Content list available at https://icarneh.icar.gov.in/; ISSN: 0970-6429

Indian Journal of Hill Farming

June 2024, Volume 37, Issue 1, Pages 78-82

Development of a convolutional neural networks-based model to classify the rice varieties

Simardeep Kaur[1*] • Naseeb Singh[2*] • Amit Kumar[2] • Binay Kumar Singh[2] • Suresh Kumar[1]

[1]Division of Biochemistry, ICAR-Indian Agricultural Research Institute, New Delhi, 110012, India
[2]ICAR-Research Complex for North Eastern Hill Region, Umiam, 793103, Meghalaya, India

ARTICLE INFO

***Article history*:**
Received: 06 January, 2024
Revision: 08 January, 2024
Accepted: 08 January, 2024

Key words: *Convolutional neural networks; Computer vision; Rice; Classification*

DOI: 10.56678/iahf-2024.37.01.11

ABSTRACT

Rice (*Oryza sativa* L.) is an important staple food in India, particularly in the North Eastern Hill (NEH) region. The surge in demand for rice, along with its significance in international trade, highlights the need for accurate identification of rice varieties. To address this, we constructed a dataset having four rice varieties—Bhalum-5, Shahsarang, Nagina-22, and IR-64. Our dataset comprises high-quality images of rice seeds, captured with smartphones having different seed counts, including 1-seeded, 5-seeded, and 10-seeded per image. A novel classifier was developed for the classification of rice. Using Convolutional Neural Networks (CNNs) with an architecture comprising 5 layers, the developed model demonstrates significant efficacy in accurately categorizing rice seeds. Experimental results revealed significant achievement and a maximum classification rate of 91.0% for the prominent rice variety (Shahsarang) cultivated in the region. This outstanding accuracy of the developed CNN-based classifier emphasizes its potential applicability in the identification of rice varieties. Thus, this research addresses the pressing need for reliable classification methods for the identification of rice varieties.

1. Introduction

Rice (*Oryza sativa* L.) holds significant importance in India, serving as a dietary staple for a vast population and contributing significantly to nutritional security. As a key source of energy and essential nutrients, rice plays a pivotal role in addressing the nutritional needs of millions (Kaur et al., 2023b). In addition to its domestic significance, India's robust rice production has positioned the country as a major exporter, contributing substantially to the global food market. The export of various rice varieties, not only bolsters the economy but also enhances India as a crucial player in the international food trade (Kaur et al., 2023a). Thus, rice stands as a cornerstone in India's food security, nutrition landscape, and economic vitality. The rice holds the distinction of being the most widely consumed staple food for a significant portion of the global population. With its ready availability, rice contributes 21.0% and 15.0% to global human per capita energy and protein, respectively. The quality of rice seeds, determined by properties such as color, texture, shape, size, and the prevalence of broken kernels, plays a vital role in shaping dietary preferences and nutritional outcomes. However, beyond the consumer and trade considerations, varietal identification and classification hold paramount importance for plant breeders (Gilanie et al., 2021). In this context, seed identification becomes a critical factor in production, necessitating meticulous inspection of rice seeds. This study considered rice seeds cultivated in the North and North East India region, an area where rice cultivation has witnessed a significant surge in demand, underlining the need for precise identification of rice varieties to prevent fraudulent practices in labelling as well as cultivation practices. To address this challenge, we have constructed a comprehensive dataset encompassing four predominant rice varieties—Bhalum-5, Shahsarang, Nagina-22, and IR-64. The dataset features high-resolution images of rice seeds captured using a smartphone having different seed counts, ranging from 1-seeded to 10-seeded configurations.

To classify the rice variety, we developed a novel model specifically designed for rice seed classification, using Convolutional Neural Networks (CNNs) with an architecture

*Corresponding author: naseeb501@gmail.com

comprising convolutional layers, pooling layers, and dense layers. These CNN-based models offer an innovative approach to the classification of rice seed varieties, using their ability to automatically extract hierarchical features from complex image data. The intricate characteristics of rice seeds, such as colour, texture, and shape, are efficiently captured by CNNs, leading to highly accurate and reliable classification outcomes. The developed CNN-based model can expedite the identification process of rice and holds immense potential in enhancing agricultural practices, ensuring the authenticity of rice varieties, and addressing the challenges of fraudulent labelling.

