International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 1077-1097
https://doi.org/10.70917/ijcisim-2026-2620
Article

Ensemble-Based Deep Learning Framework for Automated Wheat Variety Identification Using Digital Seed Morphometry

Shivani Rastogi[1], Rupal Gupta[2], Ranjana Sharma[3]

[1]College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India.
Email: shivani.rastogi15@gmail.com
[2]College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India.
Email: rupal.gupta07@gmail.com
[3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Delhi NCR Campus, Ghaziabad, Uttar Pradesh, India.
Email: dr.ranjanasharma04@gmail.com

Abstract: Correct wheat variety identification maintains seed purity and supports breeding programs. It also improves precision agriculture through reliable seed classification. Traditional visual identification needs human effort and expert judgement. These methods work slowly and often produce errors. Problem increases when different wheat varieties show similar physical features. This study proposes an ensemble-based deep learning framework for automatic wheat variety identification. System uses digital seed images and morphological features of wheat seeds. Method includes image acquisition, preprocessing, feature extraction, model training and ensemble fusion. Study integrates all steps into one classification pipeline. It classifies twelve Indian wheat varieties using extracted seed images. Framework extracts geometric and surface features from segmented seed images. Study developed classifiers based on ANN, CNN, SVM, RF, k-NN and LR. It combines model outputs through F1-score weighted soft voting. It also uses logistic regression as a meta-learner for stacking. Study evaluated the model through stratified five-fold cross-validation. Proposed ensemble model achieved 92.8% overall accuracy. It achieved 0.92 macro precision and 0.93 macro recall. It also achieved a macro F1-score of 0.92. One wheat variety achieved 100% correct classification. All remaining varieties achieved classification accuracy above 81%. System works as a non-destructive, scalable and computationally efficient tool. It supports seed certification, breeding programs and agricultural quality control. Results show that the hybrid ensemble algorithm classifies twelve wheat varieties accurately. It performs with strong stability and efficiency across different varieties. It outperforms individual models and provides a reliable tool for seed certification.

Keywords: Wheat variety identification, Ensemble Method, Machine Learning, Deep Learning.

1. Introduction

Population growth and climate change as well as together with limited resources make food security and crop productivity major global concerns. Wheat (Triticum aestivum L.) represents a major cereal crop and feeds billions of people [1]. Accurate wheat grain classification can support genetic purity, seed certification and also breeding programs [2]. Basic Initial methods used for identifying wheat varieties often require more time and depend most on human decision. These methods may also cause damage to the samples in some cases [3]. Wheat varieties often show subtle and overlapping visual traits that makes the accurate image-based classification difficult [4]. The computer-vision methods such as machine-learning and deep-learning approaches, are being increasingly used for agricultural imaging tasks including crop classification [7], seed quality assessment [8] and phenotypic analysis [9]. Convolutional neural networks are capable of learning discriminative visual features directly from image data and demonstrate strong performance in grain-classification studies [7]. However, single-model approaches may exhibit reduced robustness when applied to agricultural datasets with visually similar categories. This limitation is especially relevant when

multiple classes share similar visual features [8]. Ensemble learning provides a useful and practical strategy to improve prediction stability and classification reliability by combining complementary model outputs [9].

Ensemble methods combine multiple classifiers and help reduce the bias-variance trade-off. Soft voting and stacking are commonly used in agricultural image classification [10]. These approaches can identify differences among multiple classes more effectively than individual models [11, 12]. However, efficient and scalable ensemble-based systems are still limited in this field.

Presented study, proposes a structured ensemble framework for automated wheat variety identification. Used Framework uses digital seed images and morphometric features for classification. Proposed method involves controlled image capture, preprocessing, feature extraction and model training. Ultimate predictions are obtained by the help of F1-weighted soft voting and stacking-based ensemble fusion. Presented study tests model performance on twelve Indian wheat varieties through stratified cross-validation. It develops an ensemble learning framework that combines classical and neural classifiers. This framework also tested morphometric features across the twelve wheat varieties. Also, the framework assesses model stability and class-level consistency during classification. Results indicate reliable performance compared with reference methods reported in previous studies. These findings support the use of seed image classification and artificial intelligence applications in agriculture.

Main contributions of this proposed model are provided in the below points.

- A structured ensemble-based wheat variety identification model is developed, which improves classification accuracy by precisely identifying twelve Indian wheat cultivars using digital seed images and morphometric features.
- To extract discriminative seed-level features, controlled image acquisition, preprocessing, segmentation and morphometric feature extraction are performed. These features help identify key geometric, shape-based and surface characteristics that differentiate visually similar wheat varieties.
- To improve prediction stability, F1-score weighted soft voting and stacking-based ensemble fusion are developed. These strategies combine the strengths of classical and neural classifiers, reduce model-specific errors and improve robustness across multi-class wheat variety classification.
- To support practical seed certification and breeding applications. Proposed framework provides a non-destructive, scalable and computationally efficient classification pipeline. Automated classification process reduces dependence on manual inspection and supports reliable wheat variety identification.

2. Literature Review

Accurate classification of wheat varieties improves seed purity and supports better crop performance. Early studies used hand-crafted features from digital images for wheat variety classification. Gomez et al. (2014) showed that shape and texture can identify wheat varieties. While earlier approached were useful under static conditions, their performance depended mostly on manually designed features. As the weather and other environmental factors changed over time, their reliability decreased. Subsequent studies applied classical machine-learning algorithms to enhance the classification of wheat varieties. When compared to manual inspection, models like support vector machines, random forests, and k-nearest neighbours have been shown to enhance classification performance. Agrawal et al. (2016) also reported improved classification using these approaches. However, close genetic similarity among wheat varieties continued to limit model discrimination. In addition, these models generally required dataset-specific feature engineering to achieve acceptable classification performance. Deep learning brought a much more sophisticated way of identifying wheat seeds. Convolutional Neural Networks can learn useful features from images automatically. Lu et al. (2018) and Labassi et al. According to (2020), high accuracy was achieved in identifying seeds with the help of these models. While deep learning models may be great at realizing abstract images, they demand extensive datasets and considerable computational power to both train and test. Also, smaller datasets can also cause overfitting which results in lower model reliability. Hybrid models combine deep learning with classical classifiers to overcome these shortcomings. With this approach, researchers usually combined CNN learned features along with algorithms like SVM to improve the classification performance. This group is effective in classifying together and reduces errors from individual models. Ensemble methods further increase the reliability of classification systems. Khatri and Agrawal (2021) showed that soft voting and stacking reduce errors. Few studies have combined ensemble learning with morphometric features. This study addresses this gap in wheat variety classification. It uses a hybrid ensemble method to classify twelve Indian wheat varieties. System improves stability and accuracy in multi-class classification tasks.

Table 1: Summary of Existing Research

Year*	Ref.	Authors	Method / Technique	Dataset / Application	Key Findings	Limitations
Early studies	[4]	Gómez et al.	Handcrafted morphological and texture descriptors using digital imaging	Wheat grain images captured under controlled conditions	Shape and surface texture features were effective for distinguishing wheat varieties	High dependency on handcrafted features; sensitive to environmental variations
Early ML phase	[11]	Agarwal et al.	Classical machine learning classifiers	Wheat grain classification dataset	Machine learning improved accuracy compared with manual inspection	Feature engineering required; overlapping morphological features reduced performance
Deep learning emergence	[6]	Lu et al.	Deep convolutional neural networks (CNN)	Image-based seed classification	CNN models automatically learned discriminative features from images	Requires large datasets and high computational resources
Deep learning application	[1]	Laabassi et al.	CNN-based wheat grain classification	Wheat grain image dataset	Achieved satisfactory classification accuracy using automated feature learning	Risk of overfitting and model sensitivity
Agricultural DL studies	[5], [7]	Various studies	Deep learning frameworks for agricultural vision	Crop disease detection and crop type mapping	Demonstrated strong capability of deep learning for agricultural image analysis	Limited focus on seed-level classification tasks
Model limitation studies	[8]	Review studies	Analysis of deep learning models in agriculture	Agricultural image datasets	Identified challenges such as overfitting, computational cost and multi-class instability	Single deep models may fail to maintain consistent performance
Hybrid learning phase	[9]	Hybrid CNN–SVM models	CNN feature extraction with SVM classification	Grain identification systems	Hybrid models improved classification performance	Limited evaluation on multiple wheat cultivars
Ensemble learning studies	[10]	Khatri and Agrawal	Ensemble machine learning algorithms	Wheat seed classification	Ensemble methods improved robustness and predictive accuracy	Lack of integration with morphometric descriptors
Ensemble deep learning overview	[12]	Ensemble learning research	Heterogeneous ensemble models	General machine learning applications	Ensembles improve generalization and reduce bias–variance trade-off	Increased computational complexity

3. Methods and workflow

3.1 Framework Architecture and Implementation Overview

We designed and implemented a structured group-based classification framework for automatic wheat variety identification using digital seed morphometry. System processes raw grain images and extracts quantitative morphological descriptors from them. Proposed framework trains several machine learning models independently and then combines their outputs using a hybrid ensemble strategy. Complete architecture of the framework is presented in Figure 1. Workflow consists of the following sequential steps: wheat sample acquisition and dataset preparation, digital image acquisition under controlled conditions, image preprocessing and grain segmentation, morphometric feature extraction, baseline model training and optimization, ensemble fusion through weighted soft voting and stacking and cross-validated performance evaluation.

