. Tanveer Aslam performed all the fieldwork of data collection, wrote the manuscript, and description of the available dataset. The authors declare no conflict of interest.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

"This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors."
References

[1] Firoz Shamsi, A.; Rizwana Bashir, L. A.; Ashraf, H. Food Insecurity of Rice in Pakistan. *IPB J.* January 2014, 6(2): 133–145.

[2] Qadri, S., Furqan Qadri, S., Husnain, M., Saad Missen, M., Khan, D., Muzammil-Ul-Rehman, Razzaq, A. and Ullah, S. Machine Vision Approach for Classification of Citrus Leaves Using Fused Features. *Int. J. Food Prop.* 2019, 22(1), 2071–2088. December.

[3] Ahmad, I. Pakistan Economic Survey 2019-20. Economic Adviser’s Wing, Finance Division Government of Pakistan, Islamabad, Economic Survey 2019-20, June 2020.

[4] Uddin Siddiqui, S.; Kumamaru, T.; Satoh, H. Pakistan Rice Genetic Resources-i: Grain Morphological Diversity and Its Distribution. *Pak J Bot.* 2007, 39(3), 841–848.

[5] Junnat, T.; Haque, M.; Hossain, M.; Hosain, M.; Nizam, R. and Islam, M. Evaluation of Growth Performance of Some Rice Varieties in Relation to Their Economic Yield. *Asian Res J Agric (ARJA).* 2017 December, 7(2), 1–8. doi: 10.9734/ARJA/2017/37394.

[6] Giraud, G. *ENITA de Clermont, “BASMATI RICE IN PAKISTAN,” FOOD QUALITY AND SAFETY.* France, France: WPS Case studies FP6-2003-SSP-3 – 006522. September 2007.

[7] Hussain, I., Shah, H., Khan, M. A., Akhtar, W., Majid, A. and Mujahid, M. Y. Productivity in Rice-Wheat Crop Rotation of Pakistan: An Application of Typical Farm Methodology. *Pak J Agric Res.* 2012, 25(1), 1–13.

[8] Rehman, A.; Jingdong, L.; Chandio, A. A.; Shabbir, M.; Hussain, I. Economic Outlook of Rice Crops in Pakistan: A Time Series Analysis (1970-2015). *Springer Open Financ Innovation J.* July 2017, 3(13): 1–9.

[9] Alaa, S.; Waisy, A.; Qahwaji, R.; Ipson, S.; Al-Fahdawi, S. A Multimodal Deep Learning Framework Using Local Feature Representations for Face Recognition. *Mach Vision Appl.* September 2017, 29(1): 35–54.

[10] Calingacion, M., Fang, L., Quiatchon-Baeza, L., Mumm, R., Riedel, A., Hall, R. and Fitzgerald, M. Delving Deeper into Technological Innovations to Understand Differences in Rice Quality. *Springer Open Rice J.* 2015, 8(6), 1–10. January.

[11] Abdullah, M.; Jia, L.; Ghazanfar, S.; Ahmed, J.; Khan, I. Growth and Instability Analysis of Rice Production and Export of Pakistan. *Eur J Econ Stud.* November 2015, 11(1), 4–15. DOI: 10.13187/es.2015.11.4.

[12] Li, K., Wang, D., Gong, L., Lyu, Y., Guo, H., Chen, W., Jin, C., Liu, X., Fang, C. and Luo, J. Comparative Analysis of Metabolome of Rice Seeds at Three Developmental Stages Using a Recombinant Inbred Line Population. *Plant J SEB.* 2019, 482, 908–922. July.

[13] Meng, W., Xu, L., Du, Z., Wang, F., Zhang, R., Song, X., Lam, S., Shui, G., Li, Y. and Chye, M. RICE ACYL-COA-BINDING PROTEIN6 Affects Acyl-CoA Homeostasis and Growth in RICE. *Springer Open Rice J.* 2020, 13(75), 1–17.

[14] Timsina, J.; Connor, D. J. Productivity and Management of Rice-wheat Cropping Systems: Issues and Challenges. *Elsevier Field Crops Res J.* January 2001, 1(69), 93–132. DOI: 10.1016/S0378-4290(00)00143-X.

[15] Abdullah, Kouser, S. and Mushtaq, K. Environmental Efficiency Analysis of Basmati Rice Production in Punjab, Pakistan: Implications for Sustainable Agricultural Development. *Pak Dev Rev March.* 2010, 49(1), 57–72. DOI: 10.30541/v49i1pp.57-72.

