IDENTIFICATION OF RICE PURITY LEVEL FROM MIXED RICE VARIETIES USING DEEP LEARNING

Khalid Abbas [1], Ayesha Hakim * [2], Nasir Nadeem [3], Adnan Altaf [4] and Hafiz Muhammad Rizwan Iqbal [5]

[1,5] Researcher, [2] Assistant Professor, [4] Lecturer, Department of Computer Science, [3] Professor, Department of Agribusiness & Applied Economics, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan

*Corresponding author's email: ayesha.hakim@mnsuam.edu.pk

ABSTRACT

The current study was conducted in Multan, Pakistan to investigate an automated appearance based system for purity level identification of seven common rice (*Oryza sativa L.)* varieties from mixed rice grain samples. Adulteration is a major hurdle that affects rice export in Pakistan that refers to the mixing of premium rice grain varieties with the low grade rice grains to be marketed at a high cost. This study was based on the dataset collected from Rice Research Institute, Kala Shah Kaku during the study year 2018-2020. Three Pakistani premium rice varieties (*Basmati Shaheen, Basmati Super, and Basmati Pak*) were mixed with four low quality varieties (*Basmati 198, Basmati 2000, Basmati 370 and Basmati 385*) in weight ratios of 10%, 15%, 20%, 25% and 30%. Classification and recognition of purity level of Basmati rice achieved average accuracy of 89.88% using convolutional neural network. The proposed system has the potential to be used at a commercial scale to test the purity level of exported rice.

KEYWORDS: *Oryza sativa*; purity level; deep learning; convolutional neural network; adulteration; Pakistan

INTRODUCTION

Rice (*Oryza sativa L.*) is Pakistan's second most important crop after wheat that can bring economic prosperity to producers and generate billions of rupees of income by exporting to other countries. It adds 0.6% to overall GDP and 3.0% to the Pakistan's agricultural economy (GoP, 2019). It is cultivated in several areas across Pakistan, including Punjab, where it is sown in Gujranwala, Sheikhupura, Wazirabad, Sialkot, Faisalabad, Sargodha, Kasur, and Gujrat districts. In Sindh, Thatta, Shikarpur, Jacobabad, Dadu, Larkana, Badin districts are important for rice farming (PARC, 2021). In Pakistan, various rice varieties are grown including *IRRI, Kashmir Nafes, DR82, Sada Hayat, Swat, Ks-282, Shadab, Nayab,* and *Basmati*. *Basmati* is the premium quality Pakistani rice having good quality characteristics. Various rice varieties are cultivated in Pakistan, but there are limited numbers of varieties that are preferred for export. Rice grain quality can be damaged before and after the cultivation processes. After cultivation, cracking of rice grain occurs in the field due to variation in temperature and moisture level that affect the grain quality. Drying is another type of post-cultivation process that affects the grain and becomes a cause of breaking of rice grains. The ratio of breakage of the chalky rice grain is higher than that of the transparent rice grain because the chalky rice grain(opaque area of rice grain) is softer than the transparent rice grain. Due to cracking head, rice yield and quality of cooking rice grain is reduced that makes it inappropriate for export (Kaur and Singh, 2013).

Mostly damaged or low grade quality grains are mixed with high quality grains to earn optimal price, this process is known as adulteration. Premium rice varieties are mixed with low grade look-alike rice varieties to get a high price of premium rice variety (Fayyazi *et al.*, 2017). Rice importing countries allow 5-8% of the adulteration level caused by mixing during harvesting and post-harvesting but do not allow if the limit exceeds 8% of the adulteration level (Anami *et al.*, 2019). Several methods are used for the classification of rice grain but there is no adequate technology and methods used for adulteration level recognition and classification in Pakistan. There are two ways of adulteration; one is an adulteration that involves foreign or non-quality factors in the grain sample including dirt, stones, pieces of earth, straw, stalks or straws and any other impurity. The second way of adulteration involves the mixing of low grade quality rice with good quality rice grain. Adulteration causes serious illness