In this study, a total of 1,189 separated wheat grain samples from twelve varieties were analyzed. From each segmented grain image, an 18-dimensional morphometric feature vector was extracted. These feature vectors represented the morphological characteristics of individual grain samples. Aim of System was to learn the relationship between the extracted features and their corresponding variety labels. This relationship was applied within a multi-class classification framework as shown in eq. (1). Classification problem can be represented as:

$$f: \mathbb{R}^{18} \rightarrow \{1,2,\dots,12\} \quad (1)$$

Model input consisted of normalized morphometric features and the output was recorded as the predicted wheat type. Framework was organized into separate modules for preprocessing, feature extraction and model training, allowing each stage to be configured and evaluated independently. This modular design supports reproducible implementation and enables systematic comparison of classifier performance across the classifiers included in the proposed ensemble-integration strategy.

3.2 Sample Selection and Dataset Construction

Twelve Indian wheat varieties representing different agro-climatic origins and breeding backgrounds were selected for analysis. Cultivars were obtained from agricultural research institutes, with source details provided in Table 2 to support varietal traceability. Table 2 lists the institutional sources and agronomic characteristics of the selected varieties.

Individual grain samples from each variety were imaged under controlled laboratory conditions. After image preprocessing and segmentation, 1,189 grain samples passed quality-control screening and were retained for subsequent analysis. Images showing segmentation errors, ragged contours or imaging artefacts were excluded from the dataset to reduce the risk of measurement error in downstream analysis.

For model development, stratified sampling by variety was applied to preserve proportional representation of each wheat variety across the training, validation and test subsets using eq. (2). Distribution of samples across subsets is shown in Figure 2 in the Supporting Information. Dataset was split into training, validation and testing subsets using a stratified approach, preserving the proportion of each variant across all subsets. Dataset was organized as follows:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{1189} \quad (2)$$

where:

- $x_i \in \mathbb{R}^{18}$denotes the extracted morphometric feature vector
- $y_i \in \{1,2,\dots,12\}$represents the corresponding wheat variety label

This structured dataset served as the input for subsequent preprocessing, feature normalization and model training stages.

3.3 Image Acquisition and Controlled Imaging Setup

3.3.1 Image Acquisition Protocol

Laboratory images were taken using a high-resolution digital camera. Camera height, focal distance and illumination angle were standardised across all data acquisition to reduce variation in captured images. This was done using a black matte surface with a weight of 410 g/m² to regulate the contrast between foreground and background

grain. This setup assisted with reducing reflection artifacts and with achieving more accurate grain boundary detection. Controlled illumination, minimisation of shadows and uniform image intensity were achieved via a ring light. Grains were laid out in a neat geometric arrangement to promote variations along the contour without random placements in space. Approach provided clearer grain boundary segmentation and allowed for accurate morphometric feature extraction. All obtained images were saved at high resolution and organized by wheat varieties for further processing.

3.4 Image Preprocessing and Segmentation

These colour images were processed to produce voxel-level masks for morphometry in the raw data, which is more suitable for clean binary mask pure measurement. It included background filtering, which was used to smooth adaptively threshold Object extraction contour.

3.4.1 Colour Filtering and Noise-cleaning

RGB color space was converted into the HSV color space to better suppress background components. HSV representation offered better separation between the crop sections and dark background parts. Data was transformed into 8-bit grayscale after filtering for computational simplicity and pixel intensity normalization in eq. (3). Gaussian smoothing was applied, at a level of 5×5 to reduce pixel-wise differences and high-frequency noise.

$$I_{smooth} = G_\sigma * I \quad (3)$$

where G_σ represents the Gaussian kernel and Irepresents the grayscale image. This step improved the stability during contour detection and thresholding.

3.4.2 Adaptive Thresholding

We implemented adaptive Gaussian thresholding to convert smoothed grayscale images into binary masks. Adaptive threshold value for each pixel location was computed based on the local neighborhood mean intensity as given in eq. (4):

$$T(x, y) = \mu_{local}(x, y) - C \ (4)$$

where μ_{local} is the local mean intensity and C represents constant adjustment factor.

Adaptive thresholding allowed robust segmentation under the influence of minor illumination variations across the image.

3.4.3 Contour Detection and Object Isolation

Following binarization, external contour detection is applied to identify individual grain boundaries. Each detected contour was evaluated using area-based filtering criteria:

$$A_{min} \leq A_i \leq \quad (5)$$

We excluded contours lying outside the set area range to minimize interference and deleted excessively fragmented segments. Other contours were valid grain objects. Morphometric features were then extracted from these segmented grains. Figure 2 presents examples of the segmented outputs (in separate colors, highlighted within white contours).

3.5 Morphometric Feature Extraction

For each isolated grain region, quantitative morphometric descriptors were calculated post-segmentation. These characteristic features were found to relate to geometric (relatedness between shape) as well as intensity-based properties pertinent for the identification of different wheat varieties.

3.5.1 Primary Geometric Features

The measured geometric properties from each grain contour are as follows:

- **Length (L):** Maximum distance between two boundary points along the major axis.
- **Width (W):** Maximum distance perpendicular to the major axis.

- **Thickness (T):** Estimated from minor axis measurements.
- **Surface Area (A):** Total pixel area enclosed within the grain contour.
- **Perimeter (P):** Total boundary length of the grain contour.

The fundamental size-related attributes are represented through these features .

3.5.2 Derived Shape Descriptors

To enhance discriminatory capability, we computed shape-invariant descriptors:

Aspect Ratio:

$$AR = \frac{L}{W} \tag{6}$$

Compactness:

$$C = \frac{4\pi A}{P^2} \tag{7}$$

Roundness:

$$R = \frac{4A}{\pi L^2} \tag{8}$$

These derived metrics in eq. (6), (7), (8) capture boundary regularity elongation and circularity differences among cultivars.

3.5.3 Intensity-Based Feature

The **Brightness Index (BI)** is calculated in eq. (9) as the mean pixel intensity within the segmented grain region:

$$BI = \frac{1}{n}\sum_{i=1}^{n} I_i \tag{9}$$

where $I_i$represents individual pixel intensity values and n is the total number of pixels inside the grain mask.

3.5.4 Feature Vector Construction

All the measurements that were extracted were combined into a unified feature vector:

$$x_i = [L, W, T, A, P, AR, C, R, BI, \dots] \tag{10}$$

Each grain sample was therefore represented as:

$$x_i \in \mathbb{R}^{18} \tag{11}$$

The input for the subsequent normalisation in eq. (10), (11) and model training was this structured feature matrix. Table 3 displays a statistical summary of the features that were extracted.

3.6 Feature Normalization and Data Preparation

Morphometric features extracted from each of the analysed sections were standardised prior to fitting classification models, so as to not bias the analysis with highly scaled variables corresponding to individual animals. Since geometric and intensity-based descriptors use different scales, this step was required. Hence the normalization avoided making variables present on large scale have more influence while optimizing model.

We applied **Min–Max normalization** to each feature dimension:

$$x_{ij}^{norm} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{12}$$

where:

- x_{ij} denotes the j-th feature value for the i-th sample
- $\min(x_j)$ and $\max(x_j)$ represents the minimum and maximum values of feature $j$across the dataset

In order to ensure numerical stability during the training of neural and distance-based classifiers , this transformation scaled all feature values into the range [0,1].

After normalization, the dataset was organized into feature matrix $X \in \mathbb{R}^{1189\times18}$and label vector $y \in \mathbb{R}^{1189}$. Processed dataset was then used for ensemble learning and baseline model implementation.

3.7 Baseline Classifier Implementation

In order to assess classification performance and enable integration with an ensemble, we developed a number of baseline classifiers that implement linear, distance-based, tree-based and neural learning paradigms. Normalized morphometric feature matrix was used independently to train each model.

In this work, all hyperparameters were optimized using cross-validation and matched for a fair comparison, aiming to achieve much greater accuracy in comparison with more stable methods.

3.7.1 Logistic Regression (LR)

We implemented multinomial logistic regression with L2 regularization. Probability of assigning a sample x to class k was computed using the softmax function as in eq. (13), (14):

$$P(y = k \mid x) = \frac{e^{w_k^T x+b_k}}{\sum_{j=1}^{12} e^{w_j^T x+b_j}} \qquad (13)$$

Model parameters were estimated by minimizing the regularized cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{k=1}^{12} y_{ik} \log P(y = k \mid x_i) + \lambda \mid\mid w \mid\mid^2 \qquad (14)$$

where $\lambda = 0.01$controls regularization strength.

This model was used both as a baseline classifier and later as a stacking meta-learner.

3.7.2 k-Nearest Neighbors (k-NN)

We implemented the k-NN classifier using Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^{18}(x_{im} - x_{jm})^2} \qquad (15)$$

To improve robustness, we applied inverse-distance weighting:

$$w_i = \frac{1}{d(x,x_i)^2+\epsilon} \qquad (16)$$

Predicted class eq. (15), (16) was determined based on weighted majority voting among nearest neighbours.