[16] Irshad, M. S.; Xin, Q.; Arshad, H. Competitiveness of Pakistani Rice in International Market and Export Potential with Global World: A Panel Gravity Approach. *Cogent Econ Finance J.* July 2018, 669(90): 1–18.

[17] Shaheen, S.; Fatima, H.; Khan, M. A. Technical Efficiency Analysis of Rice Production in Pakistan under Dry and Puddle Conditions: A Case Study of Selected Districts of Punjab Province, Pakistan. *Sarhad J Agric.* September 2017, 33(3), 447–458. DOI: 10.17582/journal.sja/2017/33.3.447.458.

[18] Ali, R. I., Awan, T. H., Manzoor, Z., Ashraf, M. M., Safdar, M. E. and Ahmad, M. Screening of Rice Varieties Suitable for Direct Seeding in Punjab. *J Anim Plant Sci.* 2007, 17(1–2), 24–26. January.

[19] Ahmad, S., Ahmad, A., Ali, H., Hussain, A., Garcia y Garcia, A., Khan, M., Zia-Ul-Haq, M., Hasanuzzaman, M. and Hoogenboom, G. Application of the CSM-CERES-Rice Model for Evaluation of Plant Density and Irrigation Management of Transplanted Rice for an Irrigated Semi-arid Environment. *Springer Irrig Sci J.* 2012 February, 31 (491), 491–506. doi: 10.1007/s00271-012-0324-6.

[20] Basavraj, S. A.; Malvade, N. N.; Palaiiah, S. Deep Learning Approach for Recognition and Classification of Yield Affecting Paddy Crop Stresses Using Field Images. *Artif Intell Agric KeAi Publ J.* April 2020, 3(1): 12–20.

[21] Rabbani, M. A.; Pervaiz, Z. H.; Masood, M. S. Genetic Diversity Analysis of Traditional and Improved Cultivars of Pakistani Rice (*Oryza Sativa* L.) Using RAPD Markers. *Plant Biotechnol Electron J Biotechnol.* July 2008, 11(3): 1–8.

[22] Shi, D., Dari, B., Yan, Z., Sharma, D., Pathak, H., Sharma, O. and Nain, L. Assesment of Water Quality in IHasanuzzamanando-Gangetic Plain of South-Eastern Asia under Organic Vs. Conventional Rice Farming. *MDPI Water J.* 2020, 12(960), 1–9 March.

[23] Abbaspour-Gilandeh, Y., Molaei, A., Sabzi, S., Nabipur, N., Shamshirband, S. and Mosavi, A. A Combined Method of Image Processing and Artificial Neural Network for the Identification of 13 Iranian Rice Cultivars. *MDPI Agron J.* 2020, 117(10), 1–21. January.

[24] Cinar, I. and Koklu, M. Classification of Rice Varieties Using Artificial Intelligence Methods. *Int J Intell Syst Appl Eng (IIJISAE).* September 2019, 7(3), 188–194. DOI: 10.18201/ijisae.2019355381.
[48] Aslam, T., Qadri, S., Shehzad, M., Qadri, S. F., Razzaq, A., Muhammad, S. S., Ahmad, S., Nawaz, S. A. and Ahmad, N. Emotion Based Facial Expression Detection Using Machine Learning. *Life Sci. J.* 2020, 17(8), 35–43. August.

[49] Trokielewicz, M., Czajka, A. and Maciejewicz, P. Post-mortem Iris Recognition with Deep-learning-based Image Segmentation. *Image and Vision Computing*, 94. December, 2019, 1–11.

[50] Aslam, T., Qadri, S., Saher, N., Nawaz, S. A., Shehzad, M., Qadri, S. F., Razzaq, A., Shahzad, F. and Khan, D. M. A Spatial Model of K-Nearest Neighbors for Classification of Cotton (Gossypium) Varieties Based on Image Segmentation. Research Journal of Computer Science and Information Technology, Lahore Garrison University, 2021, 5(1), 24–39. Jan-March.

[51] Korbeogo, G. Reinventing New Systems of Crop Production in Time of Agro-ecological Change in Burkina Faso. *Afr. J. Environ. Sci. Technol.* May 2015, 9(5), 464–472. DOI: 10.5897/AJEST2012.046.

[52] Aslam, T.; Qadri, S.; Shah, A. N.; Shehzad, M. (2021, April) Muhammad Nawaz Shareef University of Agriculture, Multan. [Online]. https://www.mnsuam.edu.pk/