if it is used as a food on daily basis and it must be identified to improve quality of grain. Automatic rice variety classification and purity level identification has been an interesting research topic in the past decade. Seven rice varieties (Basmati 2000, Chenab Basmati, KSK 133, Kissan Basmati, KSK 434, PK 1121 Aromatic, and Punjab Basmati) were identified using Convolutional Neural Network. However, they did not perform identification of mixed rice varieties (Gilanie *et al.*, 2021). Anami *et al.* (2019) identified adulteration level from rice grain varieties based on texture and colour feature. Principal component analysis (PCA) for feature reduction and selection and Back propagation Neural Network (BPNN) approach was used to classify the adulteration level that accounts for 93.31% accuracy. In the literature, a computer vision and fuzzy logic based novel technique is used to detect fraudulent labelling of rice grains, DNA and Non-DNA based approach is used with 10 gram sample and achieved more than 90% accuracy (Ali *et al.*, 2017). The MLP neural network classifier was used to classify the rice varieties in mixed lots of three and two varieties using the selected features. For three mixed varieties, 93.02% and for two mixed varieties, 96.08 % maximum accuracy was attained (Fayyazi *et al.*, 2017).

Rice quality is determined based on shape, colour, size, head rice, broken and brewers' factors. Multi-class Support Vector Machine (SVM) algorithm was used to categorize the grade of rice into premium grade A, grade B, and grade C. A total of 800 rice grains images were captured, 400 images were randomly selected for training and 400 images were selected for testing and around 86% accuracy was achieved (Kaur and Singh, 2013). Similarly, Kambo and Yerpude (2014) classified three Indian Basmati rice varieties (*Basmati Classic, Basmati Rozana, and Basmati Mini*) based on textures and structural features including major-axis length, minor-axis length, perimeter, eccentricity and area. Principal component analysis (PCA) was used for feature reduction and k-NN was used for classification with an average accuracy of 79%. It is hard to predict grain variety through simple statistical functions as the grain has several structures, textures, and colours. In the literature, Artificial neural network (ANN) was used for grain quality prediction using images captured with digital mobile camera (Guzman and Peralta, 2008; Pazoki *et al.,* 2014). Shantaiya and Ansari (2010) used ANN to classify grain variety based on structural and colour factors on a dataset of 60 image of Basmati grains captured with digital camera, achieving an average accuracy of 95% accuracy. In another article, a neural network based algorithm was used to process coloured bulk sample images of grain

varieties and extracted 150 colour features and texture features achieving 90% classification accuracy (Visen *et al.,* 2003).

Similar approach has been used by Silva and Sonnadara (2013), where neural network based algorithm was deployed to classify nine various rice varieties based on six colors, thirteen morphological elements and fifteen texture features extracted from the color images of the individual seed samples achieving 92% classification accuracy. Machine learning and image processing techniques have also been used to classify rice grain varieties and grading based on head rice and broken rice of each variety. Several algorithms were used for classification of rice grain variety but K-NN with generalization delivered maximum accuracy of 90.5% (Ozan *et al.*, 2015). Another article presents a content-based approach to classify rice grains based on features extracted through histogram. RGB, CMYK, YUV, YCBCR, HSV, HVC and YIQ color spaces were used for feature extraction. For classification 500 images were used achieving an average accuracy of 85% (Agrawal *et al.*, 2011).

An image processing algorithm was implemented for analysis and classification of rice grains in 3 categories based on real field characteristics and KNN classifiers. For each category, 30 images were tested and obtained classification accuracy in the range of 83-100% for all categories (Wah *et al.,* 2018).Research also has been carried out to differentiate between two proprietary rice species. A total of 3810 images of rice grains were taken for both species and seven morphological characteristics were obtained for each category. These features were used in the Decision Tree (DT), LR, Support Vector Machine (SVM), MLP, NB, k-NN and Random Forest (RF) machine learning approaches. Average accuracy for classification, 92.86% (MLP), 93.02% (LR), 92.49% (DT), 92.39% (RF), 92.83% (SVM), 91.71% (NB), 88.58% (k-NN) was obtained (Cinar and Koklu, 2019).

