International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Seeds in Symphony: Unveiling Varietal Diversity with Hybrid Feature Classification

Ramalinga Reddy[1], **Dr. Suma R**[2]

[1]Research Scholar, Department of Computer Science, Sree Siddaganga Evening College of Arts and Commerce, Tumkur.
Email: *rreddypvg[at]gmail.com*

[2]Associate Professor, Department of Information Science and Engineering, S. S. I. T, Tumkur, India
Email: *sumar[at]ssit.edu.in*

Abstract: ***In the agricultural realm, the classification of seed varieties plays a pivotal role in enhancing crop management and fostering agricultural sustainability. Traditional methods of seed classification are often time - consuming and labor - intensive. This paper proposes a hybrid feature classification approach to unveil varietal diversity in seeds more efficiently which combines the power of computer vision and machine learning techniques. We first extract a set of visual features from seed images using state - of - the - art computer vision algorithms. These features capture various characteristics such as shape, color, and texture, which can help differentiate different seed varieties. This abstract invites readers to delve into the world of seed varietal classification, where the harmony of hybrid features orchestrates a symphony of insight into the rich diversity of agricultural seeds. "Seeds in Symphony" stands as a testament to the potential of hybrid feature classification in advancing precision agriculture and contributing to the sustainable future of crop cultivation.***

Keywords: CNNs, Data Extraction, Machine Learning, Testing, Seeds, Classifiers

1. Introduction

A large portion of the world's population is associated with the agriculture sector. Many developing and underdeveloped countries' economies rely on agriculture. This sector has experienced multiple transformations with the increase in the world's population. In the agriculture sector, the farming of crops relies heavily on seeds. Without seeds, there is no chance of producing or harvesting any crops. The human population has been increasing rapidly for many years. Due to this population growth, agricultural land is reducing day by day, which causes a decline in the production of food. To balance consumption rate with production rate, crop production must be increased. In this regard, people have started growing crops and vegetables in their homes. However, not everyone possessed the knowledge needed to do this. Only a person who has expertise in identifying seeds can cultivate them. To eliminate this dependency, there is the need for an automated system that can assist in identifying and classifying the different types of seeds. Several studies have been conducted in which various issues related to seeds have been addressed by using AI techniques, ranging from simple object identification - based techniques to complex texture and pattern identification. In recent studies, machine learning techniques have been observed more frequently to perform seed classification of various crops, fruits and vegetables. Most of these studies have been conducted on a single genre of seed (e. g., weed seeds [**3**], cottonseeds [**4**], rice seeds [**5**, **6**], oat seeds [**7**], sunflower seeds [**8**], tomato seeds [**9**] and corn [**10**, **11**]) with varying purposes. These included observing germination and vigour detection, purification and growth stages.

There are hardly any studies that apply convolutional neural networks (CNNs) [**12**] in their models to address seed identification/classification problems. CNNs are deep learning models consisting of several layers: convolutional, pooling and fully connected layers. The convolutional layers perform feature extractions, the pooling layers perform compression and fully connected layers are for classification. Their main use is image recognition and classification. A general representation of CNNs model can be seen in **Figure 1**. CNN can bring efficiency and accuracy in visual imagery analysis using precise feature extractions.