2. Methodology

Image dataset collection

In the present study, four rice varieties namely, Bhalum-5, Shahsarang, Nagina 22, and IR 64 were considered for the classification. These four varieties were selected based on their common cultivation practices in the North Eastern Hill (NEH) Region of India (except Nagina 22 which served as a standard) and the need for the requirement of skilled vision for their identification. For robust classification of these rice seeds, single as well as multiple seeds were placed randomly on a white background and captured using a smartphone in natural illumination. One hundred images were captured for each variety consisting of one, five, and ten rice seeds per image. Captured images were kept in separate folders variety-wise after cropping to a size of 400 × 400 pixels. Fig. 1 shows the sample images captured for the training of the convolutional neural network.

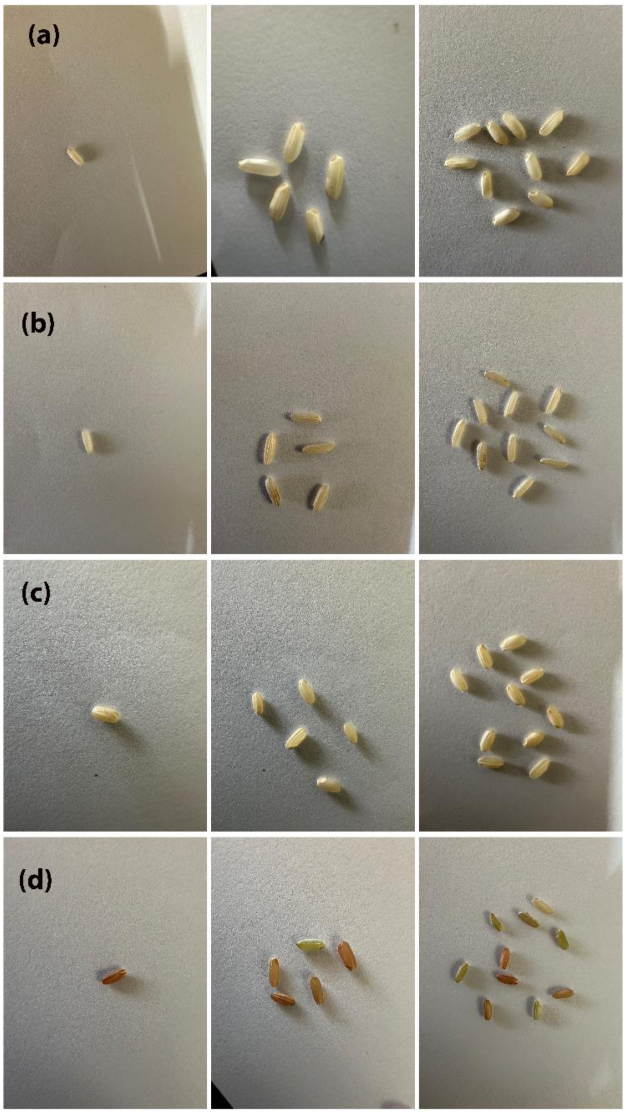

Figure 1. Sample of images captured for training of convolutional neural network for rice identification. (a: Bhalum-5; b: IR-64; c: Nagina-22; d: Shahsarang)

Development and training of CNN model

The CNN architecture for image classification in this study is designed for real-time identification of rice varieties through a mobile application. The model comprises a feature extractor and a classifier as shown in Fig. 2. The feature extractor consists of five convolutional blocks with 3×3 filters, each followed by batch normalization (Ioffe and Szegedy, 2015) and ReLU activation. Max-pooling layers reduce output dimensions after each block. The classifier includes three fully connected layers with 300, 150, and 4 neurons, respectively. ReLU activation is used in all layers except the last dense layer, where softmax activation is applied for classification. The simple and compact design is suitable for real-time mobile application use, focusing on efficiency and accuracy. The CNN was trained on a rice seed dataset split into a 70:20:10 ratio for training, validation, and testing. During training, weights and biases were adjusted based on the Adam optimizer (Kingma and Ba, 2017) with a learning rate of 0.001 and a batch size of 16 images. Convergence was observed within 46 epochs; however, the model was trained for 50 epochs. The categorical cross-entropy loss function addressed the multi-class classification problem. The machine learning library Keras with TensorFlow (Abadi et al., 2016) as the backend was employed, and the training was conducted in Google Colaboratory.