To grade rice grain quality, shape descriptors were used to determine the quantity of head rice, broken-head, and brewers based on geometric features. Color histogram was used to extract 24 color features and a probabilistic neural network (PNN) classifier was used achieving 94% accuracy (Agustin and Oh, 2008). Another research has been done to inspect the rice quality of 21 rice varieties based on shape and chalkiness. Improved multi-threshold method based on maximum entropy was used for examining rice chalkiness. Minimum enclosing rectangle method was used for assessing rice shape and system delivered satisfactory results (Yao *et al.,* 2009). In Pourreza *et al.* (2012), three Iranian rice varieties (*Shiroodi,*

Tarom, Fajr) were classified on a dataset of 666 rice seed images, 222 images of each variety using Linear Discriminant Analysis (LDA) for classification based on selected characteristics and achieved classification accuracy of 98.15%. Similarly, in (Singh and Chaudhury, 2020), authors classified four different rice varieties based on four sets of characteristics: color, morphology, texture, and wavelet of rice grain. It was found that morphological features were more suitable for rice core classification compared to other characteristics. A classifier was also tested against standard datasets from the University of California, Irvine (UCI). The results were compared with other classifiers and claimed to perform better in terms of classification accuracy (Singh and Chaudhury, 2020).

Based on the literature review, it was identified that none of the published work has presented automated purity level identification system from bulk rice sample on appearance-based characteristics of rice grains. In this paper, we are presenting a deep learning-based approach to automatically classify purity level (%) used for export quality inspection. The system has been evaluated on mixed rice grain samples of Premium Basmati Rice with low grade rice grains with promising results.

MATERIALS AND METHODS

This section presents methods and techniques used for rice grain sample collection and the proposed methodology for rice purity level identification. Graphical illustration of these stages is presented in Fig. 1.

Rice grain sample collection

The rice grain samples comprising seven pure verities (*Basmati Pak, Basmati Shaheen, Basmati* Super, Basmati 198, Basmati 2000, Basmati 370, and Basmati 385) were collected from Rice Research Institute, Kala Shah Kaku during the study year 2018-2020. All samples were sealed and labelled by rice experts. In this work, three premium Basmati rice varieties namely Basmati Pak, Basmati Shaheen and Basmati Super and four low grade rice varieties namely Basmati 198, Basmati 2000, Basmati 370, and Basmati 385 have been used for experiments. Digital Images of above-mentioned rice grain varieties have been captured using a standard digital mobile phone (Huawei Mate 10 lite 13 MP) camera that was mounted on a stand at a fixed location at a distance of 3 inches from the lens and sample. All images were captured with a black background and uniform light intensity during at noon (12:00 pm to 2:00 pm) to ensure uniform image quality.
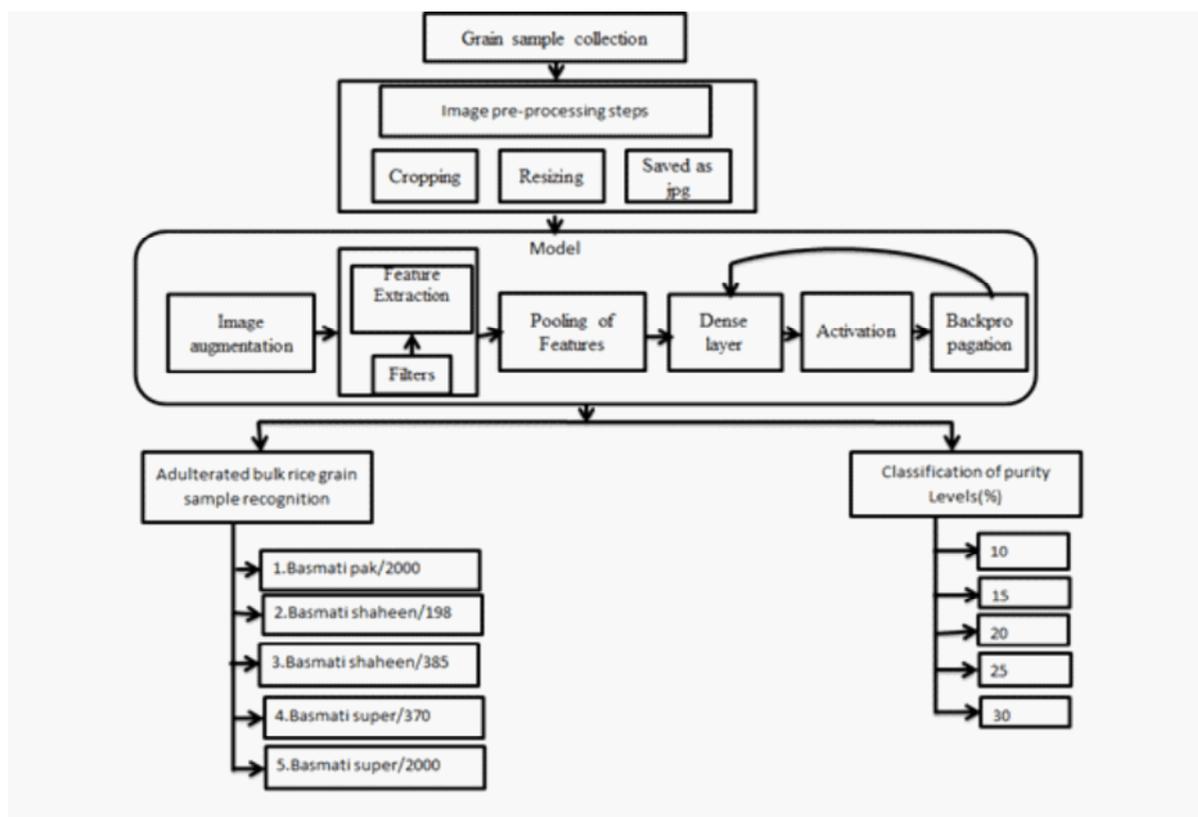