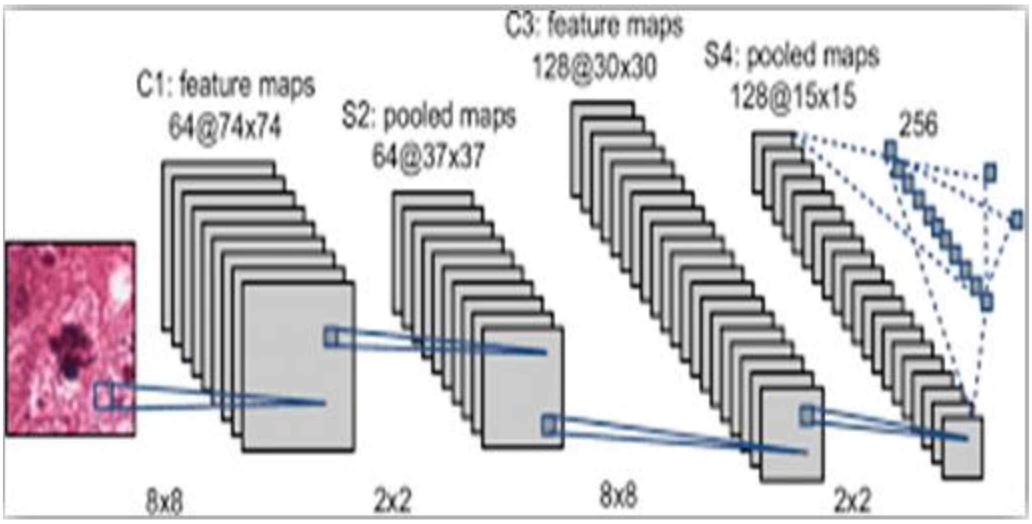

Figure 1: General Representation of CNN Model

Although these traditional techniques/approaches have been successfully applied in many practical studies, they still have many limitations in certain real - world scenarios [**14**].

This research proposed an efficient approach for seed identification and classification based on a deep learning CNN model and the application of symmetry. The model is trained by implementing transfer learning, which focuses on transferring knowledge across domains and is considered a promising machine learning methodology. It is an important machine learning (ML) tool in solving problems related to insufficient training data, which can encourage the idea of not training the model from scratch and significantly reducing the training time [**15**]. In this research, a dataset comprised of symmetric images of various seeds was used to carry out the seed classification process. Many symmetric images from each seed class were captured and then augmented to train the proposed model. Further details pertaining to the dataset are presented in **Section 3**. This research work adopted a VGG16 - trained model in order to accomplish successful seed identification and classification [**16**].

Seeds, the fundamental units of crop propagation, exhibit a remarkable array of genetic, environmental, and physical characteristics that define their distinct varieties. However, the visual homogeneity often presented by a bulk of seeds poses a challenge for conventional classification methods. Enter the "Seeds in Symphony" model—a novel approach that transcends traditional boundaries by harnessing the power of hybrid features.

"Seeds in Symphony" seeks to contribute to the arsenal of tools available to farmers, researchers, and agricultural stakeholders. The potential impact of accurate seed variety classification extends beyond the immediate identification of seeds; it serves as a catalyst for informed decision - making in crop planning, resource allocation, and biodiversity preservation.

The study aims to evaluate the effectiveness of the hybrid feature classification approach using a diverse collection of seed images from various crop species. The results obtained from this research have the potential to revolutionize seed research, enabling rapid analysis of large seed collections and the discovery of previously unknown varietal diversity.

2. Related Work

AI has made many contributions to agriculture, as it brings flexibility, high performance, accuracy and cost - effective solutions to solve various problems [**17**]. Image processing aligned with computer vision has remained the main area of interest for many researchers. Image processing was introduced in agriculture less than a decade ago to address seed sorting and classification for the quality of crop production. Image processing involves feature extraction, which is considered a complex task and helps in object identification and classification. In AI, traditional ML techniques have been widely applied in object identification and classification.

Jamuna et al. [**4**] employed machine learning techniques (e. g. Naive Bayes classifier, a decision tree classifier and MLP) to train the model in feature extraction using a sample of 900 cotton seeds. They reported that the decision tree classifier and MLP gave the same accuracy in classifying the seed cotton, with a rate of 98.7%, and Naive Bayes classifier had an accuracy rate of 94.22%. Their results show that Naive Bayes classifier had the highest error rate, as it made incorrect classification 52 times, whereas the decision tree classifier and MLP made 11 incorrect classifications each.

Feature extraction in traditional ML techniques mainly relies on user - specified features that may cause the loss of some important information, due to which researchers are then faced with difficulty in getting accurate results. Deep learning techniques determine the features of the images in different layers instead of relying on the self - made features of the images [**2**]. For example, in a study by Rozman and Stajnko [**9**], the quality of tomato seeds was reported in terms of their vigor and germination. In their study, they proposed a computer vision system and reported a detailed procedure for image processing and feature extraction by incorporating a Gaussian filter, segmentation, and Region of Interest (ROI). This study incorporated machine learning classification algorithms including Naive Bayes classifiers (NBC), k - nearest neighbors (k - NN), and decision tree classifiers, support vector machines (SVM) and artificial neural networks (ANN) to sort a sample of 700 seeds. Among these algorithms, the ANN (MLP architecture) showed the best performance in seed classification, with an accuracy of 95.44%. Other accuracy rates were NBC at 87.89%, k - NN at 91.66%, DT at 93.66% and SVM at 93.09% [**9**].

