

Enhancing Pumpkin Seed Classification Using Deep Multilayer Perceptron with Early Stopping

Muhammad Lawal Tijjani
E-extension Department,
National Agricultural Extension
and Research Liaison Services
Ahmadu Bello University, Zaria
Zaria, Nigeria
mltbnahmad@gmail.com

Abdullahi Mohammed
Department of Computer
Science.
Ahmadu Bello University,
Zaria, Nigeria
abdullahiwafu@abu.edu.ng

Mustapha A. Bagiwa
Department of Computer
Science
Ahmadu Bello University,
Zaria, Nigeria
mstphaminu@yahoo.com

Umar Mohammed Aminu
Department of Computer Science.
Ahmadu Bello University,
Zaria, Nigeria
mohammedaminuumar@gmail.com

Abstract—Pumpkin seed classification plays a vital role in various agricultural and food processing industries. Accurate and efficient classification of pumpkin seeds is crucial as it contributes significantly to the oversight of seed quality and the assessment of seed grades, ultimately enhancing crop production efficiency. Traditional methods of classification often rely on manual inspection, which is time-consuming, subjective, and prone to errors. Therefore, the development of an automated and efficient classification system is essential to address these challenges. Multilayer perceptron (MLP) is a widely used neural network model for pattern recognition and classification tasks. The proposed deep MLP structure has been developed with the purpose of extracting complex attributes from the dataset of pumpkin seeds. This is achieved by employing the deep learning paradigm, which enables the capturing of both overall and specific characteristics. Early stopping, a regularization technique, is integrated into the training process to mitigate overfitting and optimize the performance of the model. The use of early stopping ensures that the deep MLP generalizes well to unseen data, thus resulting in reliable and robust classification outcomes. The experimental evaluation of the proposed enhanced MLP is conducted on a validation dataset comprising morphological measurements of 2500 pumpkin seeds of the Urgup Sivriksi and Cercevelik varieties. The performance of the deep MLP with early stopping is compared against traditional MLP models and other commonly used classification algorithms. The final MLP configuration achieves values of 92.20% for accuracy, 94.58% for precision, 89.72% for recall, 94.74% for specificity, and 92.09% for F1-score. The results demonstrate that the deep MLP with early stopping significantly outperforms traditional MLP models and other classification algorithms in terms of classification accuracy and efficiency.

Keywords— Agriculture, multilayer perceptron, Machine Learning, pumpkin seed

I. INTRODUCTION

Seeds are the foundation of crop production, carrying genetic traits that determine yield and resistance to pests. High-quality seeds are therefore very vital for ensuring good harvests and improving farmers' economic outcomes [1]. The response of all additional input within the realm of agriculture is largely contingent upon the calibre of these indispensable seeds [2]. For producing quality seeds, seed analysis and classification offer a wealth of additional information. Seed classification is a procedure whereby diverse seed varieties are classified into distinct categories based on their morphological characteristics [2]. Yet, this critical process has long relied on manual inspection by specialists who meticulously scrutinize seed samples [3]. However, despite receiving standardized training, the classification process lacks uniformity. It depends not only on the skills of technicians but also on individual circumstances, leading to issues like eye fatigue and variations in judgment among inspectors [4]. These challenges, inherent to human involvement throughout the procedure, call for a solution that introduces more and more automation to enhance efficiency.

Pumpkins belong to the genus *Cucurbita* [5], which includes various gourd and squash species are no exception to the above mentioned problem. They are typically classified as *Cucurbita pepo*. The pumpkin seed, known for its nutritional significance, holds a special place in this realm. Comprising 37 percent of carbohydrates, 35 to 40 percent of fats and proteins, along with a spectrum of essential minerals including calcium, potassium, phosphorus, magnesium, iron, and zinc, pumpkin seeds represent a valuable resource [6].

In the field of agriculture, where seed quality stands as a pivotal determinant of productivity [2], the latest innovations are ushering in a new era. The fusion of computational technologies with traditional agricultural practices is transforming how farming is approached. Seed classification, a process traditionally based on visual attributes such as colour, shape, and texture, is undergoing a digital revolution. This innovation promises to bring unprecedented convenience and efficiency to agriculture.