Fig. 1 Graphical description of proposed methodology

Table 1. List of premium rice varieties mixed with low-grade rice varieties

Sr. No.	Premium rice variety	Low-grade rice variety
1.	Basmati Pak	Basmati 2000
2.	Basmati Shaheen	Basmati 198
3.	Basmati Super	Basmati 370
4.	Basmati Super	Basmati 2000
5.	Basmati Shaheen	Basmati 385

(1) Basmati Pak Mixed with Basmati 2000

(b) Basmati Shaheen Mixed with Basmati 198

(c) Basmati Super Mixed with Basmati 370

(d) Basmati Super Mixed with Basmati 2000

(e) Basmati Shaheen Mixed with Basmati 385

Fig. 2: Mixed rice grain samples (premium and low-grade rice) for dataset collection

Three premium rice varieties (*Basmati Pak, Basmati Shaheen and Basmati Super*) were mixed with four low grade rice varieties (*Basmati 2000, Basmati 198, Basmati 370, Basmati 2000, and Basmati 385*). The pure and mixed varieties were shown in Table 1 and samples of images in Fig. 2. The weight of rice grains in a mixed sample was fixed to 20 grams in order to maintain a uniform image sample. Each variety was mixed with other at a weight ratio of 10%, 15%, 20%, 25%, and 30% to test the robustness of the proposed system. 500 images were captured for each mixture (100 images for each weight ratio) making a total of 2500 images in the dataset. A detailed description of the proposed methodology has been discussed in this section. In the first step, pre-processing of the acquired data sets was performed. The resolution of originally captured images was 3456 × 4608, which was resized to 280×260 for preparation of dataset. Then image augmentation of the training dataset was performed using *Keras* library of Python. Convolutional Neural Network (CNN) architecture from deep learning networks has been used for feature extraction, classification of rice grains. CNN accepts images as input and automatically extract features from images using adjacent pixel information to sample the image effectively by convolution layer and then by using prediction layer at the end (Hoang *et al.*, 2020). In this study, filters were applied to the original image and a feature map was created. 3 × 3 filters were applied in this model and a feature map was developed. Polling function was applied to reduce the dimensionality of features to get a specific pixel value of the input image. It significantly reduced the training time. In polling, 2

× 2 filters were applied using the *Keras* library and the average pooling technique was used. In the fully connected layer, back-propagation was performed and weights were updated to decrease loss ratio and improve performance. In the proposed system, CNN VGG16 model has been used. As the network proceeds towards fully connected layers, the number of filters in convolutional layers increased and spatial size decreased. The greater number of filters in convolution layers means the networks had more depth.