A comparative study conducted by Agrawal and Dahiya [**18**] on ML algorithms aimed to classify various grain seeds by using logistic regression (LR), Linear Discriminant Analysis (LDA), k - Nearest Neighbors classifier (kNN), a decision tree classifier (CART), Gaussian Naïve Bayes (NB) and support vector machine (SVM). This study reported performance rates for both linear (LR and LDA) and non - linear (kNN, CART, NB and SVM) algorithms. The accuracy rates for these six algorithms were as follows: the rate of LR was 91.6%, the rate of LDA was 95.8%, the rate of kNN was 87.5%, the rate of CART was 88%, the rate of NB was 88.05% and the rate of SVM was 88.71%. From these results, it can be seen that LDA had the superior performance [**18**].

Another classification study conducted on corn seeds incorporated a probabilistic neural network (PNN). This analysis has been done on waveform data and aligned the terahertz time - domain spectroscopy (THz - TDS) with the machine learning algorithm PNN. The result of their classification rate showed 75% accuracy with 5 - fold cross - validation [**10**].

Moreover, on feature extraction, research carried out by Vlasov and Fadeev [**19**] on grain crop seeds by used a machine learning approach over mechanical methods. They elaborated the details of feature extraction by using traditional machine learning, which included image feature extraction, descriptors retrieval, clustering and finishing with a vocabulary of visual words. Although their major focus was on the traditional ML approach, they also reported deep learning as a second method for seed classification and purification. Their results showed that the deep learning approach reached 95% classification accuracy, where traditional learning had a rate of around 75% [**19**].

In another study, the feature extraction of rice, which was based on physical properties like shape, colour and texture, was done for classification. The researchers used four methods (LR, LDA, k - NN and SVM) of statistical machine learning techniques, and five pre - trained models (VGG16, VGG19, Xception, InceptionV3 and InceptionResNetV2) with deep learning techniques were used for the classification performance comparison. The best accuracy rate was obtained from the SVM method (83.9%), while the best accuracy from the deep learning techniques was obtained from the InceptionResNetV2 model (95.15%) [**6**].

In a study conducted on corn seed purification, a hybrid dataset preparation was reported. The authors performed a tedious procedure to extract features from corn seeds. The hybrid feature dataset was comprised of a histogram, texture features and spectral features. The classification models used in their study included random forest (RF), BayesNet (BN), LogitBoost (LB) and multilayer perceptron (MLP), along with optimized multi - feature using the (10 - fold) cross - validation approach. Among these classifiers, MLP reported outstanding classification accuracy (98.93%) on ROIs size (150 × 150) [**11**].

With regards to traditional AI - based algorithms, they involve detailed steps in the feature extraction technique. They also need assistance from experts, which has a negative impact on the efficiency of the algorithms. Deep learning (DL), unlike traditional classification learning methods, is not limited to shallow structure algorithms. It can perform complex functions with limited samples, extracting the most essential features from only a small number of training samples. A study conducted by Xinshao [**3**], based on a sample of weed seeds, and used the deep learning technique principal component analysis network (PCANet). This study minimised the limitation of manual feature extraction by learning features from the dataset.

DL, which mainly focuses on machine learning, has brought advancements to the producing of results within various data analysis tasks, especially in computer vision [**5**]. DL is capable of representing data by automatically learning abstract deep features of a deep network. It is widely applied in various visual tasks. A study conducted on oat seed variety implemented a deep convolutional neural network (DCNN) in combination with some traditional classifiers, namely logistic regression (LR), support vector machine with RBF kernel (RBF_SVM) and linear kernel (LINEAR_SVM) [**7**].

Another study [**5**] compared the performance of k - nearest neighbours (kNN), a support vector machine (SVM) and CNN models on spectral data of rice. They reported that CNN outperformed the other two models with 89.6% and 87% accuracy rates on a training set and testing set, respectively.

The complexity of sunflower seed features was addressed in a study with a core focus on CNN. The authors claim that the impurities in sunflower seeds are difficult to recognise due to their texture. They reported that, among all available methods, CNN achieved great success in object detection and identification. Therefore, they used it to address their research problem. They developed an eight - layer CNN model to extract image features. The results of their extensive experiments affirmed the model’s accuracy to be much higher than any other traditional model [**8**].