The role of automation in seed classification cannot be overstated. It has the potential to significantly enhance seed quality control, streamline production processes, and identify impurities with unprecedented accuracy. What was once a labour-intensive and time-consuming task, reliant on human inspectors to meticulously examine each seed sample, is now transitioning into an era of automated precision, offering a more efficient approach to seed sorting.

Unlike prior studies that applied shallow MLP or conventional classifiers, this work introduces a deeper MLP model integrated with early stopping to enhance generalization and prevent overfitting. The architecture was designed and tuned specifically for morphological pumpkin seed data, demonstrating that even without complex image-based CNNs, a properly optimized deep MLP can achieve comparable or superior performance. This provides an efficient and interpretable approach for resource-constrained agricultural applications.

II. RELATED WORKS

The research into seed classification using machine learning has been a prominent area of study. These investigations employed diverse machine learning classifiers and attained a notably high level of precision in their works. Demir et al. [7] determined some physical properties of pumpkin seeds which include volume, geometric mean diameter, and some other properties of three types of pumpkin seeds and were classified using a Neural Network. The neural network used is back propagation neural network BPNN and Radial Basis Neural Network RBNN structures. Using root mean squared error as a measurement metric, the error rate of BPNN was found to be 0.6875 and that of RBNN was observed to be 0.0025 hence, it was concluded that the BPNN structure in the classification of pumpkin seeds was insufficient in predicting the morphological features of the seeds whereas the RBNN should be used because it gives low error rate. Therefore, an accurate classification which is vital is reached through the appropriate use of the classifier or its structure.

Ali et al. (2020) conducted research on digital images of different corn varieties, such as Sygenta ST-6142, Pioneer P-1429, Desi Makkai, Kashmiri Makkai, Neelam Makkai, and ICI 339. They introduced a novel threshold-based region growing segmentation (TRGS) technique, which divided each corn image into four regions of interest. From these regions, they extracted hybrid features, consisting of 10 histogram features, 26 texture features, and 19 spectral features. Due to the large number of extracted features, the researchers employed correlation-based feature selection (CFS) with the best first search algorithm to identify the most relevant features. They narrowed down the feature set to nine optimized features. These optimized features were used for classification with machine learning algorithms, including Random Forest (RF), BayesNet (BN), LogitBoost (LB), and Multilayer Perceptron (MLP). This approach aimed to enhance

the accuracy of corn variety classification by selecting the most pertinent features from the extensive dataset.

Koklu et al., [5] examines the effectiveness of various machine learning methods in classifying two types of pumpkin seeds. In the research, twelve morphological features were identified from the image of the pumpkin seeds taken in two dimensions. The methods assessed include Logistic Regression (LR), Multilayer Perceptrons (MLP), Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbor (k-NN). The study reveals accuracy rates ranging from 87.56% to 88.64% for these classifiers, with SVM achieving the highest accuracy at 88.64%, making it the most successful approach for classifying pumpkin seed varieties.

In another study by Bili, [8], rice variety classification was examined. The study included rice types such as Arborio, Basmati, Ipsala, Jasmine, and Karacadag. The researchers extracted three feature sets: twelve morphological features, four shape features, and ninety color features. Initially, they achieved impressive classification accuracy exceeding 97% using various machine learning algorithms (KNN, DT, LR, MLP, RF, and SVM) with RF delivering the highest accuracy. To improve classification, shape features were added to the morphological features, resulting in 98.04% accuracy with RF. Subsequently, when the 90 color features were included, accuracy exceeded 99% across all classifiers, with MLP achieving the highest accuracy at 99.91%.