RESULTS AND DISCUSSION

Performance of CNN has been evaluated on the dataset of 2500 images using 10-fold cross validation. Dataset has been divided into two parts: 85% dataset for training and 15% used for validation. Hyper-parameter tuning was performed to get best performance of the proposed system using three hyper-parameters: batch size, learning rate, and number of epochs. The learning rate (η) was set to 0.001 and batch size was set to 16 whereas, *softmax* activation function was used in hidden layers. Model training and testing was done by using 100 numbers of epochs, where epoch refers to a hyper parameter that specified number of times the algorithm will operate through the entire set of training data. The overall average accuracy of 89.88% has been achieved at 100 epochs for purity level classification as presented in Table 2. These parameters have been tuned to achieve optimum performance of the model on the given dataset. The average accuracy and test loss has been calculated as shown in Fig. 3 and Fig. 4 respectively**.**

Table 2 Classification of rice purity level using CNN 100 epochs

| Sr. No.	Premium variety	Mixed (low-grade) variety	Adulteration level (%)	Average μ accuracy (%)
1	BasmatiPak	Basmati 2000	10	93
	BasmatiPak	Basmati 2000	15	100
	BasmatiPak	Basmati 2000	20	87
	BasmatiPak	Basmati 2000	25	93
	BasmatiPak	Basmati 2000	30	100
2	BasmatiShaheen	Basmati 198	10	100
	BasmatiShaheen	Basmati 198	15	100
	BasmatiShaheen	Basmati 198	20	87
	BasmatiShaheen	Basmati 198	25	67
	BasmatiShaheen	Basmati 198	30	87
3	Basmati Super	Basmati 370	10	100
	Basmati Super	Basmati 370	15	100
	Basmati Super	Basmati 370	20	100
	Basmati Super	Basmati 370	25	100
	Basmati Super	Basmati 370	30	93
4	Basmati Super	Basmati 2000	10	93
	Basmati Super	Basmati 2000	15	100
	Basmati Super	Basmati 2000	20	73
	Basmati Super	Basmati 2000	25	80
	Basmati Super	Basmati 2000	30	87
5	BasmatiShaheen	Basmati 385	10	87
	BasmatiShaheen	Basmati 385	15	73
	BasmatiShaheen	Basmati 385	20	80
	BasmatiShaheen	Basmati 385	25	87
	BasmatiShaheen	Basmati 385	30	80
Overall average accuracy				89.88 $\mu \pm$ 4.5 σ
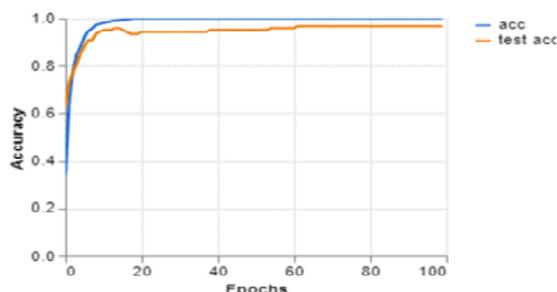
Fig. 3. Test accuracy per epoch using CNN 100 epochs
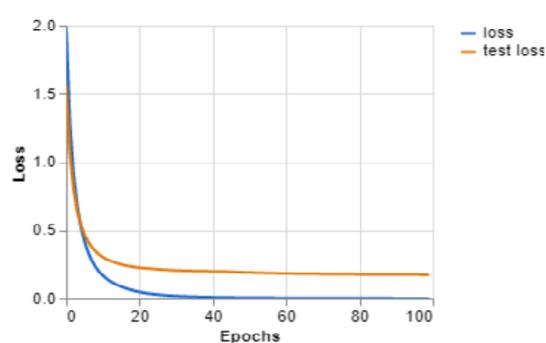