There are not many studies found that incorporated CNN to identify and classify varieties of seeds. In their study, Maeda - Gutiérrez et al. [**20**] reported comparisons among CNN - based architectures, including AlexNet [**21**], GoogleNet [**22**], Inception V3 [**23**] and Residual Network (ResNet 18 and 50) [**24**]. The data set used in their research contained a single genre (tomato plant seeds), whereas, in our research, we have proposed an efficient model for seed identification and classification based on CNN, which is a deep learning model that possesses a high precision level in image features extraction. Unlike most of the relevant studies, the dataset of this research contains 12 types of seed. The training of this model was carried out through transfer learning, which made

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

the focus of this research more about validation and the testing of the model.

3. Proposed Methodology

We proposed a new model based on CNN [12], which provides an efficient approach for seed identification and classification. The whole process consists of three phases:

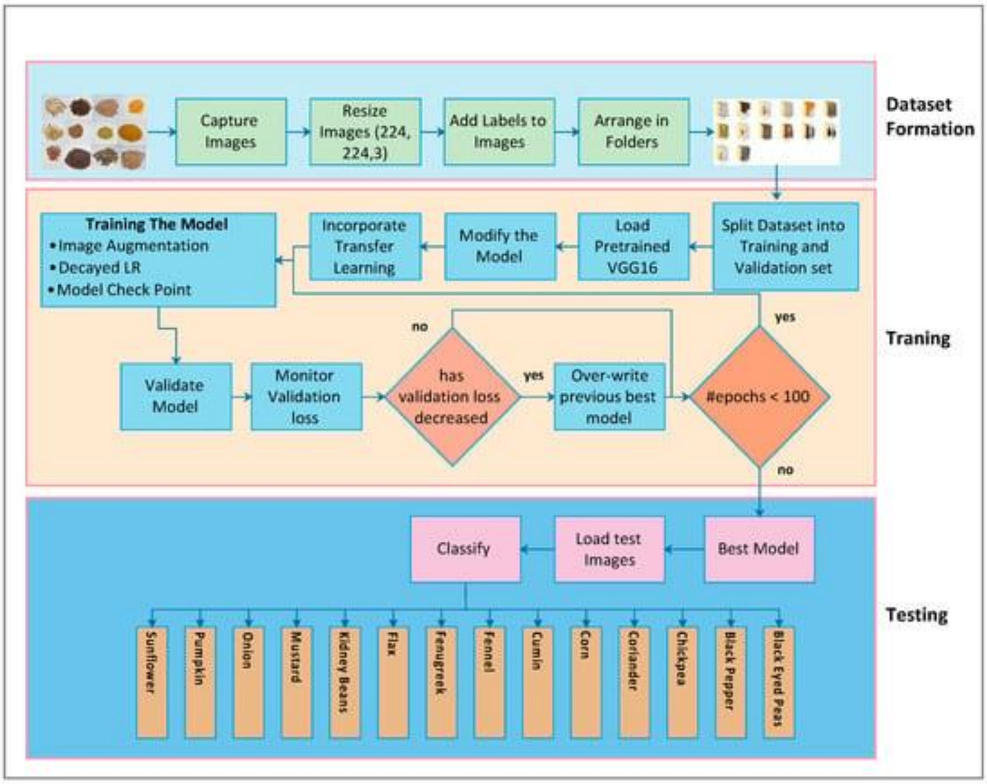

Figure 2: Flow Diagram

(i) Data Set Extraction:

A wide range of seed samples is collected, representing different varieties within the target crop species. These seeds may come from various sources, such as agricultural research centers, seed banks, or experimental fields. The collected seeds need to be prepared for imaging. This may involve cleaning, sorting, and ensuring that the seeds are in a suitable condition for capturing high - quality images. Any damaged or irregular seeds may be excluded from the dataset.

