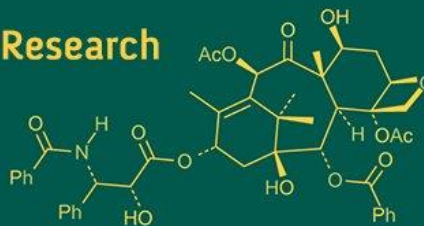
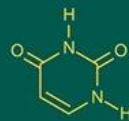
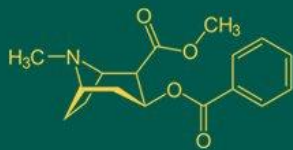


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2025): 5.29
IJABR 2025; SP-9(11): 764-771
www.biochemjournal.com
Received: 23-09-2025
Accepted: 28-10-2025

Vidya Theertha VP
P.G. Scholar, Department of
Seed Science and Technology,
College of Agriculture,
Vellayani, Kerala, India

Arivusudar S
P.G. Scholar, Department of
Seed Science and Technology,
College of Agriculture,
Vellayani, Kerala, India

Beena R
Associate Professor and Head,
Department of Seed Science
and Technology, College of
Agriculture, Vellayani, Kerala,
India

Nivedhitha MS
Assistant Professor,
Department of Seed Science
and Technology, College of
Agriculture, Vellayani, Kerala,
India

Corresponding Author:
Vidya Theertha VP
P.G. Scholar, Department of
Seed Science and Technology,
College of Agriculture,
Vellayani, Kerala, India

Integration of machine learning in seed quality analysis

Vidya Theertha VP, Arivusudar S, Beena R and Nivedhitha MS

DOI: <https://www.doi.org/10.33545/26174693.2025.v9.i11Sj.6313>

Abstract

Seed quality assessment is essential for ensuring high crop productivity, maintaining varietal integrity, and strengthening global food security. Traditional seed testing methods such as germination testing, purity analysis, and moisture determination are often destructive, labor-intensive, and prone to human error, limiting their efficiency and reliability. Recent advancements in artificial intelligence (AI), machine learning (ML), deep learning (DL), and computer vision have transformed seed quality evaluation into a faster, more accurate, and non-destructive process. This review summarizes the principles of ML, key algorithmic categories, and their applications in seed quality analysis. Modern ML models including Support Vector Machines, Random Forests, Naïve Bayes, k-Nearest Neighbours, Artificial Neural Networks, and Convolutional Neural Networks demonstrate strong performance in seed classification, defect detection, vigor estimation, and variety identification. Integration of ML with multispectral, hyperspectral, terahertz, X-ray, and RGB imaging has enabled precise germination prediction, moisture estimation, and physical purity discrimination across diverse crop species. AI-based platforms such as SeedGerm, ASeed, and deep learning-driven phenotyping tools further enhance high-throughput and real-time analysis. Overall, the convergence of machine learning and advanced imaging technologies offers a transformative approach to seed quality assessment, delivering rapid, reliable, and non-destructive solutions that surpass traditional methods and support more efficient, data-driven agricultural systems.

Keywords: Machine learning, seed quality, germination, moisture content, and physical purity

Introduction

The main requirement of farmers for the cultivation of crop is high quality seeds. Seed quality analysis is crucial for enhancing agricultural productivity and ensuring food security. High quality seeds significantly influence the yield of the crop and poor seed selection leads to yield loss (Mogilicharla, *et al.*, 2024) [25]. As a result, different traditional seed quality tests including mechanical tests, tests of genetic purity, seed germination and vigour, and seed health tests are used. (McDonald, 1998) [23]. But traditional seed quality analysis methods face several limitations including irreversible destruction of seeds, high costs, complexity of the method and inability to perform real time assessment which are important for modern agricultural practices (Li *et al.*, 2023) [20]. These methods also often rely on human intelligence and lack validation leading to unstable seed quality prediction (Srinivasaiah *et al.*, 2023) [32]. To overcome these challenges, innovative methods using machine learning and advanced imaging techniques have been developed for seed quality assessment (Mekala, *et al.*, 2024) [24]. Machine learning is increasingly being utilized in seed analysis, enhancing the efficiency and accuracy of seed quality assessment (Cutler *et al.*, 2024) [7]. Machine learning algorithms especially convolutional neural networks are used to predict seed quality facilitating better crop yield outcomes (Srinivasaiah *et al.*, 2023) [32]. These advancements collectively enhance the seed quality assessment and decision-making in agriculture.