Javanmardi et al., [9] developed a model for classifying corn seed varieties, specifically nine different types (KSC 201, KSC 704, KSC 290, KSC 380, KSC 301, KSC 400, KSC 260, KSC 647, and KSC 410). They employed the extraction of eighty-seven features, consisting of 6 color features, 17 morphological features, 5 Gy level co-occurrence matrix (GLCM) features, and 59 local binary patterns (LBP) features. To address the challenges posed by interclass similarity within the dataset, they also extracted high-level features using a convolutional neural network (CNN). The research involved the utilization of various machine learning methods for feature classification, including artificial neural network (ANN), cubic support vector machine (SVM), quadratic SVM, weighted k-nearest-neighbor (kNN), boost+d tree, bagged tree, and linear discriminant analysis (LDA). These classifiers were trained using both the hand-crafted features and high-level features, leading to distinct accuracy results for each classification method.

[10], leverages a deep learning framework, DAFFnet that combines dual attention mechanism and feature fusion for seed identification. The methodology combines spatial and channel attention modules to enhance discriminative feature extraction and fuses multi-level features for an improved classification by focusing on critical features while suppressing irrelevant ones, using an image dataset of a soybean seed. It achieved **97.56% classification accuracy**, outperforming other baseline models such as ResNet, DenseNet, and EfficientNet. The study contributes by providing an accurate and scalable approach for seed variety

classification, with limitations such as reliance on high-quality labelled image and also further validation real-world conditions.

III. METHODOLOGY

A. Data collection and Preparation

The dataset used in this study was sourced from Kaggle and consist of 2,500 pumpkin seed samples drawn from two commercial varieties Cercevelik (1,300 samples) and Urgup Sivrisi (1,200 samples) cultivated in the Urgup and Karacaoren regions of Nevsehir, Turkey. The dataset is publicly available at: <https://www.muratkoklu.com/datasets/>. Each record in the dataset represents a single seed and consists of **12 morphological attributes**, which describe the seed's geometric and structural characteristics. These features are as follows:

- (i) **Area:** Represents the total number of pixels within the seed boundary and indicates the overall seed size.
- (ii) **Perimeter:** Denotes the total length of the seed's boundary.
- (iii) **Major Axis Length:** The longest diameter of the fitted ellipse representing the seed shape.
- (iv) **Minor Axis Length:** The shortest diameter of the fitted ellipse.
- (v) **Eccentricity:** Measures how elongated the seed shape is; values closer to 1 indicate more elongation.
- (vi) **Convex Area:** The area of the convex hull surrounding the seed, useful for detecting irregular shapes.
- (vii) **Equivalent Diameter:** The diameter of a circle having the same area as the seed.
- (viii) **Extent:** The ratio of the seed area to its bounding box area, indicating how compact the seed is.
- (ix) **Solidity:** The ratio of the seed area to its convex area; it shows how smooth or concave the seed edges are.
- (x) **Roundness:** Describes the degree of circularity of the seed.
- (xi) **Aspect Ratio:** Represents the proportional relationship between the major and minor axes of the seed.
- (xii) **Compactness:** Expresses how tightly the seed area fills its boundary, indicating uniformity in shape.

B. Feature Scaling

Prior to model training, all input features were normalized to ensure that the contribution from each feature is uniform and also to accelerate convergence process during model learning.

This process prevents features with larger numeric ranges from dominating others. It is aimed at giving all the attributes equal weight and also speeding up the learning process [11], particularly in our model where we will have more than one dense layer. In the dataset, if we consider features like the Area, convex_Area, Eccentricity, solidity, Aspect_Ratio etc, it was observed that features like Solidity, and Eccentricity will be dominated as they do not have the same scale and hence produce an incorrect result. Consequently, we must perform feature scaling to eliminate this problem. To achieve the feature scaling, we used the normalization method of feature scaling. The normalization equation is as presented in the equation

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where X' is the rescaled values ranging between 0 and 1 inclusively, X_{max} is the maximum value of the feature, and X_{min} is the minimum value of the feature.

C. Splitting of the Dataset

After normalization, the dataset was divided into **training** and **testing** subsets to enable performance evaluation and generalization assessment. In the data preprocessing phase, this particular step entails splitting the dataset into two distinct sets. The first subset is utilized in train the model. The second subset serves the purpose of validating the trained model [12]. In this research, 80% of the data is allocated to the training set, while the remaining 20% is reserved for the testing set.