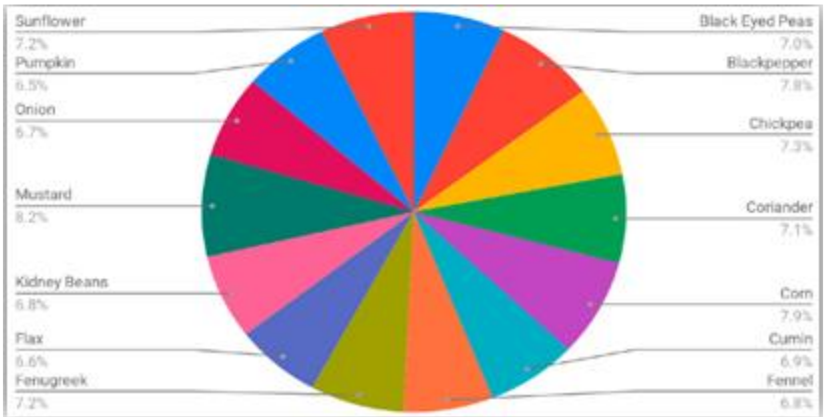

Figure 3: Frequency distribution of dataset (seeds).

Figure 4: Setup for capturing images of seeds

Figure 5: Sample images of 12 different seeds.

(ii) Training

The visual features are extracted from the seed images using computer vision algorithms. These features capture important characteristics such as shape, color, and texture. The extracted features serve as input for the machine learning algorithms. The appropriate machine learning algorithm and model architecture are selected for the classification task. This choice depends on factors such as the nature of the dataset, the complexity of the problem, and the resources available.

Commonly used models include decision trees, support vector machines (SVM), random forests, or deep learning models like convolutional neural networks (CNN) he selected model is trained using the labeled dataset. During training, the model learns to recognize patterns and make predictions based on the extracted features. The training process involves feeding the input features and corresponding labels to the model, which adjusts its internal parameters to minimize the prediction errors. The trained model's performance is evaluated using the validation set. Metrics such as accuracy, precision, recall, and F1 score are calculated to assess the model's classification performance. This evaluation helps fine - tune the model's hyperparameters, such as learning rate or regularization, to improve its performance. Based on the evaluation results, the model may undergo iterative refinement. This involves adjusting the model's architecture, hyperparameters, or feature extraction techniques to further improve its performance. This process continues until satisfactory performance is achieved.

The training process aims to optimize the model's ability to classify seeds accurately based on their visual features. By iteratively adjusting the model's parameters and architecture, the system learns to make more accurate predictions, ultimately enabling the unveiling of varietal diversity in seeds.

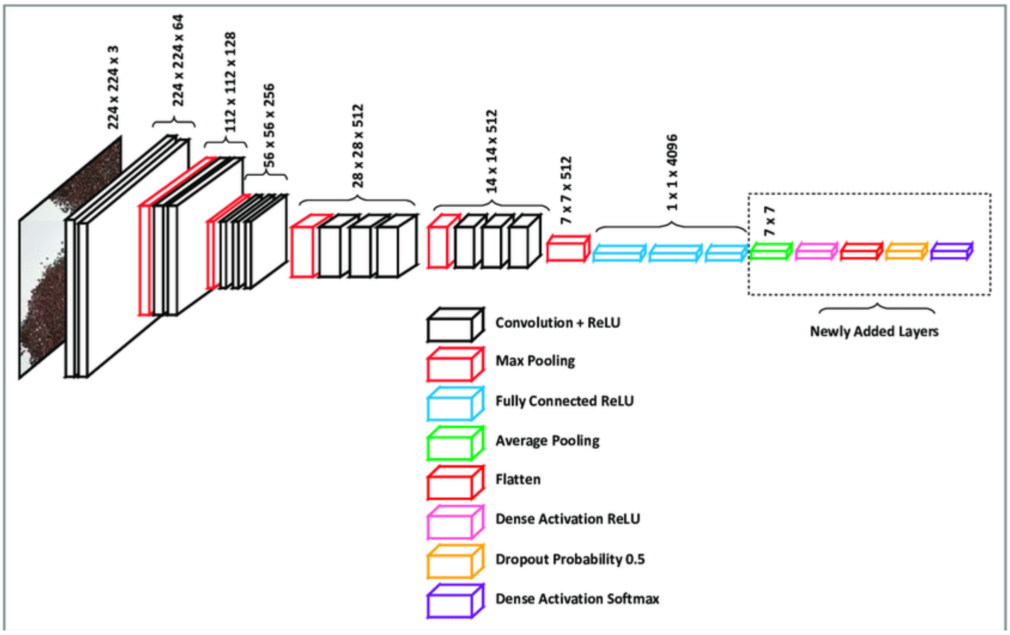

Figure 6: Proposed CNN model for seed classification

(iii) Testing

Unveiling Varietal Diversity with Hybrid Feature Classification, " the testing phase is an important step to evaluate the performance and generalization ability of the trained seed classification system. The testing process involves applying the trained model to unseen seed images and assessing its ability to accurately predict the seed variety.