Artificial intelligence (AI) is rapidly becoming the fundamental aspect of current research. The field of artificial intelligence involves designing machines that can perform tasks that usually require human cognitive functions such as learning, decision making and problem solving. Artificial intelligence and machine learning are often used synonymously but machine learning is a broader subset within the field of AI. Machine learning is a subset of artificial intelligence that involves techniques that enable software to learn independently and optimize its performance through exposure to increasing amounts of data. It is seen as an

approach to achieve artificial intelligence. Deep learning is a subset of machine learning based on deep neural network. (Misilmani *et al.*, 2019)^[8]

Machine learning can be defined as the process of developing computer systems that have the capacity to improve through experience and implement a learning process. (Ayodele, 2010)^[4]. The aim of machine learning is to build computer programs that can learn from data. A machine learning system stores the acquired knowledge in a structure called hypothesis, which typically takes the form of a model. According to Ockham's razor principle, the hypothesis should generalize the training data in the form of the simplest explanation. For a valid generalization, the hypothesis must be simpler than the data itself. The learning algorithm outlines how to update the hypothesis with new experience, that is training data, to optimize performance on the specific task.

A machine learning model is composed of six components and each component has a specific task.

Collection and preparation of data

The main focus of this component is to collect and prepare the data. Significantly large amount of data may be available for addressing any problem. The unstructured nature of the web data makes it irrelevant and redundant. This generates the need to clean and pre-process the data to obtain a structured format that can be given to the algorithm as input.

Feature selection

The learning process does not require all the features present in the data obtained from the first step. The irrelevant features are removed and only the most important features are retained.

Table 1: The three main categories of learning (Misilmani *et al.*, 2019)^[8]

Learning category	Description
Supervised Learning	A model trained on a data set. Predictions are made on new inputs
Unsupervised Learning	A pattern is derived from the data after exploring it
Reinforcement Learning	Model takes decisions and learns from its actions

Supervised Learning

In supervised learning the models are trained using data where a label is provided for each observation in the dataset. This label may be categorical variable or a continuous variable. (Maleki *et al.*, 2020). Supervised learning is also known as learning through examples or learning from exemplars because under SL a set of examples or training modules are provided with the correct outputs and on the basis of these training sets, the algorithm learns to respond more accurately by comparing its output with those that are given as input. (Alzubi *et al.*, 2018)^[3].

Supervised learning algorithms include linear regression, logistic regression, artificial neural networks and support vector machines.

- **Linear regression:** It consists of including a continuous linear function through the data from which the algorithm can make prediction on new inputs.
- **Logistic regression:** This is used in classification tasks where it predicts the probability that a particular input corresponds to one of the known classes.
- **Artificial Neural Networks:** Neural networks consist of large interconnections of neurons, which are computer cells, and these neurons are used to achieve good performance. Neural networks are composed of multiple layers, the input layer, output layer and hidden layers between the input and output layers. Complex

Choice of algorithm

The machine learning algorithm that is best suited for addressing the problem at hand to get the best possible results is selected.

Selection of models and parameters

Most of the machine learning algorithms demand some initial manual intervention to set appropriate values for various parameters.

Training

After selecting the appropriate algorithm and the suitable parameter values, the model is trained. The dataset is divided into data for training, validation and testing. The model is trained using the dataset set aside as training data.

Performance evaluation

The model is then tested using a new dataset to evaluate the performance of the system in terms of accuracy, precision and recall before real-time implementation of the system. (Alzubi *et al.*, 2018)^[3]

Types of machine learning algorithms

Machine learning is based on algorithms that can learn from data without relying on rule-based programming. It can be divided into three categories. (Misilmani *et al.*, 2019)^[8]

- Supervised learning (SL)
- Unsupervised learning (UL)
- Reinforcement learning (RL)

functions with many features can be carried out using neural networks as they offer an alternate path to perform machine learning.