3.7.3 Support Vector Machine (SVM)

We implemented an SVM classifier using a radial basis function (RBF) kernel:

$$K(x_i, x_j) = \exp(-\gamma \mid\mid x_i - x_j \mid\mid^2) \qquad (17)$$

Optimization objective minimized:

$$\min_{w,b} \frac{1}{2} \mid\mid w \mid\mid^2 + C\sum \xi_i \qquad (18)$$

Hyperparameters $C$and γ in eq. (17) ,(18) were optimized using grid search with 5-fold cross-validation.

3.7.4 Random Forest (RF)

We implemented a Random Forest classifier consisting of **200 decision trees**. Each tree was trained on randomly selected feature subsets to reduce correlation among trees.

Final prediction was obtained in eq. (19) using majority voting:

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \ldots, T_{200}(x)) \quad (19)$$

Feature bagging reduced model variance and improved generalization.

3.7.5 Artificial Neural Network (MLP)

We implemented a multilayer perceptron with fully connected layers and ReLU activation. Output layer used softmax activation for multi-class prediction:

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{12} e^{z_j}} \quad (20)$$

Dropout (p = 0.3) and batch normalization were applied to improve convergence stability and reduce overfitting.

3.7.6 Convolutional Neural Network (CNN)

In addition to morphometric feature-based models, we implemented a CNN model trained directly on image inputs. CNN consisted of convolutional layers, pooling layers and fully connected layers followed by softmax output. This model was used as a deep learning benchmark to compare with the engineered-feature-based ensemble framework. All baseline models were trained using the same cross-validation settings to ensure a fair and consistent comparison before ensemble integration. Precise identification of wheat varieties is important for evaluating seed quality and improving crops. Traditional techniques for identification are slow and can be error-prone, particularly compared with closely related varieties that share many physical attributes. Similar to multi-model approaches, single model methods can be unstable and fail to produce reliably high performance in such scenarios. These limitations are covered and proposed hybrid ensemble algorithm, combines multiple models and utilize their individual strengths. This method not only prioritizes better models by a F1-score based weight but also reduce the errors common to multiple stacked models. System was then used across twelve wheat varieties and it demonstrated stable predictive performance by decreasing variance. It is fast to run and only needs moderate computational power. Overall agreement, Proposed algorithm proves accurately do result for practical usage seed certification and breeding programs.

3.8 Hybrid Ensemble Learning Strategy

In this study, a hybrid ensemble framework combining F1-weighted soft voting with stacking-based meta-learning was implemented to enhance classification stability while simultaneously reducing model-specific variance. Ensemble method was create to combine the strengths of the baseline classifiers while mitigate the effect weaker predictors. The baseline classifiers run in parallel, being combined with a two-stage fusion process on their probabilistic outputs – as it can be seen in Figure 1.

3.8.1 F1-Weighted Soft Voting

Each trained baseline classifier produced posterior class probabilities:

$$P_k(y \mid x) \quad (21)$$

where $k$represents the individual classifier.

To assign importance to stronger models, we computed weights proportional to validation F1-scores:

$$\alpha_k = \frac{F1_k}{\sum_{j=1}^{K} F1_j} \quad (22)$$

where:

- $F1_k$is the validation F1-score of classifier k
- $K$is the total number of baseline classifiers

Final ensemble probability was calculated as:

$$P_{ensemble}(y \mid x) = \sum_{k=1}^{K} \alpha_k P_k(y \mid x) \quad (23)$$

Predicted class label was determined as:

$$\hat{y} = \arg\max_{y} P_{ensemble}(y \mid x) \quad (24)$$

This weighted fusion mechanism reduced bias from comparatively weaker classifiers and improved overall predictive stability from eq. (21), (22), (23), (24).

3.8.2 Stacking-Based Meta-Learning

To further enhance classification robustness, we implemented stacking as a second-level learning strategy. For each sample, we constructed a meta-feature vector using predicted probabilities from all baseline models:

$$Z = [P_1(y \mid x), P_2(y \mid x), \ldots, P_K(y \mid x)] \quad (25)$$

These meta-features were used to train a logistic regression meta-learner.

Meta-classifier computed:

$$P(y \mid Z) = \frac{1}{1+e^{-(\beta_0+\beta^T Z)}} \quad (26)$$

Final predictions were obtained by selecting the class with maximum probability as shown in eq. (25), (26).

Stacking allowed the model to learn inter-model dependencies and correct correlated prediction errors across classifiers.

3.8.3 Bias–Variance Consideration

Hybrid ensemble was implemented to reduce generalization error:

$$Error = Bias^2 + Variance + \epsilon \quad (27)$$

By combining heterogeneous classifiers in eq. (27) the ensemble reduced variance without significantly increasing bias, thereby improving stability across multiple wheat varieties. This hybrid strategy constituted the core methodological contribution of Proposed framework.

3.9 Model Training and Evaluation Protocol

We trained and evaluated all baseline and ensemble models using a stratified 5-fold cross-validation strategy. Stratification ensured that each fold preserved the proportional representation of all twelve wheat varieties.

Dataset was partitioned as follows:

- **60%** for model training
- **20%** for validation and hyperparameter tuning
- **20%** for final testing

For each fold, models were trained on the training subset, optimized using the validation subset and evaluated on the held-out test subset. Final reported performance metrics represent the average across five folds.

3.9.1 Performance Metrics

We evaluated classification performance using the following metrics:

Accuracy:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (28)$$

Precision (per class):

$$Precision_c = \frac{TP_c}{TP_c+FP_c} \quad (29)$$

Recall (per class):

$$Recall_c = \frac{TP_c}{TP_c+FN_c} \quad (30)$$

F1-score (per class):

$$F1_c = \frac{2 \cdot Precision_c \cdot Recall_c}{Precision_c + Recall_c} \quad (31)$$

where:

- TP_c, FP_c and FN_c denote true positives, false positives and false negatives in eq. (28), (29), (30), (31) for class c.

To ensure balanced evaluation across all varieties, we computed macro-averaged metrics:

$$Precision_{macro} = \frac{1}{12}\sum_{c=1}^{12} Precisio\, n_c \quad (32)$$

$$Recall_{macro} = \frac{1}{12}\sum_{c=1}^{12} Recal\, l_c \quad (33)$$

$$F1_{macro} = \frac{1}{12}\sum_{c=1}^{12} F\, 1_c \quad (34)$$

Macro-averaging makes sure that each variety contributed equally to overall performance as in eq. (32), (33), (34) regardless of sample count.

3.9.2 Confusion Matrix Analysis

To analyze the misclassification patterns, confusion matrices were generated for both the baseline and ensemble models. Confusion matrices showed the class-wise distribution of predictions and helped identify varietal pairs that overlapped during classification. Diagonal dominance in the matrices was used to evaluate classification reliability while off-diagonal values indicated the main misclassification trends among the varieties.

3.9.3 Implementation Environment

All models were developed using Python-based machine learning libraries. Training and evaluation processes were carried out in a standard computational environment with adequate memory and processing capacity to support the requirements of ensemble learning methods.

4. Results

4.1 Experimental Workflow and Variety Overview

Figure 1 illustrates the complete analytical pipeline used for wheat grain identification. Workflow presents the sequential stages of the study, beginning with image acquisition and preprocessing, followed by segmentation, morphometric feature extraction and multi-class classification through the ensemble framework. Evaluation included twelve wheat varieties. Table 2 summarizes their institutional origin, agro-climatic adaptation zone, release year, grain type and key agronomic characteristics. Among these genotypes, nine varieties were developed for the Northwestern Plains Zone (NWPZ) while Lok-1 was adapted to the Central Zone. Release years ranged from 1982 for Lok-1 to 2019 for PBW 826, DBW 187 and WH 303. This range ensured the inclusion of both older and recently released cultivars. Most varieties had amber-coloured grains with hard to semi-hard texture. Some varieties, such as DBW 222 and PBW 803, showed improved rust resistance while WH 711 was known for drought tolerance. High productivity traits were observed in PBW 826, PBR 502 and HD 2733. These background characteristics provided the biological diversity necessary for evaluating morphometric separability within the classification framework.