Evaluation of developed model

The proposed convolutional neural network's classification performance for predicting rice varieties is quantitatively assessed using four matrices: accuracy, precision, recall, and F1-score (Singh et al., 2022). Accuracy (Eq. 1) represents the ratio of correctly classified images to the total test images. Precision (Eq. 2) is the ratio of correctly classified images of class i to the total images classified as class i by the model. Recall (Eq. 3) is the ratio of correctly classified images of class i to the total images truly belonging to class i in the test dataset. The F1-score (Eq. 4), a harmonic mean of precision and recall, is utilized to account for false positives and false negatives in the evaluation process.

$$\text{Accuracy}=\frac{TP_i+TN_i}{TP_i+FN_i+FP_i+TN_i} \quad \ldots (1)$$

$$\text{Precision}=\frac{TP_i}{TP_i+FP_i} \quad \ldots (2)$$

$$\text{Recall}=\frac{TP_i}{TP_i+FN_i} \quad \ldots (3)$$

$$\text{F1-score}=\frac{2\times\text{Precision}^i\times\text{Recall}^i}{\text{Precision}^i+\text{Recall}^i} \quad \ldots (4)$$

In Eqs. (1-4), TP_i represents images correctly classified as class i, FP_i are incorrectly classified as class i, FN_i are images of class i misclassified as another, and TN_i are images not belonging to class i correctly classified by the CNN. Additionally, the confusion matrix was utilized to assess classification performance. The confusion matrix, an n×n matrix (n being the number of classes), describes classifier performance on a test dataset with known true labels.

3. Results and Discussion

The proposed CNN underwent training for 50 epochs, utilizing 70.0% of the image dataset for training and 20.0% for validation. The Fig. 3 shows the training curves i.e., accuracy and loss curves wherein it can be observed that after 50 epochs no significant decreased in loss values occurred. Unbiased evaluation was conducted on a remaining test dataset, comprising 44 randomly selected images from the remaining 10.0% of the rice seed dataset.

In Fig. 4, the confusion matrix is illustrated, where each row corresponds to instances in a true class, and each column signifies instances in a predicted class. This matrix facilitates a meticulous examination of class-specific model confusion. The developed CNN model exhibits a higher accuracy in identification of rice varieties. However, classifier resulted into significant incorrect classifications for Nagina-22, as the model erroneously classifies Nagina-22 as Bhalum-5 four times from 16 instances of Nagina-22 images. This misclassification may be attributed to potential similarities in color or shape between these two varieties.

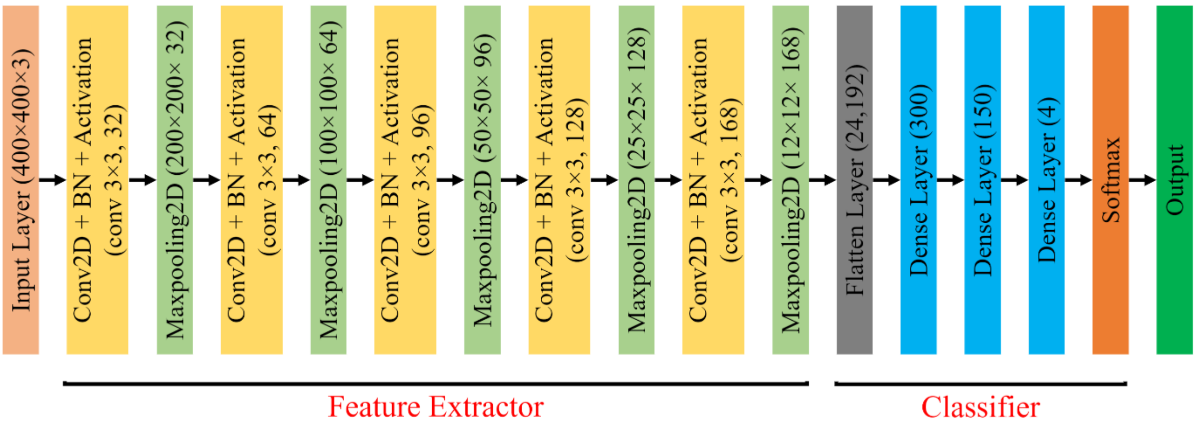

Figure 2. The architecture of the proposed CNN-model for the identification of rice varieties