- **Support Vector Machines (SVM):** This type of supervised learning algorithm is used in classification and deals with more difficult cases of non-linearly separable patterns by using kernel methods.
- **k Nearest Neighbors:** This is considered as one among the simplest of all machine learning algorithms. After memorizing the training set, the output of any new input is predicted by the algorithm based on the output of its closest neighbors in the training set. (Misilmani *et al.*, 2019)^[8]

Unsupervised Learning

The aim of unsupervised learning is to find associations and patterns among input datasets (Maleki *et al.*, 2020). In unsupervised learning, the algorithm derives a structure from the data after identifying similarities in the input. (Misilmani *et al.*, 2019)^[8]. This technique is used when the categories of data are unknown. The training data for unsupervised learning is not labeled. Unsupervised learning is a statistic-based approach of learning and addresses the problem of identifying hidden structure in unlabeled data. (Alzubi *et al.*, 2018)^[3].

Unsupervised learning algorithms include K-Means Clustering and Dimensionality Reduction Algorithms.

- **K-Means Clustering:** It is a type of unsupervised learning where variables in the data are grouped together based on relationships among them.
- **Dimensionality Reduction Algorithms:** The goal of this algorithm is to minimize projection errors by reducing every feature's distance to a certain projection line. E.g. Principle Component Analysis Algorithm (PCA) (Misilmani *et al.*, 2019)^[8].

Reinforcement learning

Reinforcement learning is an intermediate type of learning because the algorithm is only provided with a response that tells whether the output is correct or not. The algorithm has to explore and rule out various possibilities to get the correct output. It is regarded as learning with a critic as the algorithm doesn't propose any suggestion or solution to the problem. (Alzubi *et al.*, 2018)^[3]. In this type of learning, the machine does not receive any labeled data. The information is collected after interacting with the environment through different actions. The machine is rewarded after each action and its objective is maximizing this expected average reward where the action would become optimal. An example of this type of learning is Markov Decision Process (MDP) (Misilmani *et al.*, 2019)^[8].

Important algorithms in seed quality analysis

Artificial Neural Networks (ANNs)

One of the most broadly used algorithm in seed recognition research is ANN (Chaugule, 2021)^[6]. Though there are various ANN models, the basic principle is similar. An ANN model consists of several neurons which act as components that process the signals and these neurons are connected by synapses to form a channel. An ANN receives input signals, processes the signal and finally produces an output signal. Different types of ANNs have been developed for particular problems and applications. The most popular ANNs used in seed recognition and classification are Kohonen networks, multilayer perceptron (MLP), deep neural networks (DNN) and convolutional neural networks (CNN) (Basheer, *et al.*, 2000)^[5].

Kohonen network

This algorithm is also known as self-organizing maps (SOM). It is a type of unsupervised learning algorithm. It comprises a two-layer network in which the input and output layers are fully connected. The primary function of SOM is to map the input data into a two-dimensional grid of output neurons. During this process, the neurons in close proximity on the grid represent similar pattern or feature from the input data. This results in the formation of a 2D map where the similar patterns are located near each other (Kohonen, 1989)^[18]. Kohonen networks are employed in clustering, pattern recognition, image segmentation and classification. Ghamari (2012) used SOM to identify chickpea seed varieties and it showed better performance with an accuracy of 79 percent compared to supervised ANN which had an accuracy of 73 percent. The training and identification time of the Kohonen network is moderately short but the error rate is very high since they are applied without supervision.

Multilayer perceptron (MLP)

MLP is a class of feed-forward artificial neural network. MLP is a non-linear computational model effective in handling complex classification and regression task. It solves non-linear classification problems by introducing multiple layers of neurons between the input and output layer known as hidden layers. This layer processes the information received from the input layer and pass it on to the output layer. The output layer develops perceptrons (complex decision boundaries) to resolve non-linear classification problems (Hecht- Nielsen, 1989). However, MLP is a time-consuming method and it only has few hidden neurons so it is not suitable for modeling and predicting (Geetha, 2020)^[9].

Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN) operates in stages in which each layer goes through three main steps: convolution, detection and pooling.

- **Convolution stage:** In this stage the network takes the input data and process it by applying filter. These filters slide over the input data to create feature maps, which highlight certain features like edges or textures in the data. All the units in the filter map share the same filter.
- **Detection stage:** After convolution the network applies a non-linear function called Rectified Linear Unit (ReLU) to the result. This adds non-linearity to the model which helps to capture more complex patterns in the data.
- **Pooling stage:** Finally pooling function simplifies the output by summarizing the information from nearby units in the feature maps. This reduce the dimensionality of the data (Goodfellow *et al.*, 2016)^[12]. CNN have successful applications in agriculture especially for the identification and classification of unclear feature data like the delicate features of small seeds (Patrício, *et al.*, 2018)^[19]. In 2018 Uzal and coworkers used CNN for computing the number of seeds in soybean pods. CNN and SVM model were compared in this project and the result showed that CNN has higher accuracy (86.2%) than SVM (50.4%) in seed-per-pod estimation.