Fig. 4. Test loss per epoch using CNN 100 epochs

The proposed system has been validated at 15 images for each class and achieved highest accuracy of 9 classes. It has made 15 right predictions out of 15 for each 9 classes and 4 classes has achieved 14 right prediction out of 15 as shown in confusion matrix (Fig.5). The system achieved lower accuracy for samples of Basmati Shaheen mixed with 25% of Basmati 198 as it got confused with samples of Basmati Shaheen mixed with 30% of Basmati 198 (as presented in Confusion matrix - Fig. 5).

In this case, out of 15 it has made 10 right predictions. The proposed system has made 11 right predictions and 4 wrong predictions for samples of Basmati Super mixed with 20% of Basmati 2000. It got confused with samples of Basmati Super mixed with 25% of Basmati 2000 and with 30% of Basmati 2000. For samples of Basmati Shaheen mixed with 20%, 25%, and 30% of Basmati 385, the system made 12 correct and 3 wrong predictions because it confused with its own classes of different adulteration level as shown in confusion matrix Fig. 5.

At the starting epoch 85.50%μ accuracy is achieved that was improved to 89.88 μ $\pm$ 4.5 std (δ) at 100 epochs. Similarly, at initial epoch test loss was greater than 0.5that reduced to 0.37 at 100 epochs. Similarly, at initial epoch test loss was greater than 0.5that reduced to 0.37 at 100 epochs. For the performance evaluation of the proposed system, the measures of recall, precision, and F-1 score has been calculated as shown in Table 3. Similar results were noted by Pazoki *et al.* (2014) and Wah *et al.* (2018).
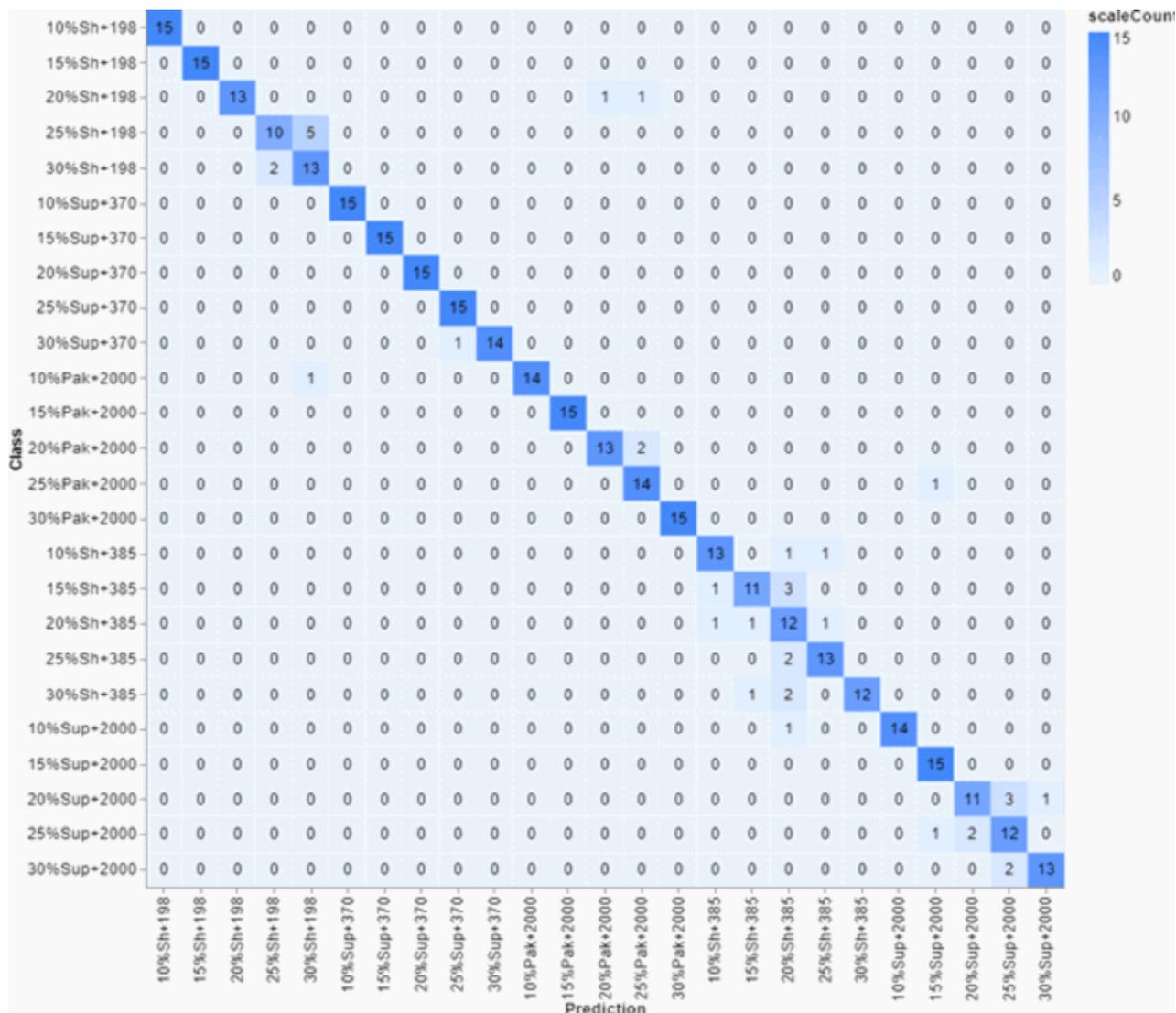
Fig. 5. Confusion matrix for rice purity level classification