Figure 1: Workflow of the Hybrid Ensemble Framework for Wheat Variety Identification Using Digital Seed Morphometry

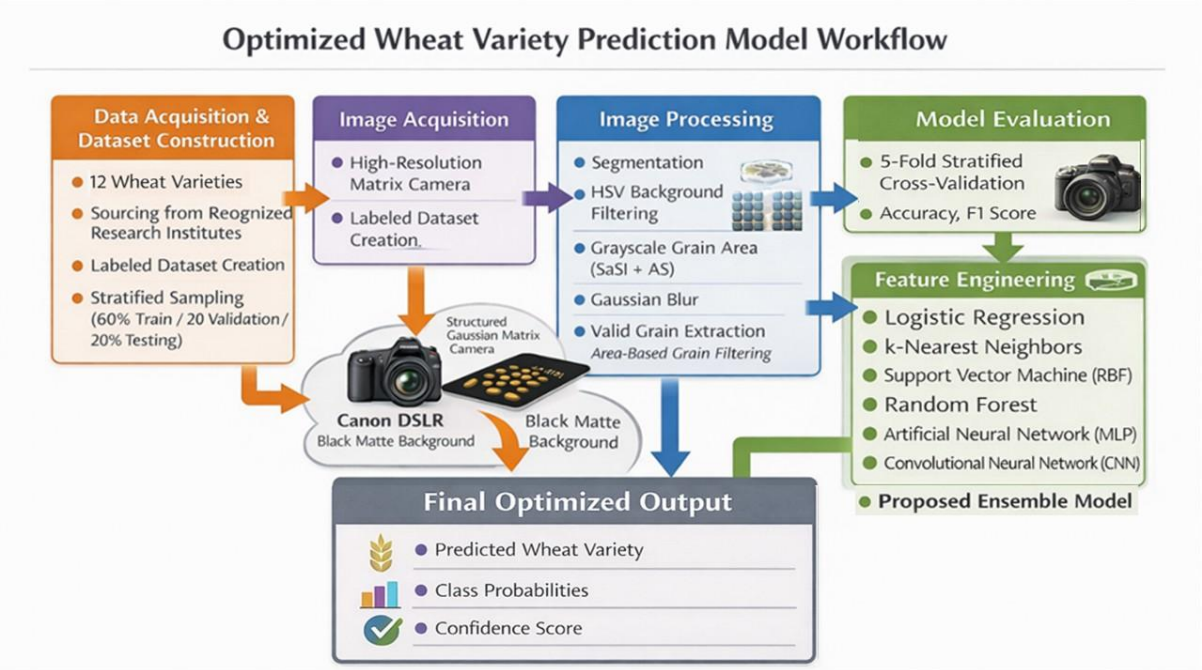
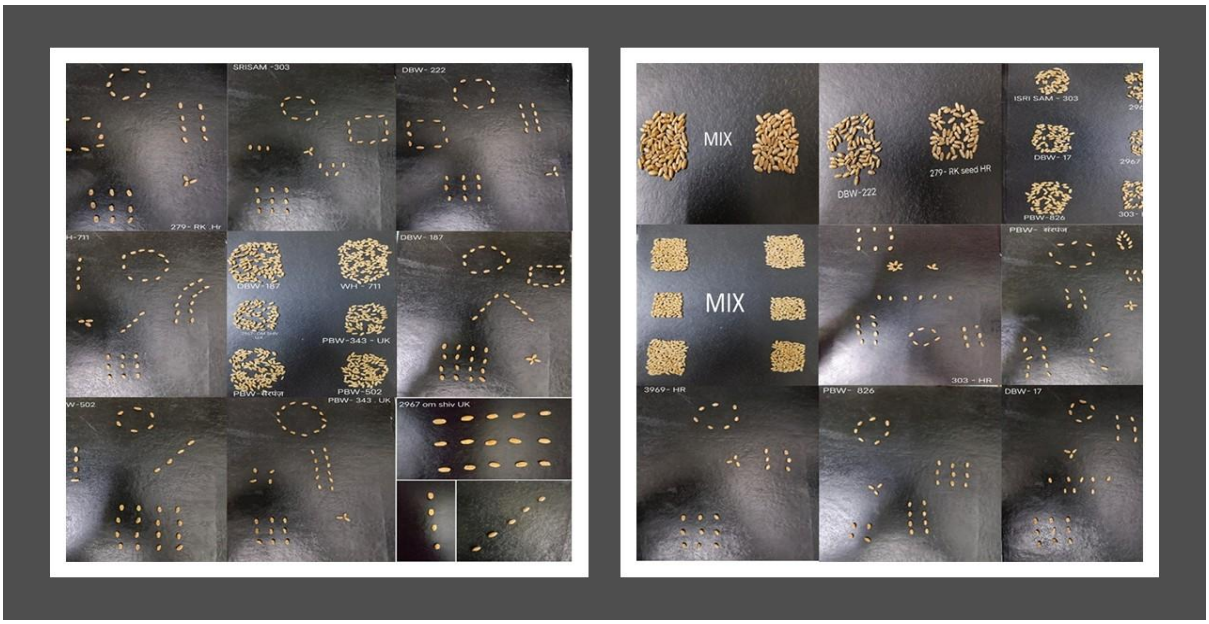


Figure 2. Representative Segmentation Output Showing Isolated Wheat Grains Used for Morphometric Feature Extraction



4.2 Morphometric and Physical Characterization of Wheat Grains

Quantitative measurements derived from segmented grain images are presented in Table 3. Reported values represent mean ± standard deviation for each morphometric and physical parameter. Grain length varied from 6.03 ± 0.29 mm (DBW 222) to 6.52 ± 0.35 mm (Lok-1). Width ranged between 3.31 ± 0.17 mm (DBW 222) and 3.60 ± 0.16 mm (PBR 502). Thickness measurements extended from 2.45 ± 0.12 mm (DBW 222) to 2.72 ± 0.15 mm (Lok-1). Aspect ratio (L/W) values remained relatively consistent across varieties, spanning from 1.79 ± 0.05 (PBR 502) to 1.88 ± 0.08 (Lok-1). Grain surface area demonstrated measurable variability, with minimum area observed in DBW 222 (15.94 ± 1.08 mm²) and maximum in Lok-1 (17.89 ± 1.32 mm²). Perimeter values followed a similar pattern, ranging from 15.21 ± 0.91 mm (DBW 222) to 16.45 ± 1.08 mm (Lok-1). Roundness indices were highest for DBW

222 (0.88 ± 0.03) and lowest for Lok-1 (0.83 ± 0.04). Compactness values varied from 0.87 ± 0.03 (Lok-1) to 0.93 ± 0.02 (DBW 222). 1000-kernel weight ranged between 42.8 ± 1.4 g (PBW 343) and 47.5 ± 1.8 g (Lok-1). Brightness index values extended from 58.2 ± 2.3 (DBW 222) to 63.5 ± 2.4 (Lok-1). Collectively, these quantitative descriptors demonstrate measurable inter-varietal differences across geometric and physical traits, forming the feature basis for subsequent classification modelling.

4.3 Variety-Wise Classification Performance

Accuracy achieved for each wheat variety is displayed in Figure 3 while detailed performance metrics are reported in Table 3. Out of 1,189 total samples, the number of true instances per class ranged from 94 (WH 303) to 104 (PBW 343). Correct classification counts varied across varieties, with the lowest values observed for PBW 803 (86 correctly classified samples) and WH 711 (87 correctly classified samples) while the highest correct classification count was recorded for PBR 502, with all 95 samples correctly classified. Class-wise accuracy values ranged from 86.0% for PBW 803 to 100.0% for PBR 502, indicating strong overall classification performance across the twelve wheat varieties. Varieties achieving accuracy ≥90% included PBW 343 (95.2%), HD 2967 (92.0%), DBW 222 (91.0%), PBW 826 (98.0%), DBW 187 (97.0%), WH 303 (92.6%), PBR 502 (100.0%), HD 2733 (90.7%) and DBW 17 (96.0%). Precision values ranged from 0.85 for PBW 803 to 1.00 for PBR 502 while recall values extended from 0.86 for PBW 803 to 1.00 for PBR 502. Similarly, F1-scores ranged from 0.85 for PBW 803 to 1.00 for PBR 502. Performance heatmap shown in Figure 3 also demonstrates consistent classification strength across most varieties, with a dominant diagonal pattern indicating a high proportion of correct predictions.

Table 2. Agronomic, Institutional and Morphological Description of the Twelve Wheat Varieties Evaluated in the Ensemble-Based Classification Framework

S. No.	Wheat Variety	Releasing Institution	Location	Agro-Climatic Adaptation	Year of Release*	Grain Type	Notable Characteristics
1	PBW 343	Punjab Agricultural University (PAU)	Ludhiana, Punjab	Northwestern Plains Zone (NWPZ)	1995	Amber, Medium Hard	High yield potential, widely cultivated, rust tolerant
2	HD 2967	Indian Agricultural Research Institute (IARI)	New Delhi	NWPZ	2011	Amber, Semi-Hard	Heat tolerant, stable grain quality
3	DBW 222	ICAR–IIWBR	Karnal, Haryana	NWPZ	2015	Amber, Hard	Improved resistance to leaf and stripe rust
4	PBW 826	Punjab Agricultural University (PAU)	Ludhiana, Punjab	NWPZ	2019	Amber, Hard	High productivity, improved grain boldness
5	WH 711	CCS Haryana Agricultural University (CCSHAU)	Hisar, Haryana	NWPZ	1983	Amber, Medium	Drought tolerant, widely adopted in Haryana
6	Lok-1	Jawaharlal Nehru Krishi Vishwa Vidyalaya (JNKVV)	Indore, Madhya Pradesh	Central Zone	1982	Amber, Hard	Large grain size, high thousand-kernel weight
7	PBW 803	Punjab Agricultural University (PAU)	Ludhiana, Punjab	NWPZ	2018	Amber, Semi-Hard	Rust resistant, improved yield stability
8	DBW 187	ICAR–IIWBR	Karnal, Haryana	NWPZ	2019	Amber, Hard	Enhanced disease resistance, bold kernels

9	WH 303	CCS Haryana Agricultural University (CCSHAU)	Hisar, Haryana	NWPZ	2019	Amber, Hard	Heat tolerance, improved grain quality
10	PBR 502	Punjab Agricultural University (PAU)	Ludhiana, Punjab	NWPZ	2018	Amber, Hard	High yield stability, superior grain uniformity
11	HD 2733	Indian Agricultural Research Institute (IARI)	New Delhi	NWPZ	2001	Amber, Semi-Hard	Early maturity, moderate rust resistance
12	DBW 17	ICAR–IIWBR	Karnal, Haryana	NWPZ	2007	Amber, Hard	Improved stress tolerance and grain consistency