D. Architecture of the Deep MLP Used

The core of this research is the implementation of a Deep MLP architecture designed to classify pumpkin seed varieties based on their morphological attributes. (MLPs) are layered architecture that comprises a hierarchy of processing levels called neurons or nodes, input layer, an output layer, and also hidden layers sandwiched between the input and output layer [130]. MLPs, known as universal approximators, are employed to map sets of appropriate output values. The architectural dimensions of the MLPs were determined prior to commencing the training and evaluation procedures, informed by the collective expertise of the authors and contemporary findings that underscore the implications of training state-of-the-art neural networks within an overparameterized framework. Unlike a shallow MLP that contains only one hidden layer, the deep MLP used here consists of multiple hidden layers capable of learning hierarchical and non-linear feature representations. Each layer transforms the input features into higher-level abstractions, improving the network's ability to discriminate between seed varieties [14]. It enhances classification accuracy by extracting informative and latent features from the data contained in the dataset. These deep neural networks offer the additional benefit of being compatible with dropout, a technique that mitigates the risk of overfitting during training by preventing excessive learning [15]. Moreover, early stopping is used in the training to monitor the performance of the model on validation dataset. Figure 1.1 shows the structure of the MLP used for the pumpkin seed classification.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 32)	416
dense_1 (Dense)	(None, 128)	4224
dropout (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 512)	66048
dropout_1 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 128)	65664
dense_4 (Dense)	(None, 32)	4128
dense_5 (Dense)	(None, 1)	33
=====		
Total params: 140513 (548.88 KB)		
Trainable params: 140513 (548.88 KB)		
Non-trainable params: 0 (0.00 Byte)		

Fig. 1. structure of the proposed model

The recommendation of [16], as well as other available information in the literature was used in developing above architecture and also in choosing the activation function and the optimization algorithm. The hyperbolic tangent activation function, frequently denoted as tanh, is a mathematical function that is widely employed in the field of artificial neural networks. This particular function serves as a non-linear activation function, effectively transforming input values into a range that spans from -1 to 1. The optimization algorithm selected for the MLP architecture was an Adam-based stochastic gradient descent optimizer, recognized as one of the more recently developed and widely adopted solvers, notable for its efficiency in terms of training expenses. Moreover, the experimented model is compared with state-of-the-art work of [5] where cross validation was used to ensure that their models do not overfit. Table 1 and table 2 show the Summary of the Deep MLP Model Training Parameters and the Implementation software/library and hardware respectively:

TABLE I. IMPLEMENTATION SOFTWARE/LIBRARY AND HARDWARE

CATEGORY	Details
Programming Language	Python 3.10
Deep Learning Library	TensorFlow 2.10
Operating System	Windows 10
Hardware	Intel Core i5 CPU, 16 GB RAM,

TABLE II. SUMMARY OF THE DEEP MLP MODEL TRAINING PARAMETERS

STEP / COMPONENT	Description	Parameters / Values
Input Layer	Accepts preprocessed input features for classification	–
Hidden Layer 1	Fully connected layer	32 neurons, tanh activation
Hidden Layer 2	Fully connected layer	128 neurons, tanh activation
Dropout Layer 1	Regularization layer to prevent overfitting	Dropout rate = 0.2
Hidden Layer 3	Fully connected layer	512 neurons, tanh activation
Dropout Layer 2	Regularization layer	Dropout rate = 0.4
Hidden Layer 4	Fully connected layer	128 neurons, tanh activation
Output Layer	Final prediction layer	Sigmoid
Learning Rate	Controls the step size during optimization	0.001
Batch Size	Number of samples per gradient update	32
Epochs	Number of complete training iterations	250
Optimizer	Optimization algorithm	Adam
Loss Function	Error measurement during training	Binary Cross-Entropy

E. The Loss Function Used in Penalizing and Updating Model

To guide the learning process, binary cross-entropy is the error metric chosen for the MLP neural network. So, when the MLP is run on the dataset, the output produced indicates probability or confidence in possible labels. Using these probabilities, the loss function penalizes the MLP's deviation from the target label. In back-propagation, the MLP trainable weights undergo partial derivative calculations for the loss function, and the weight is adjusted by these derivatives. The equation below shows the mathematical expression for the binary cross entropy loss function.

$$J = \frac{1}{M} \sum_{i=0}^M li * \log(p(li)) + (1 - li) * \log(1 - p(li))$$

Where J is the binary cross-entropy (log loss) function,

M is the number of trainable examples,

li is the label of a given observation i (i = 0, 1),

And $p(li)$ is the predicted probability of an observation being *cercevelik*.