Table 3. Recall, precision and F1-score

Purity level classification (%)		
Recall rate	Precision rate	F-1Score rate
0.9492	0.9439	0.9465

CONCLUSION AND RECOMMENDATIONS

This work presented an automated system for purity level classification of rice grains using neural network architecture. The performance of CNN has been evaluated to classify purity level of rice grains from mixed rice varieties on a specially designed dataset of rice grain images. For purity level classification, a dataset comprising a total of 2500 images was captured at five levels of each mixed variety. Dataset is divided into two parts: training and testing of 85:15% ratios. The learning rate and batch size is set to 0.001 and

16 that achieved an average classification accuracy of 89.88 ± 4.5 std for purity level identification. The model achieved relatively less accuracy in predicting purity level of *Basmati Shaheen* mixed with *Basmati 198* at 25% mixing ratio and *Basmati Shaheen* with *Basmati 385* at 15% mixing ratio. It is anticipated that the results would be improved further by increasing the size of training dataset. Further, this work is being extended to a more robust approach of training the model on pure varieties instead of mixed varieties. It would need a more optimised object detection system that is able to detect the purity level in a mixed rice sample based on individual pure varieties.

The ultimate goal of this study was to develop a localized automated system to detect purity level and presence of any adulterant in exported rice grains. This has significant potential to improve the export of

rice from Pakistan that would have a huge impact on agriculture-based economy. Similar model can be used to identify purity level and quality for other food grains such as maize and wheat that are important staple crops in the regions of Pakistan and India.
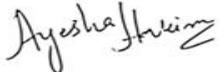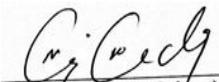
REFERENCES

Agrawal, S., N.K. Verma, P. Tamrakar and P. Sircar. 2011. Content based color image classifiction using SVM. 8th International conference on information technology: New Generations. p. 1090-1094.

Agustin, O.C. and B.J. Oh. 2008. Automatic milled rice quality analysis. Second International Conference on Future Generation Communication and Networking. p. 112-115.

Ali, T., Z. Jhandhir, A. Ahmad, M. Khan, A.A. Khan and G.S. Choi. 2017. Detecting fraudulent labeling of rice samples using computer vision and fuzzy knowledge. Multimedia Tools and Applications 76:24675-24704.