Class Activation Maps (CAM) can be a valuable tool for visualizing and interpreting the activations of a Convolutional Neural Network (CNN) at different spatial locations. CAM highlights the regions in an input image that contribute the most to a particular class prediction. Here's how you can incorporate CAM for testing your CNN - based seed variety classification model:

1) **Load the Trained CNN Model:** Load the pre - trained CNN model that includes the layers necessary for generating Class Activation Maps.
2) **Load Test Data:** Load a set of test data, including seed images and their corresponding true labels.
3) **Preprocess Test Data:** Preprocess the test data, ensuring it undergoes the same preprocessing steps as during training.
4) **Generate CAMs:** Modify your model to output CAMs for visualizing class activation.
5) **Visualize CAMs:** Use CAMs to visualize the areas of interest in the test images.

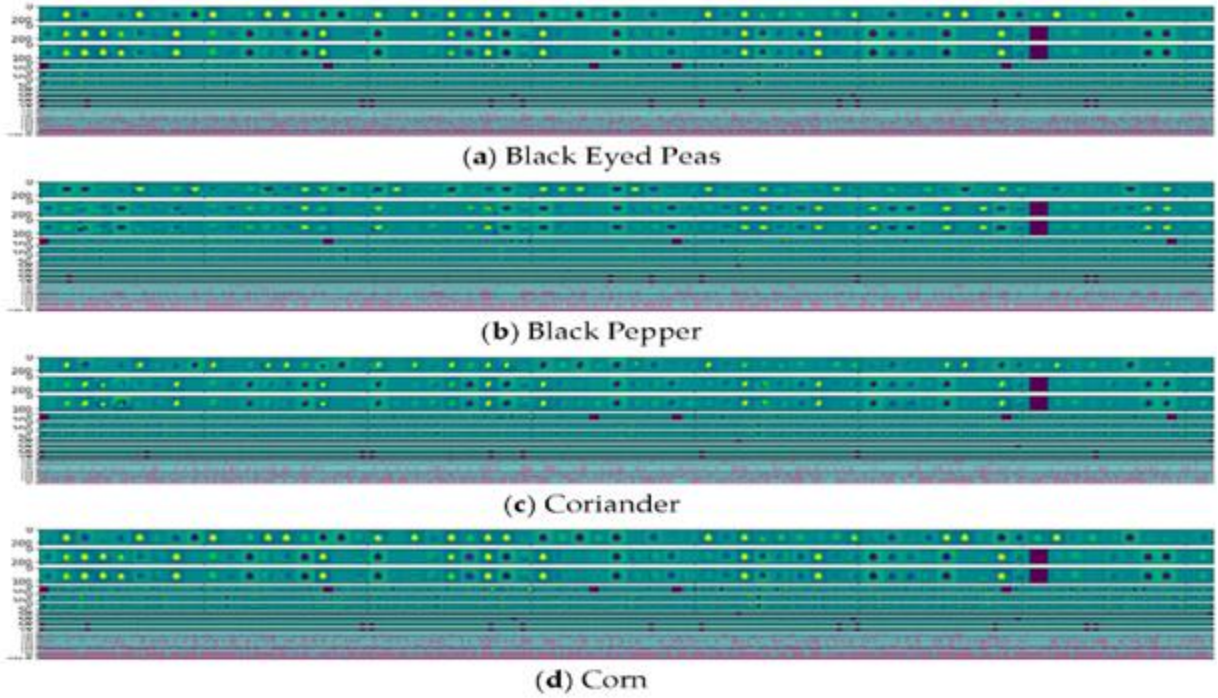

Figure 7: Sample of visual representation of CNN features of seed images

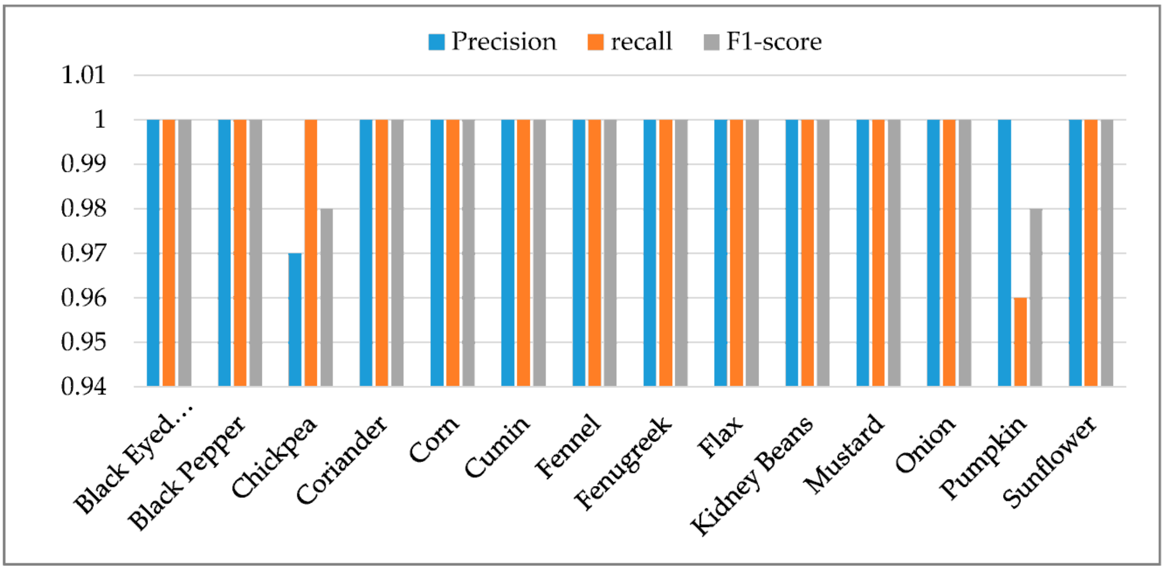

Figure 8: The training results of the proposed model.

4. Conclusion

The "Seeds in Symphony" project stands as a testament to the synergy between advanced technologies and agricultural innovation. By harmonizing hybrid features and cutting - edge machine learning, this research has illuminated the path toward a more precise and informed approach to seed variety classification. As we envision a future where agricultural decisions are guided by intelligent algorithms, the symphony of seeds continues to play a crucial role in nurturing a sustainable and bountiful harvest.

The proposed hybrid feature classification system, when applied to seed classification, enables the unveiling of varietal diversity. By leveraging visual features and machine