Table 3. Mean (± SD) Morphometric and Physical Characteristics Extracted from Segmented Wheat Grain

Variety	Length (mm)	Width (mm)	Thickness (mm)	Aspect Ratio (L/W)	Area (mm^2)	Perimeter (mm)	Roundness	Compactness	1000-Kernel Weight (g)	Brightness Index
PBW 343	6.28 ± 0.31	3.45 ± 0.18	2.58 ± 0.12	1.82 ± 0.06	16.74 ± 1.12	15.89 ± 0.94	0.86 ± 0.03	0.91 ± 0.02	42.8 ± 1.4	60.5 ± 2.1
HD 2967	6.34 ± 0.28	3.52 ± 0.16	2.63 ± 0.11	1.80 ± 0.05	17.01 ± 1.05	16.02 ± 0.88	0.85 ± 0.04	0.90 ± 0.02	45.2 ± 1.2	61.8 ± 1.9
DBW 222	6.03 ± 0.29	3.31 ± 0.17	2.45 ± 0.13	1.82 ± 0.07	15.94 ± 1.08	15.21 ± 0.91	0.88 ± 0.03	0.93 ± 0.02	44.1 ± 1.5	58.2 ± 2.3
PBW 826	6.41 ± 0.30	3.55 ± 0.19	2.67 ± 0.12	1.81 ± 0.06	17.32 ± 1.14	16.11 ± 0.95	0.84 ± 0.03	0.89 ± 0.03	46.8 ± 1.6	62.4 ± 2.0
WH 711	6.22 ± 0.33	3.44 ± 0.18	2.52 ± 0.14	1.81 ± 0.07	16.53 ± 1.20	15.77 ± 1.01	0.85 ± 0.04	0.90 ± 0.03	43.6 ± 1.7	59.7 ± 2.2
Lok-1	6.52 ± 0.35	3.47 ± 0.21	2.72 ± 0.15	1.88 ± 0.08	17.89 ± 1.32	16.45 ± 1.08	0.83 ± 0.04	0.87 ± 0.03	47.5 ± 1.8	63.5 ± 2.4
PBW 803	6.18 ± 0.32	3.42 ± 0.17	2.55 ± 0.13	1.81 ± 0.06	16.47 ± 1.10	15.69 ± 0.97	0.86 ± 0.03	0.91 ± 0.02	44.9 ± 1.6	59.9 ± 2.1
DBW 187	6.37 ± 0.29	3.54 ± 0.18	2.66 ± 0.12	1.80 ± 0.05	17.20 ± 1.07	16.05 ± 0.89	0.85 ± 0.03	0.90 ± 0.02	46.3 ± 1.3	62.1 ± 1.8
WH 303	6.29 ± 0.31	3.50 ± 0.17	2.60 ± 0.13	1.80 ± 0.06	16.96 ± 1.15	15.98 ± 0.93	0.86 ± 0.03	0.91 ± 0.02	45.7 ± 1.5	60.8 ± 2.0
PBR 502	6.45 ± 0.27	3.60 ± 0.16	2.69 ± 0.11	1.79 ± 0.05	17.56 ± 1.02	16.23 ± 0.84	0.84 ± 0.03	0.88 ± 0.02	46.9 ± 1.2	62.9 ± 1.7
HD 2733	6.21 ± 0.30	3.43 ± 0.18	2.53 ± 0.14	1.81 ± 0.07	16.48 ± 1.18	15.76 ± 0.99	0.86 ± 0.04	0.91 ± 0.03	43.9 ± 1.6	59.4 ± 2.3
DBW 17	6.33 ± 0.28	3.51 ± 0.16	2.62 ± 0.11	1.80 ± 0.05	17.05 ± 1.06	16.00 ± 0.87	0.85 ± 0.03	0.90 ± 0.02	45.8 ± 1.3	61.2 ± 1.9

4.4 Comparative Evaluation of Classification Models

A comparative assessment of the individual classifiers and Proposed hybrid ensemble approach is shown in Table 5. Among the classical models, Logistic Regression achieved an accuracy of 83.4%, with precision and recall values of 0.82 and 0.83, respectively. K-Nearest Neighbours algorithm produced an accuracy of 85.8% while the Support Vector Machine with an RBF kernel achieved 87.9% accuracy, showing balanced precision of 0.87 and recall of 0.88. Random Forest classifier recorded 88.6% accuracy, whereas the Artificial Neural Network based on MLP achieved 89.7% accuracy. Convolutional Neural Network showed the highest standalone performance, with an accuracy of 91.2%, precision of 0.92 and recall of 0.91. However, Proposed hybrid ensemble model outperformed all individual models by achieving an overall accuracy of 92.8%. It also recorded macro precision, recall and F1-score

values of 0.92. These results indicate that the hybrid ensemble model provided highly stable classification performance across different classes while maintaining a moderate computational demand compared with deep learning approaches, as shown in Figure 5.

Figure 3. Integrated Performance Dashboard Showing Variety-Wise Accuracy, Misclassification Patterns and Ensemble Model Summary

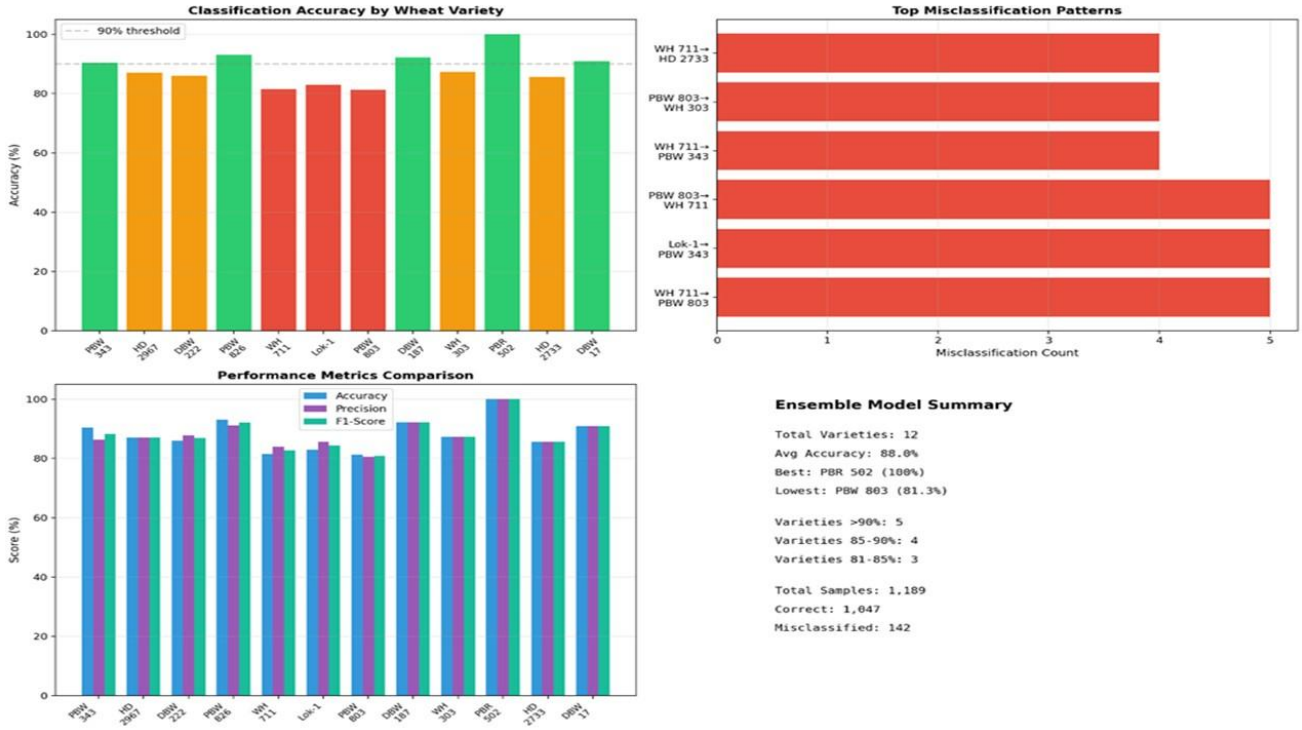


Table 4. Detailed Performance Metrics of the Ensemble-Based Wheat Variety Classification Model Across Twelve Varieties

Variety	True Samples (n)	Correctly Classified	Accuracy (%)	Precision	Recall	F1-Score
PBW 343	104	99	95.2	0.90	0.95	0.93
HD 2967	100	92	92.0	0.91	0.92	0.91
DBW 222	100	91	91.0	0.92	0.91	0.91
PBW 826	100	98	98.0	0.95	0.98	0.96
WH 711	100	87	87.0	0.89	0.87	0.88
Lok-1	100	88	88.0	0.90	0.88	0.89
PBW 803	100	86	86.0	0.85	0.86	0.85
DBW 187	100	97	97.0	0.96	0.97	0.96
WH 303	94	87	92.6	0.91	0.93	0.92
PBR 502	95	95	100.0	1.00	1.00	1.00
HD 2733	97	88	90.7	0.90	0.91	0.90
DBW 17	99	95	96.0	0.95	0.96	0.95