Anami, B.S., N.N. Malvade and S. Palaiah. 2019. Automated recognition and classification of adulteration levels from bulk paddy grain samples. Information processing in agriculture 6:47-60.

Cinar, I. and M. Koklu. 2019. Classification of Rice Varieties Using Artificial Intelligence Methods. International Journal of Intelligent Systems and Applications in Engineering 7:188-194.

Fayyazi, S., M.H. Abbaspour-Fard, A. Rohani, S.A. Monadjemi and H. Sadrnia. 2017. Identification and classification of three iranian rice varieties in mixed bulks using image processing and MLP neural network. International Journal of Food Engineering. 2:13-20.

Gilanie, G., N. Nasir, U.I. Bajwa and H. Ullah. 2021. RiceNet: convolutional neural networks-based model to classify Pakistani grown rice seed types. Multimedia Systems. 17:1-9.

Guzman, J.D. and E.K. Peralta. 2008. Classification of Philippine rice grains using machine vision and artificial neural networks. World conference on Agricultural information and IT. p. 41-48.

Hoang, V.T., Van Hoai, D.P., Surinwarangkoon, T., Duong, H.T. and Meethongjan, K., 2020. A comparative study of rice variety classification based on deep learning and hand-crafted features. ECTI Transactions on Computer and Information Technology, 14(1):1-10.

Kambo, R., and A. Yerpude. 2014. Classification of basmati rice grain variety using image processing and principal component analysis. arXiv preprint arXiv. 1:1405-7626.

Kaur, H. and B. Singh. 2013. Classification and grading rice using multi-class SVM. Scientific and Research Publications. 3:1-5.

Ozan, A., A. Güllü and E. Uçar. 2015. Classification of rice grains using image processing and machine learning techniques. International Scientific Conference, Gabrovo. p. 20-21.

GoP. 2019. Economic survey of Pakistan. Available online withdates at https://www.finance.gov.pk/survey/chapters_19/Economic_Survey_2018_19.pdf.

PARC. 2021. Cultivation of Rice, Pakistan Agriculrture Research Council. Available online with updates at http://www.parc.gov.pk/

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

Pourreza, A., H. Pourreza, M.H. Abbaspour-Fard and H. Sadrnia. 2012. Identification of nine Iranian wheat seed varieties by textural analysis with image processing. Computers and electronics in agriculture 83:102-108.

Shantaiya, S. and U. Ansari. 2010. Identification of food grains and its quality using pattern classification. International Conference on Communication Technology, India. p. 3-5.

Silva, C.S. and U. Sonnadara. 2013. Classification of rice grains using neural networks. Proceedings of Technical Sessions, Srilanka. p. 9-14.

Singh, K.R. and S. Chaudhury. 2020. A cascade network for the classification of rice grain based on single rice kernel. Complex and Intelligent Systems, 6(2):321-334.

Visen, N.S., J. Paliwal, D. Jayas and N. White. 2003. Image analysis of bulk grain samples using neural networks. Annual Meeting. American Society of Agricultural and Biological Engineers, Michigan. p. 1-20.

Wah, T.N., P.E. San and T. Hlaing. 2018. Analysis on Feature Extraction and Classification of Rice Kernels for Myanmar Rice Using Image Processing Techniques. International Journal of Scientific and Research Publications, 8(8):603-606.

Yao, Q., J. Chen, Z. Guan, C. Sun and Z. Zhu. 2009.

Inspection of rice appearance quality using machine vision. Global Congress on Intelligent Systems. p.274-279.

CONTRIBUTION OF AUTHORS

Sr. No.	Author's name	Contribution	Signature
1.	Khalid Abbas	Performed the research work and wrote-up the manuscript	
2.	Ayesha Hakim	Supervisor	
3.	Nasir Nadeem	Co-supervisor	
4.	Adnan Altaf	Co-supervisor	
5.	Hafiz Muhammad Rizwan Iqbal	Contributed in research work and write-up	