Paper ID: SR231120163403 DOI: https://dx.doi.org/10.21275/SR231120163403

learning techniques, the system offers a valuable tool for identifying and categorizing seed varieties accurately.

This research contributes to the understanding and study of seed diversity, which is crucial for agricultural and botanical research, plant breeding, and conservation efforts. The findings highlight the potential for using machine learning and computer vision to enhance seed classification and analysis, paving the way for further advancements in seed research and agricultural practices.

References

[1] Santos, L.; Santos, F. N.; Dos Oliveira, P. M.; Shinde, P. Deep Learning Applications in Agriculture: A Short Review. In *Iberian Robotics Conference*; Springer: Cham, Switzerland, 2020; pp.139–151. ISBN 9783030359898. [**Google Scholar**]

[2] Barman, U.; Choudhury, R. D.; Sahu, D.; Barman, G. G. Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. *Comput. Electron. Agric.*2**020**, *177*, 105661. [**Google Scholar**] [**CrossRef**]

[3] Xinshao, W. Weed Seeds Classification Based on PCANet Deep Learning Baseline. In Proceedings of the 2015 Asia - Pacific Signal and Information Processing Association Annual Summit and Conference APSIPA ASC, Hong Kong, China, 16–19 December 2015; pp.408–415. [**Google Scholar**]

[4] Jamuna, K. S.; Karpagavalli, S.; Vijaya, M. S.; Revathi, P.; Gokilavani, S.; Madhiya, E. Classification of Seed Cotton Yield Based on the Growth Stages of Cotton Crop Using Machine Learning Techniques. In Proceedings of the ACE 2010–2010 International Conference on Advances in Computer Engineering, Bangalore, India, 20–21 June 2010; pp.312–315. [**Google Scholar**]