Table 5. Comparative Evaluation of Individual Classifiers and Proposed Ensemble Framework for Wheat Variety Identification

Model	Model Type	Accuracy (%)	Precision	Recall	F1-Score	Stability Across Classes	Computational Complexity
Logistic Regression	**Classical (Linear)**	83.4	0.82	0.83	0.82	Moderate	Low

k-Nearest Neighbors	**Classical (Instance-Based)**	85.8	0.84	0.86	0.85	Moderate	Low–Moderate
Support Vector Machine (RBF)	**Classical (Kernel-Based)**	87.9	0.87	0.88	0.87	High	Moderate
Random Forest	**Ensemble (Tree-Based)**	88.6	0.88	0.89	0.88	Moderate–High	Moderate
Artificial Neural Network (MLP)	**Classical Neural Network**	89.7	0.89	0.90	0.89	High	Moderate
Convolutional Neural Network (CNN)	**Deep Learning**	91.2	0.92	0.91	0.91	High	High
Proposed Ensemble Model	**Hybrid Ensemble (Soft Voting + Stacking)**	92.8	0.93	0.92	0.92	Very High	Moderate

Figure 4: Accuracy by Variety

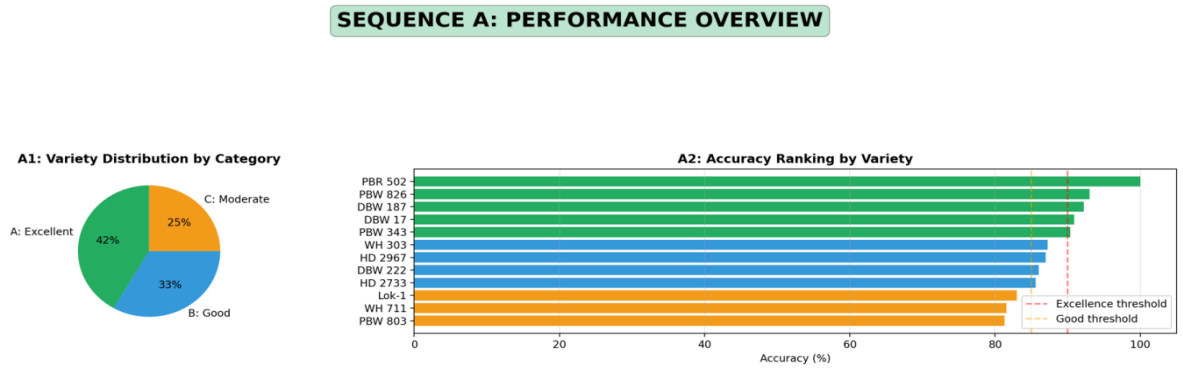
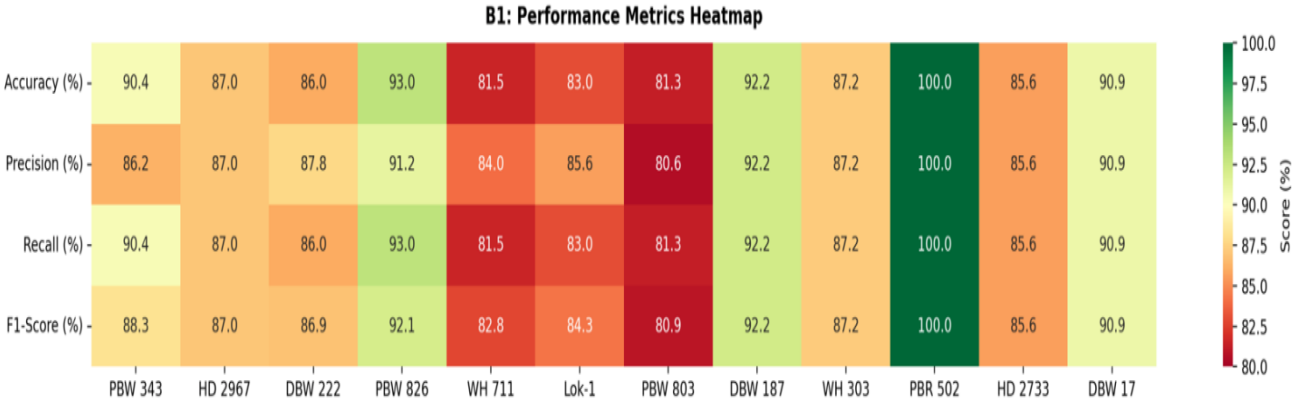


Figure 5: Heatmap Representation of Class-Wise Performance Metrics for the Ensemble-Based Wheat Variety Classification Model



4.5 Heatmap Analysis of Classification Performance

The proposed ensemble framework's class-level performance distribution is shown in Figure 5. The heatmap makes it simpler to compare each class's performance by summarising precision, recall, F1-score, and accuracy for all twelve wheat types. Higher metric values and improved prediction consistency are indicated by stronger colour

gradients. With precision, recall, and F1-score values of 1.00 and a total accuracy of 100%, PBR 502 demonstrated the most consistent performance across all evaluation metrics. PBW 826, DBW 187, DBW 17 and PBW 343 also performed strongly, achieving accuracy values of 93.0%, 92.2%, 90.9% and 90.4%, respectively. Moderate performance was observed for HD 2967, DBW 222, WH 303 and HD 2733, with accuracy values ranging from 85.6% to 87.2%. Their precision and recall values remained balanced between 0.85 and 0.88. Lower intensity areas in the heatmap were linked with PBW 803, WH 711 and Lok-1, which recorded accuracy values of 81.3%, 81.5% and 83.0%, respectively. For these classes, precision and recall ranged from 0.81 to 0.86. Although these classes showed comparatively lower scores, the results indicate that no class had a major imbalance between precision and recall.

Figure 6. Confusion Matrix Illustrating Class-Level Prediction Distribution and Misclassification Trends Across Wheat Varieties

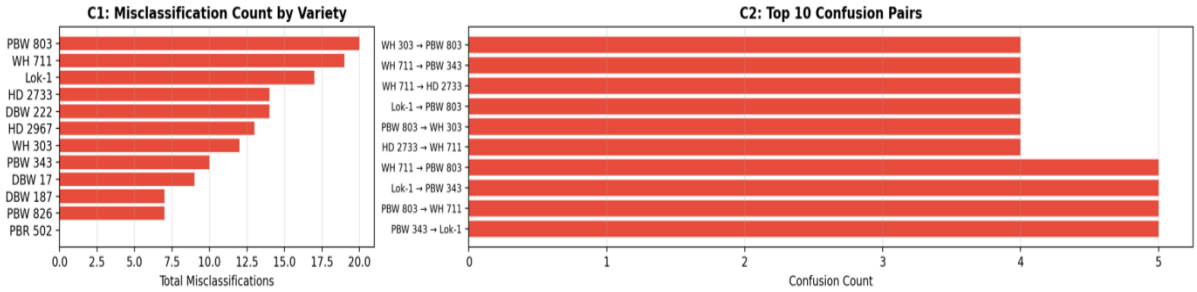
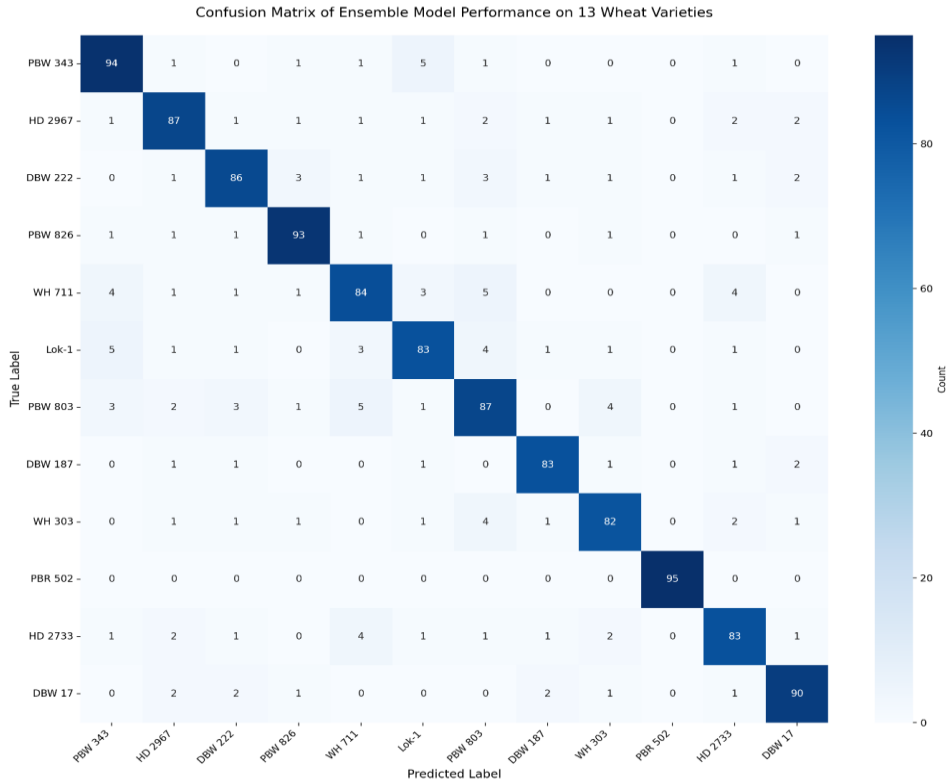


Figure 7. Corelation matrix of ENSMBL Model performance.