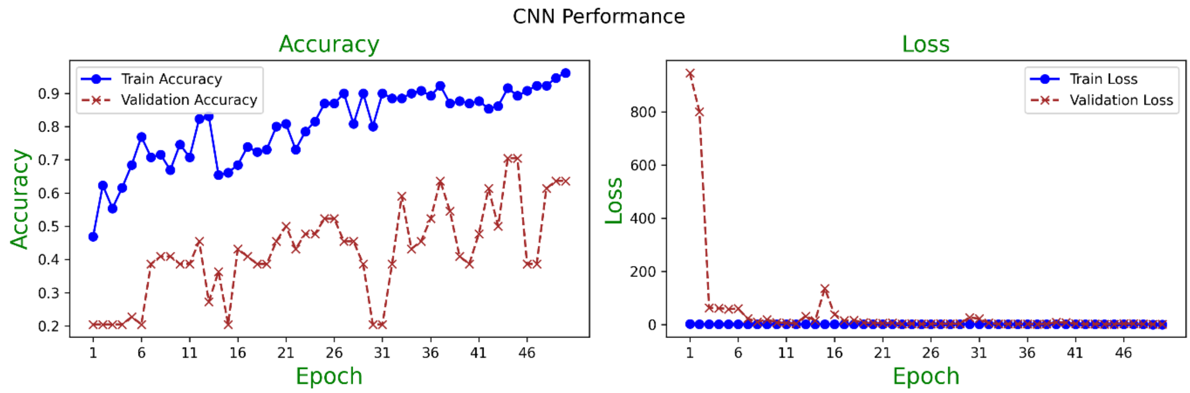

Figure 3. Training curves obtained during the training of CNN model.

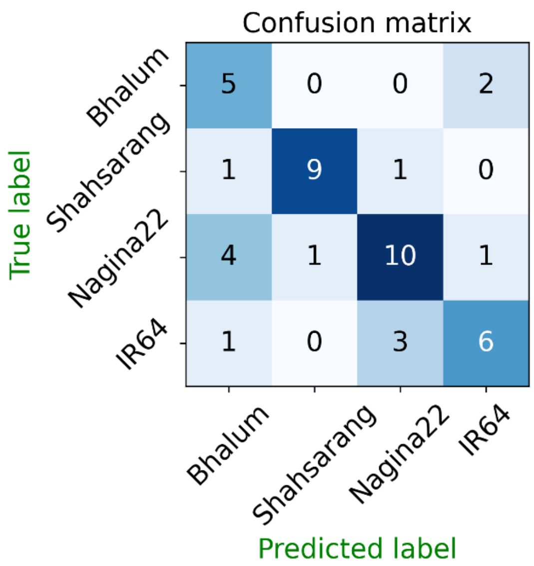

Figure 4. Confusion matrix obtained from testing of developed CNN-based rice seed classifier.

Table 1 provides a quantitative analysis of classification results for each class, including accuracy, precision, recall, and F1-score, obtained through the developed CNN in this study. The achieved performance metrics showcase significant results, with maximum values of 91.0% accuracy, 90.0% precision, 81.0% recall, and 86.0% F1 score. The developed CNN-based classifier achieved maximum accuracy for Shahsarang variety (91.0%) followed by IR-64 (81.0%) and Nagina-22 (75.0%). The higher accuracy achieved for Shahsarang may be attributable to its coloured seed pericarp which is easy to distinguish from other rice varieties considered in the present study. However, the classifier showed a lower accuracy (68.0%) for Bhalum-5. In a study conducted by Ibrahim et al. (2019), focused on the classification of rice grain, a high accuracy of 97.75% was achieved through the incorporation of morphological, texture, color, and wavelet features. In the investigation conducted by Huang and Chien, 2017, a Backpropagation Neural Network (BPNN) was employed for the classification of paddy seed varieties. The results demonstrated notable accuracy, with an average of 92.68% for Taikong-9, 97.35% for Tainan-11, and 96.57% for Taikong-14. These findings highlight the efficacy of utilizing BPNN in distinguishing between various rice varieties. Similarly, Ali et al. (2017) pursued a different approach in their study, employing fuzzy classification to categorize six distinct rice varieties. Their research yielded valuable insights, showcasing the effectiveness of fuzzy classification techniques in discerning between different rice varieties. Such diverse methodologies and outcomes contribute to the broader discourse on innovative techniques for rice variety classification.