To improve generalization and prevent overfitting, early stopping was integrated into the training process. The model's performance on the validation data was monitored continuously, and training was terminated when validation loss failed to improve after several epochs. This approach ensured that the final model captured essential data patterns without the noise.

IV. PERFORMANCE MEASURES

These are quantifiable measures we used to evaluate and gauge the performance and effectiveness of the model. These metrics provide insights into how well the model is performing and can guide further improvements. The performance measures are explained as follows:

True Positive (TP): the number of positive instances in which the model predicted correctly in a class.

True Negative (TN): the number of negative instances in which the model predicted correctly in a class.

False Positive (FP): the number of negative instances in which the model predicted incorrectly in a class.

False Negative (FN): the number of positive instances in which the model predicted incorrectly in a class.

Accuracy: It measures the ratio of correctly classified instances over the total number of instances. The equation below is its mathematical definition:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: It measures the ratio of true positive predictions over the total predicted positive instances. It is useful when the focus is on minimizing false positives. The equation below is its mathematical definition:

$$Precision = \frac{TP}{TP+FP}$$

Recall: It measures the proportion of true positive predictions over the total actual positive instances. It measures the model's ability to capture all positive instances. The mathematical definition is as seen below:

$$Recall = \frac{TP}{TP+FN}$$

F1 score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance for imbalanced class distributions. Below is its mathematical definition:

$$Recall = \frac{TP}{TP+FN}$$

F1 score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance for imbalanced class distributions. Below is its mathematical definition:

$$F1score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

Specificity: measures the proportion of actual negative instances that the model correctly classified as negative. It

indicates how well the model can avoid classifying a negative instance as positive. The equation below is its mathematical definition:

$$Specificity = \frac{TN}{TN+FP}$$

Confusion Matrix: this is another analytical tool employed in the field of machine learning and classification tasks to effectively assess the efficacy of a given classification model. It is particularly useful for assessing how well the model is doing in terms of making correct and incorrect predictions for different classes.

TABLE III. CONFUSION MATRIX

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

V. GENERALIZATION AND MODEL ROBUSTNESS

The strategy of early stopping played an essential role in averting overfitting and guaranteeing the generalization of our model. By closely monitoring the model's performance on a validation set throughout the training process and terminating when performance ceased to exhibit improvement, we successfully mitigated the potential for overfitting. To further enhance robustness, the training process incorporated dropout layers. Dropout randomly deactivates neurons during training, preventing the model from relying too heavily on specific nodes, thus maintaining a balance between bias and variance. These ensure that the model is more apt to exhibit strong performance on unseen data, rendering it a reliable instrument for real-world applications. Figure 2 shows the performance of the model on the validation and training data which was monitored:

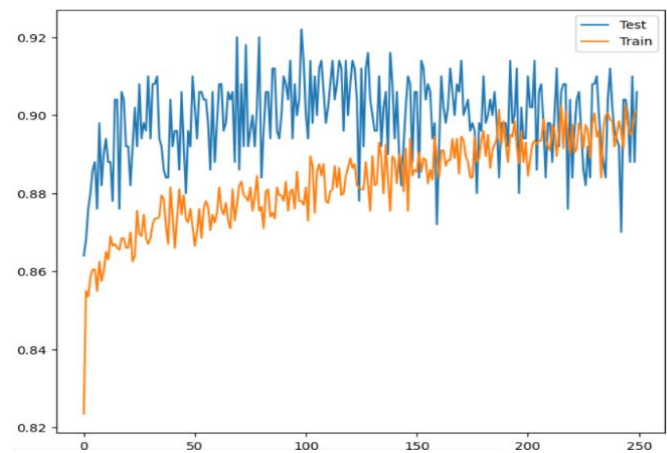


Fig. 2. Training and Validation Graph

VI. RESULT

In our experimental trials, we employed a Multi-Layer Perceptron (MLP) architecture consisting of four concealed strata, incorporating two dropout strata for regularization purposes. Additionally, we employed the implementation of early stopping techniques to mitigate the occurrence of overfitting and guarantee the applicability of the model to unobserved data. The model was subjected to training and evaluation procedures utilizing a dataset that had been meticulously preprocessed. The ensuing examination of results yielded the following performance metrics when tested on the dataset:

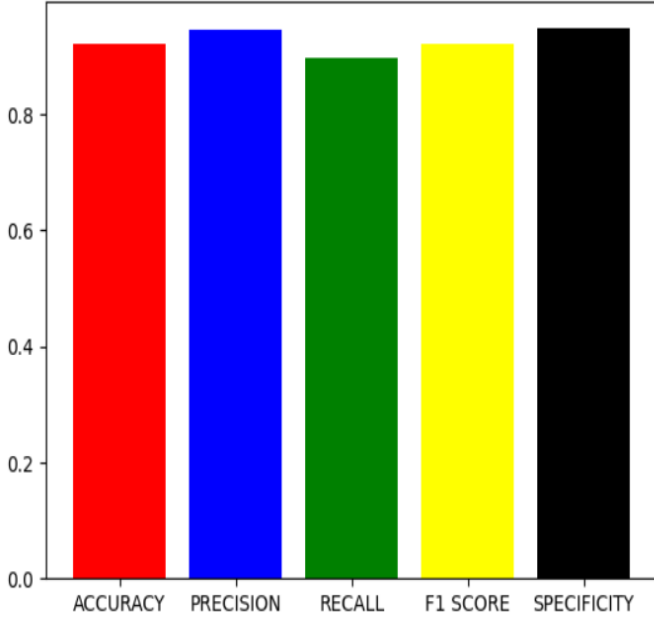


FIG. 3. BAR CHART SHOWING THE DEEP MLP'S PERFORMANCE

Fig. 3. above shows the model's performance on the dataset.

Table I shows the obtained results with the deep MLP with early stopping approach

TABLE IV. CONFUSION MATRIX

Measure	Deep MLP with early stopping
Accuracy	92.2%
Precision	94.58
Recall	89.72
F1 Score	92.09
Specificity	94.74

Table V below shows the work of (Koklu et al., 2021) in comparison with the proposed Deep mlp with early stopping:

TABLE V. COMPARATIVE RESULT OF THE PROPOSED RESULT WITH OTHER MODELS

Measure	LR	MLP	SVM	RF	KNN	Deep MLP with early stopping
Accuracy	87.92	88.52	88.64	87.56	87.64	92.20
Precision	91.08	92.31	92.77	91.15	90.69	94.58
Recall	86.42	86.51	86.39	85.80	86.25	89.72
Specificity	89.73	91.02	91.49	89.72	89.32	94.74
F1-score	88.69	89.32	89.47	88.40	88.41	92.09

VII. DISCUSSION AND CONCLUSION

This study developed and evaluated a deep MLP model integrated with early stopping for classifying pumpkin seed varieties based on their morphological features. The inclusion of early stopping played a significant role in improving generalization by preventing overfitting during training. The high precision (94.58%) indicates that the model produces few false positives, ensuring that correctly classified seeds are highly reliable. The recall (89.72%) value suggests that the model captures the majority of actual positive instances, although minor improvements could be made by optimizing the learning rate or incorporating additional features. The F1-score (92.09%) reflects a well-balanced trade-off between precision and recall, while the specificity (94.74%) demonstrates strong performance in correctly identifying negative samples. The improvement in our model's performance can be largely attributed to the early stopping mechanism. During training, deep networks are prone to overfitting where the model learns noise and specific patterns from the training set, reducing its ability to generalize to new data. The early stopping technique monitors the validation loss at each epoch and halts training once the loss stops improving. This prevents the model from overfitting, ensuring that it retains its generalization capacity and performs effectively on unseen test data. Consequently, the deep MLP achieves a balance between underfitting and overfitting, yielding superior

accuracy and robustness compared to conventional models that do not incorporate this mechanism.