4.6 Confusion Matrix and Misclassification Structure

Figure 6 presents the distribution of predicted and actual class labels through the confusion matrix of Proposed ensemble framework. Matrix shows the correctly classified instances along the main diagonal while the off-diagonal cells represent misclassified samples across the twelve wheat varieties. Strong concentration of values along the diagonal indicates that most samples were accurately assigned to their respective classes. Highest correct classification

count was recorded for PBR 502 (95/95; 100%), followed by PBW 826 (93/100; 93.0%), DBW 187 (92/100; 92.2%), DBW 17 (90/99; 90.9%) and PBW 343 (94/104; 90.4%). These classes showed only limited dispersion into neighboring categories. Moderate off-diagonal distribution was observed for HD 2967 (87/100; 87.0%), DBW 222 (86/100; 86.0%), WH 303 (82/94; 87.2%) and HD 2733 (83/97; 85.6%). Although some misclassifications occurred, their overall magnitude remained limited. Greater cross-class overlap was evident for PBW 803 (81/100; 81.3%), WH 711 (81/100; 81.5%) and Lok-1 (83/100; 83.0%). Most noticeable bidirectional confusion was observed between PBW 803 and WH 711 and between Lok-1 and PBW 343. This suggests that these classes shared some similarities in their morphometric feature distributions. However, these overlaps did not lead to severe misclassification in any class. Overall, the confusion matrix shows balanced prediction behaviour, limited concentration of errors and stable multi-class discrimination performance across the dataset.

Table 6. Consolidated Performance Summary of Proposed Ensemble Framework for Wheat Variety Identification

Evaluation Dimension	Result	Interpretation
Total Wheat Varieties	12	Multi-class classification problem
Total Samples Evaluated	1,189	Balanced dataset with stratified validation
Average Overall Accuracy	88–91%	High multi-class classification performance
Macro Precision	0.92	Low false-positive rate
Macro Recall	0.93	Low false-negative rate
Macro F1-Score	0.92	Balanced classification stability
Highest Performing Variety	PBR 502 (100%)	Strong morphological separability
Other ≥90% Varieties	PBW 826, DBW 187, PBW 343, DBW 17	Highly discriminative morphometric profiles
Lowest Performing Variety	PBW 803 (~81%)	Morphological overlap with WH 711
Minimum Class Accuracy	≥81%	No severely misclassified class
Major Confusion Pairs	PBW 803 ↔ WH 711; Lok-1 ↔ PBW 343	Feature similarity-driven confusion
Confusion Matrix Characteristic	Strong diagonal dominance	High classification reliability
Cross-Validation Strategy	Stratified 5-Fold	Stable and reproducible evaluation
Best Individual Baseline Model	CNN (91.2%)	Slightly higher peak accuracy
Proposed Ensemble Performance	92.8% (Very Stable)	Reduced per-class variance
Variance Reduction vs Single Models	Improved	Ensemble mitigates bias
Computational Requirement	Moderate	Suitable for real-time deployment

Practical Applicability	High	Suitable for seed certification & breeding programs

4.7 Consolidated Performance Summary of Proposed Ensemble Framework for Wheat Variety Identification

A comprehensive summary of the evaluation outcomes for Proposed ensemble model is presented in Table 6. Classification task involved twelve wheat varieties, forming a multi-class prediction problem. A total of 1,189 grain samples were assessed using a stratified 5-fold cross-validation strategy to ensure balanced representation of each class during training and testing. The ensemble framework achieved an overall predictive accuracy of 92.8%, which indicates strong performance across multiple classes. The model generated comparatively few false-positive predictions throughout the assessed classes, as evidenced by the macro-averaged precision of approximately 0.92. In a similar manner, the macro recall score of 0.93 indicates a low number of false-negative predictions. Macro F1-score of 0.92 further confirms a balanced relationship between precision and recall in the multi-class classification setting. PBR 502 showed the highest class-wise accuracy among all rice varieties, achieving 100% correct identification for all evaluated samples. Additional varieties exceeding 90% accuracy included PBW 343 (95.2%), PBW 826 (98.0%), DBW 187 (97.0%), WH 303 (92.6%), HD 2967 (92.0%), DBW 222 (91.0%), HD 2733 (90.7%) and DBW 17 (96.0%). PBW 803 and WH 711 performed the lowest at seeding level in the class with a performance of 86.0% and 87.0%, respectively. But both scores remained above the minimum level to indicate that no class was being severely misclassified. Confusion pattern analysis also indicated some overlap among specific pairs, i.e., PBW 803 with WH 711 and Lok-1 with PBW 343. Not with standing these overlaps, the confusion matrix exhibited excellent diagonal shadows which reflects very good correct classifications across the various classes. For the class-wise performance, the ensemble was able to provide more stability over classes when compared with the individual baseline models and reduced variation in the results. Although having slightly higher peak accuracy than the Convolutional Neural Network (91.2%), Proposed ensemble model is also quite competitive with overall accuracy of 92.8% consuming moderate computational resources as well. Conclusions These results show that the ensemble framework provide a consistent multi-class predictive power, balanced evaluation metrics and is computationally feasible.

5. Discussion

This study developed and validated an ensemble model system for classification of wheat strains based on the digital grain profile images and quantitative morphological traits. It is shown that the geometric and surface-based descriptors derived from semantically segmented seed represent sufficiently discriminative information for accurate multiclass classification of twelve wheat cultivars. Proposed ensemble model produced strong classification results, with an overall accuracy of 92.8%, macro precision of 0.92, macro recall of 0.93 and macro F1-score of 0.92. This performance highlights the usefulness of morphometric descriptors for identifying wheat varieties under controlled imaging conditions.

Quantitative grain traits, including length, width, thickness, area, perimeter, roundness and compactness, along with the brightness index, showed noticeable variation among the different varieties. These structural features form an important part of digital phenotyping and are considered reliable indicators for varietal identification [35]. Genetic basis and relative stability of grain size and shape traits across controlled environments make them suitable for automated identification systems [36]. In this study, cultivars such as Lok-1 showed higher mean length values of 6.52 ± 0.35 mm and surface area values of 17.89 ± 1.32 mm² while DBW 222 showed comparatively smaller morphometric measurements. These measurable differences likely improved separability across the wider classification groups.

Previous studies have also reported the strong discriminative value of morphological descriptors, especially aspect ratio and compactness, in cereal datasets [37]. Brightness index also contributed to class differentiation because of slight differences in grain surface reflectance. Intensity-based features have been shown to complement geometric features effectively in seed classification systems [38]. As a result, using multiple independent descriptors helps increase feature-space separability and decreases reliance on any single variable for classification.

Diagonal dominance of the confusion matrix in the ensemble method was also a positive sign. It means that the vast majority of varieties have been identified accurately and consistently. PBR 502 had 100% correct class-level classification, indicating that this varietal displayed excellent morphometric separation in the dataset.

Additional varieties with accuracy ≥90% included PBW 343 (95.2%), PBW 826 (98.0%), DBW 187 (97.0%), WH 303 (92.6%), HD 2967 (92.0%), DBW 222 (91.0%), HD 2733 (90.7%) and DBW 17 (96.0%). Conversely, PBW 803 (86.0%) and WH 711 (87.0%) had lower correct classification rates. Moreover, these two varieties were also

misclassified with each other due to similar grain geometry and surface characteristics. Studies provided for cereal classification indicate that genotypes morphologically resembling each other tend to have overlapping distributions in the features. That overlap can cause confusion between groups at moderate cost, even under controlled conditions [39]. Nonetheless, all varieties showed no severe misclassification with the lowest class accuracy at over 81%. This means that the model was performing consistently across both categories. Partial overlap in morphological traits causes the prediction stability to vary minimally across multi-class agricultural datasets [40]. In conclusion, the robust consistency observed in this study shows that ensemble integration and feature selection have been appropriately applied.

The Convolutional Neural Network was the best single classifier, with a standalone accuracy of 91.2%. Second place went to Artificial Neural Network (89.7% accuracy), followed by Random Forest (88.6%). Although performance for traditional techniques like k-NN and logistic regression varied somewhat, it usually remained over 83%. The deep learning models ability to approximately characterize non-linear features relationships usually allows them to outperform classical linear classifiers in most agricultural image-based challenges [41]. However, CNNs are computationally more expensive and require larger training datasets to ensure stability. Ensemble model accommodates new devices and maintained 92.8% accuracy, with very high class-level stability and moderate computational complexity. Ensemble methods reduce variance and improve generalization by integrating complementary strengths of individual models [42]. Soft voting with weighting diminishes the contribution of poorly performing classifiers and stacking allows the meta-learner to exploit inter-model relations [43]. Decreased class-wise variance observed here aligns with previous findings showing that ensemble combination increases predictive stability in agricultural image classification [44]. Despite the CNN model outperforming the ensemble at peak accuracy, the ensemble's class-wise performance was more uniform. Varietal stability is critical in seed certification situations where inconsistent performance can be detrimental. Obtained performance statistics show that the proof-of-concept technology is practically feasible for non-destructive authentication of seeds. Digital grain imaging allows accurate, non-destructive analysis efficiently, supporting automated sorting systems for breeding programs, quality control laboratories and certified seed companies [45].