[5] Qiu, Z.; Chen, J.; Zhao, Y.; Zhu, S.; He, Y.; Zhang, C. Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Appl. Sci.*2**018**, *8*, 212. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[6] Kiratiratanapruk, K.; Temniranrat, P.; Sinthupinyo, W.; Prempree, P.; Chaitavon, K.; Porntheeraphat, S.; Prasertsak, A. Development of Paddy Rice Seed Classification Process using Machine Learning Techniques for Automatic Grading Machine. *J. Sens.*2**020**, *2020*, 7041310. [**Google Scholar**] [**CrossRef**]

[7] Wu, N.; Zhang, Y.; Na, R.; Mi, C.; Zhu, S.; He, Y.; Zhang, C. Variety identification of oat seeds using hyperspectral imaging: Investigating the representation ability of deep convolutional neural network. *RSC Adv.*2**019**, *9*, 12635–12644. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[8] Luan, Z.; Li, C.; Ding, S.; Wei, M.; Yang, Y. Sunflower Seed Sorting Based on Convolutional Neural Network. In Proceedings of the ICGIP 2019 Eleventh International Conference on Graphics and Image Processing, Hangzhou, China, 12–14 October 2019; Pan, Z., Wang, X., Eds.; SPIE: Bellingham, WA, USA, 2020; Volume 11373, p.129. [**Google Scholar**]

[9] Rozman, Č. Denis Stajnko Assessment of germination rate of the tomato seeds using image processing and machine learning. *Eur. J. Hortic. Sci.*2**015**, *80*, 68–75. [**Google Scholar**]

[10] Taylor, J.; Chiou, C. P.; Bond, L. J. A methodology for sorting haploid and diploid corn seed using terahertz time domain spectroscopy and machine learning. *AIP Conf. Proc.*2**019**, *2102*. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[11] Ali, A.; Qadri, S.; Mashwani, W. K.; Brahim Belhaouari, S.; Naeem, S.; Rafique, S.; Jamal, F.; Chesneau, C.; Anam, S. Machine learning approach for the classification of corn seed using hybrid features. *Int. J. Food Prop.*2**020**, *23*, 1097–1111. [**Google Scholar**] [**CrossRef**]

[12] LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.*1**995**, *3361*, 1995. [**Google Scholar**]

[13] Schmidhuber, J. Deep Learning in neural networks: An overview. *Neural Netw.*2**015**, *61*, 85–117. [**Google Scholar**] [**CrossRef**] [**PubMed**] [**Green Version**]

[14] Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2020**, 1–31. [**Google Scholar**] [**CrossRef**]

[15] Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; pp.270–279. [**Google Scholar**]

[16] Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large - Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015; pp.1–14. [**Google Scholar**]

[17] Eli - Chukwu, N. Applications of Artificial Intelligence in Agriculture: A Review. *Eng. Technol. Appl. Sci. Res.*2**019**, *9*, 4377–4383. [**Google Scholar**] [**CrossRef**]

[18] Agrawal, D.; Dahiya, P. Comparisons of classification algorithms on seeds dataset using machine learning algorithm. *Compusoft* **2018**, *7*, 2760–2765. [**Google Scholar**]

[19] Vlasov, A. V.; Fadeev, A. S. A machine learning approach for grain crop's seed classification in purifying separation. *J. Phys. Conf. Ser.*2**017**, *803*, 012177. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[20] Maeda - Gutiérrez, V.; Galván - Tejada, C. E.; Zanella - Calzada, L. A.; Celaya - Padilla, J. M.; Galván - Tejada, J. I.; Gamboa - Rosales, H.; Luna - García, H.; Magallanes - Quintanar, R.; Guerrero Méndez, C. A.; Olvera - Olvera, C. A. Comparison of convolutional neural network architectures for classification of tomato plant diseases. *Appl. Sci.*2**020**, *10*, 1245. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[21] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012; pp.1097–1105. [**Google Scholar**]

[22] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp.1–9. [**Google Scholar**]

[23] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp.2818–2826. [**Google Scholar**]

[24] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp.770–778. [**Google Scholar**]