Limitations and Feature Work

Although the proposed deep MLP with early stopping achieved high accuracy, the study was limited to morphological measurements of only two pumpkin seed varieties. Including additional feature types such as color, texture, or spectral attributes, or training on image-based data could enhance the model's discriminative ability. Future work could explore transfer learning or convolutional architectures for visual seed images, as well as testing the model on field-level datasets to assess scalability.

- [17] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020, doi: 10.1109/ACCESS.2019.2962617.

REFERENCES

- [1] J. Chen, L. Zhou, Y. Pu, C. Zhang, H. Qi, and Y. Zhao, "Application of deep learning for high-throughput phenotyping of seed: a review," *Artif. Intell. Rev.*, vol. 58, no. 3, art. no. 76, Jan. 2025, doi: 10.1007/s10462-024-11079-5.
- [2] A. Chaugule, "Survey of seed classification techniques," *Int. J. Sci. Technol. Res.*, vol. 12, no. 13, pp. 1236–1260, 2021.
- [3] L. Journaux, "SS symmetry Classification System," unpublished.
- [4] A. A. Chaugule, "Seed technological development – A survey," *Int. J. Comput. Appl.*, 2013.
- [5] M. Koklu, I. Cinar, and Y. S. Taspinar, "Classification of rice varieties with deep learning methods," *Comput. Electron. Agric.*, vol. 187, p. 106285, Jun. 2021, doi: 10.1016/j.compag.2021.106285.
- [6] D. Peričin, L. Radulović, S. Trivić, and E. Dimić, "Evaluation of solubility of pumpkin seed globulins by response surface method," *J. Food Eng.*, vol. 84, no. 4, pp. 591–594, 2008, doi: 10.1016/j.jfoodeng.2007.07.002.
- [7] B. Demir, I. Eski, Z. A. Kus, and S. Ercisli, "Prediction of physical parameters of pumpkin seeds using neural network," *Not. Bot. Horti Agrobot. Cluj-Napoca*, vol. 45, no. 1, pp. 22–27, 2017, doi: 10.15835/nbha45110429.
- [8] J. A. S. Bili, "Identification of rice varieties using machine learning algorithms," *Ankara Univ. J. Agric. Sci.*, vol. 28, no. 2, pp. 307–325, 2022, doi: 10.15832/ankutbd.862482.
- [9] S. Javanmardi, S. M. Ashtiani, F. J. Verbeek, and A. Martynenko, "Computer-vision classification of corn seed varieties using deep convolutional neural network," *J. Stored Prod. Res.*, vol. 92, p. 101800, 2021, doi: 10.1016/j.jspr.2021.101800.
- [10] Z. Liu, H. Wang, and X. Wu, "DAFFnet: Seed classification of soybean variety based on dual attention feature fusion networks," *Comput. Electron. Agric.*, vol. 214, p. 108273, 2025, doi: 10.1016/j.compag.2025.108273.
- [11] S. A. Borz, "Development of a modality-invariant multi-layer perceptron to predict operational events in motor-manual willow felling operations," *Forests*, vol. 12, no. 9, pp. 1–26, 2021.
- [12] A. Al Bataineh and S. Manacek, "MLP-PSO hybrid algorithm for heart disease prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 1, pp. 451–458, 2022.
- [13] S. Khan, H. Rahmani, S. Afaq, A. Shah, and M. Bennamoun, *A Guide to Convolutional Neural Networks for Computer Vision*. San Rafael, CA, USA: Morgan & Claypool, 2018.
- [14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [16] M. Li, M. Soltanolkotabi, and S. Oymak, "Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks," *Proc. Int. Conf. Artif. Intell. Stat.*, vol. 108, pp. 4313–4324, 2019, [Online]. Available: <http://arxiv.org/abs/1903.11680>.