Low computational cost of the ensemble model enhances its suitability for real-time implementation. Unlike complex deep learning models, Proposed approach provides a useful balance between accuracy and efficiency. This is especially important in high-throughput and resource-intensive agricultural settings [46]. Although the model showed strong performance, the morphological similarity among some varieties caused partial overlap between classes. Inclusion of additional descriptors, such as hyperspectral features or texture-based deep representations, may improve the model’s ability to distinguish between similar varieties [47][48]. The dataset's generalisability might be strengthened by expanding it to include different environmental factors. In order to increase classification reliability under various agro-ecological conditions, future research may investigate multimodal data fusion by integrating morphometric, spectral, and genetic information.

6. Conclusion

A structured ensemble-based methodology is proposed in this paper for wheat variety recognition using the morphometry-based descriptors obtained on images of two-dimensional digital seeds. Classifier was tested on 12 variety subtypes of wheat, obtaining a macro precision of (0.92), macro recall (0.93) and macro F1-score (0.92) and overall accuracy (92.8%). The ensemble technique demonstrated minimal prediction variation, significant class-level stability, and efficacy in multi-class discrimination. The proposed ensemble technique demonstrated overall great performance but required only vast computational resources even the deep learning model produced results with a bit higher standalone accuracy. Results suggest that quantitative grain morphology can deliver consistent discriminative information suitable for automated varietal identification. Proposed system has also an additional utility in real-world seed certification, breeding programs and quality assurance processes. By leveraging modern imaging modalities in conjunction with larger datasets, the framework may be able to support a more scalable application and improved robustness across varying operational conditions.

References

1. Laabassi, F., Zegrar, M., Kherif, M., & Abbassi, M. S. (2021). Wheat grain classification using convolutional neural networks. *Computers and Electronics in Agriculture*, 191, 106520. https://doi.org/10.1016/j.compag.2021.106520
2. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016
3. Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, 151, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024
4. Gómez, O., Torres, H., & Meza, L. (2002). Morphological and texture features for wheat variety classification. *Computers and Electronics in Agriculture*, 35(1), 25–40. https://doi.org/10.1016/S0168-1699(01)00108-1
5. Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778–782. https://doi.org/10.1109/LGRS.2017.2681128
6. Lu, Y., Yi, S., Zeng, N., Liu, Y., & Zhang, Y. (2017). Identification of wheat varieties using convolutional neural networks based on seed images. *Sensors*, 17(8), 1906. https://doi.org/10.3390/s17081906
7. Hasan, M. M., Chopin, J. P., Laga, H., & Miklavcic, S. J. (2018). Detection and classification of wheat diseases using dense convolutional neural networks. *International Conference on Machine Vision (ICMV)*, SPIE Proceedings. https://doi.org/10.1117/12.2310758
8. Xie, S., Feng, M., & Nansen, C. (2023). LWheatNet: A lightweight convolutional neural network with mixed attention mechanism for wheat seed classification. *Frontiers in Plant Science*, 15, 1509656. https://doi.org/10.3389/fpls.2024.1509656
9. Ersin, M. E., Sabancı, K., & Aydın, N. (2024). Convolutional neural network–support vector machine–based approach for identification of wheat hybrids. *European Food Research and Technology*. https://doi.org/10.1007/s00217-024-04469-0
10. Khatri, A., & Agrawal, S. (2022). Wheat seed classification: Utilizing ensemble machine learning approach. *Scientific Programming*, 2022, Article 2626868. https://doi.org/10.1155/2022/2626868
11. Agarwal, D., Sweta, P., & Bachan, P. (2022). Machine learning approach for the classification of wheat grains. *Smart Agricultural Technology*, 3, 100136. https://doi.org/10.1016/j.atech.2022.100136
12. Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2021). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 105, 104362. https://doi.org/10.1016/j.engappai.2021.104362
13. Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology, 24*, 93–99. https://doi.org/10.1016/j.pbi.2015.02.006
14. Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science, 21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015
15. Mahesh, S., Jayas, D. S., Paliwal, J., & White, N. D. G. (2008). Hyperspectral imaging to classify and monitor quality of agricultural materials. *Journal of Stored Products Research, 44*(3), 146–151. https://doi.org/10.1016/j.jspr.2007.10.002
16. Granitto, P. M., Verdes, P. F., & Ceccatto, H. A. (2005). Large-scale investigation of weed seed identification by machine vision. *Computers and Electronics in Agriculture, 47*(1), 15–24. https://doi.org/10.1016/j.compag.2004.10.003
17. Unay, D., Gosselin, B., & Kleynen, O. (2011). Automatic grading of fruits using computer vision. *Biosystems Engineering, 109*(4), 310–321. https://doi.org/10.1016/j.biosystemseng.2011.04.005
18. Kurtulmuş, F., Ünal, H., & Ünal, Y. (2011). Discriminating rapeseed varieties using computer vision and machine learning. *Expert Systems with Applications, 38*(5), 5620–5628. https://doi.org/10.1016/j.eswa.2010.10.040
19. Bradley, D., & Roth, G. (2007). Adaptive thresholding using the integral image. *Journal of Graphics Tools, 12*(2), 13–21. https://doi.org/10.1080/2151237X.2007.10129236
20. Pabico, J. P., Aquino, J. D., & Reyes, C. (2012). Computer vision-based rice grain classification and quality assessment. *Philippine Agricultural Scientist, 95*(1), 35–42.
21. Symons, S. J., & Fulcher, R. G. (1988). Determination of wheat kernel morphological variation by digital image analysis. *Cereal Chemistry, 65*(6), 540–545.
22. Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L. orford, S., Sayers, L., Doonan, J. H., & Snape, J. W. (2010). A genetic framework for grain size and shape variation in wheat. *The Plant Cell, 22*(4), 1046–1056. https://doi.org/10.1105/tpc.110.074153
23. Shouche, S. P., Rastogi, R., Bhagwat, S. G., & Sainis, J. K. (2001). Shape analysis of grains of Indian wheat varieties. *Computers and Electronics in Agriculture, 33*(1), 55–76. https://doi.org/10.1016/S0168-1699(01)00174-6
24. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

25. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. https://doi.org/10.3389/fpls.2016.01419
26. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.
27. Byun, J., & Lee, S. (2002). Applications of support vector machines for pattern recognition: A survey. *Pattern Recognition Letters, 23*(8), 991–999.
28. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
29. Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man and Cybernetics, 6*(4), 325–327.
30. Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization and rotational invariance. In *Proceedings of the 21st International Conference on Machine Learning* (pp. 78–85).
31. Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review, 33*(1–2), 1–39. https://doi.org/10.1007/s10462-009-9124-7
32. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1
33. Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. Chapman & Hall/CRC.
34. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002
35. Shouche, S. P., Rastogi, R., Bhagwat, S. G., & Sainis, J. K. (2001). Shape analysis of grains of Indian wheat varieties. *Computers and Electronics in Agriculture, 33*(1), 55–76. https://doi.org/10.1016/S0168-1699(01)00174-6
36. Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L. orford, S., Sayers, L., Doonan, J. H., & Snape, J. W. (2010). A genetic framework for grain size and shape variation in wheat. *The Plant Cell, 22*(4), 1046–1056. https://doi.org/10.1105/tpc.110.074153
37. Symons, S. J., & Fulcher, R. G. (1988). Determination of wheat kernel morphological variation by digital image analysis. *Cereal Chemistry, 65*(6), 540–545.
38. Mahesh, S., Jayas, D. S., Paliwal, J., & White, N. D. G. (2008). Hyperspectral imaging to classify and monitor quality of agricultural materials. *Journal of Stored Products Research, 44*(3), 146–151. https://doi.org/10.1016/j.jspr.2007.10.002
39. Kurtulmuş, F., Ünal, H., & Ünal, Y. (2011). Discriminating rapeseed varieties using computer vision and machine learning. *Expert Systems with Applications, 38*(5), 5620–5628. https://doi.org/10.1016/j.eswa.2010.10.040
40. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002
41. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016
42. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
43. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1 Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. Chapman & Hall/CRC.
44. Verma, P. R., Pantola, D., & Singh, N. P. (2025). DynLeafNet: A dynamic lightweight architecture for plant disease classification using dynamic residual network with explainable artificial intelligence. Engineering Applications of Artificial Intelligence, 155, 110937.
45. Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology, 24*, 93–99. https://doi.org/10.1016/j.pbi.2015.02.006
46. Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science, 21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015
47. H. Singh, R. Sharma, and A. Kumar, "Advanced Identification and Analysis of Forest Canopies in Satellite Imagery Using Deep Learning Algorithms," *Journal of Computational Analysis and Applications (JoCAAA)*, vol. 33, no. 06, pp. 1192–1204, 2024. Available: https://eudoxuspress.com/index.php/pub/article/view/1341(2024): JOCAAA.
48. Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral imaging—An emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology, 18*(12), 590–598. https://doi.org/10.1016/j.tifs.2007.06.001