Table 1. Evaluation of classifier developed for the identification of rice varieties

Classes	**Metrics**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Bhalum-5	0.68	0.45	0.71	0.56
Shahsarang	0.91	0.90	0.81	0.86
Nagina-22	0.75	0.71	0.62	0.67
IR-64	0.81	0.67	0.60	0.63

4. Conclusion

Our research highlights the critical role of CNNs in revolutionizing the classification of rice seed varieties as the developed CNN model, with its 5-layer architecture, has demonstrated significant efficiency, achieving a maximum classification rate of 91.0% for the prominent rice variety (Shahsarang) cultivated in the region. This success not only enhances the reliability of our model but also highlights its potential to address the pressing issue of fraudulent labelling in the rice industry. In addition, our method can be used to create a database of the rich diversity in farmers' varieties as well as rice germplasm of the North East Region. The significance of accurate rice seed classification extends beyond local concerns of farmers, impacting global trade and ensuring the authenticity of rice varieties crucial for export and import. Our work contributes to the broader field of agricultural technology, providing a reliable and efficient solution for the identification of rice types, thereby safeguarding the interests of consumers, traders, and plant breeders. Looking ahead, the application of CNN-based models in rice seed classification holds immense promise for advancing agricultural practices and ensuring the integrity of food supply chains.

5. Acknowledgements:

Simardeep Kaur acknowledges the financial assistance in the form of a Senior Research Fellowship (SRF) during her Ph.D. from the Council of Scientific and Industrial Research [CSIR Award No: 09/083(0387)/2019-EMR-I], Ministry of Science and Technology, Government of India, New Delhi, India.

6. References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, et al., TensorFlow: A system for large-scale machine learning. *Business Opp*, 2016, 10: 265-283.

Ali T, Jhandhir Z, Ahmad A et al., Detecting fraudulent labeling of rice samples using computer vision and fuzzy knowledge. *Multimed Tools Appl,* 2017, 76: 24675-24704.

Gilanie G, Nasir N, Bajwa U I, Ullah H., RiceNet: convolutional neural networks-based model to classify Pakistani grown rice seed types. *Multimedia Systems*, 2021, 27: 867-875.

Huang K Y, Chien M C., A Novel Method of Identifying Paddy Seed Varieties. *Sensors,* 2017, 17(4):809.

Ibrahim S, Zulkifli N A, Sabri N, Shari A A, Noordin M R., Rice Grain Classification using Multi-class Support Vector Machine (SVM). *IAES International Journal of Artificial Intelligence*, 2019, 8: 215-220.

Ioffe S, Szegedy C., Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015. arXiv:1502.03167 [cs].

Kaur S, Seem K, Duhan N, Kumar S, Kaundal R, Mohapatra T., Transcriptome and Physio-Biochemical Profiling Reveals Differential Responses of Rice Cultivars at Reproductive-Stage Drought Stress. *International Journal of Molecular Sciences,* 2023a, 24: 1002.

Kaur S, Seem K, Kumar S, Kaundal R, Mohapatra T., Comparative Genome-wide Analysis of MicroRNAs and Their Target Genes in Roots of Contrasting Indica Rice Cultivars under Reproductive-Stage Drought. *Genes,* 2023b, 14: 1390.

Kingma D P, Ba J, Adam: A Method for Stochastic Optimization. 2017. arXiv:1412.6980 [cs].

Singh N, Tewari V K, Biswas P K, Dhruw L K, Pareek C M, Singh H D., Semantic segmentation of in-field cotton bolls from the sky using deep convolutional neural networks. *Smart Agricultural Technology*, 2022, 2: 100045.