Support Vector Machines (SVMs)

SVM was introduced by Vapnik in 2000 and it is considered as one of the most prevailing and simple machine learning algorithms. Based on the output variable, SVM can be classified as Support Vector Classification (SVC) which classifies the data and Support Vector regression (SVR) which determine regression (Singh *et al.*, 2016)^[30]. SVM is usually applicable to a two-class problem that create a boundary between two groups in linear and non-linear parameters. In non-linear relationships, SVM discover patterns and performances. SVM was used for discrimination and classification. For soybean seed discrimination the morphological and colour attributes of different seed classes were analyzed by SVM. The results showed that colour trait had better discrimination ability than morphological traits with an accuracy of 77 percent and 59 percent respectively (Namias *et al.*, 2012)^[27]. SVM classifier was also used to detect defects in corn seeds using colour and texture analysis. The best accuracy (81.8%) in results were obtained when the combination of both colour

and texture was used for analysis than colour and texture individually (Kiratiratanapruk *et al.*, 2011)^[17].

Random forest (RF)

A random forest is composed of many individual decision trees, each of which is built to make decisions based on the data provided. All these decision trees are constructed simultaneously rather than sequentially. Bootstrapping technique is used in RF to create different subsets of the original dataset to train each decision tree. Bootstrapping is a statistical technique in which multiple random samples are taken from a dataset to create a subset of data. Bagging (short for bootstrap aggregation) is done which is a technique that improves the stability and accuracy of machine learning algorithms by combining the results of the decision trees. Each decision tree is trained on different bootstrapped sample of the dataset. Out-of-bag observations are the data points not included in the bootstrapped sample used to train a particular decision tree.

Since these data points are not used in training, they can be used to test the model's performance that can be used to estimate the accuracy of the model. The process of creating bootstrapped samples and identifying out-of-bag observations is repeated for each decision tree in the forest (Mokry *et al.*, 2013)^[26]. Random Forest along with SVM was used to discriminate soybean varieties based on colour and morphological features. The results showed that RF classifiers discriminated the colour features with a better accuracy (78%) than SVM model (77%) (Namias *et al.*, 2012)^[27]. Another research classified rice using image features of rice seed which includes colour, shape and texture. RF classifiers along with SVM were used in this research and the results showed better accuracy for the RF model (90.54%) than SVM.

Naïve Bayes (NB)

NB is a simple but effective classification algorithm based on Bayes theorem. It assumes that all the features in the dataset are independent of each other. In comparison with other classifiers like neural networks and SVMs, NB requires only small amount of data for training to build an effective model. This makes it useful in situations where data is limited. NB trains quickly compared to more complex models like neural networks. NB can struggle when the dataset contains a lot of irrelevant features. This negatively impact the model's performance. When the attributes are highly correlated NB's assumption of independence between the features is violated and can cause certain attributes to have undue influence on the model's decisions leading to classification bias and a decrease in accuracy for predictions involving these correlation features (Ali *et al.*, 2023)^[2]. NB is usually used to recognize weed seeds based on morphological, colour and texture characteristics from images. The NB classifier outperformed ANN algorithms in weed seed identification (Granitto *et al.*, 2005)^[13]. It was also used in seed classification of kama, rose and Canadian wheat varieties according to their morphological features. However, ANN showed highest performance (95.2%) than NB (94.3%) (Ajaz *et al.*, 2015)^[1].

Artificial intelligence in seed germination analysis

Use of machine learning, deep learning and computer vision have improved the accuracy of germination percent determination and seed quality prediction compared to

traditional methods. The three steps involved in germination analysis using artificial intelligence are image acquisition, feature extraction and model-based prediction. Morphological features are the most commonly used parameters in classical machine learning models that used classifiers such as k-nearest neighbours (KNN) and decision trees. Silva and coworkers (2024) developed a model to predict the germination percent of tomato with an accuracy of 94% using phenotypic characters.

Deep learning techniques such as convolutional neural networks (CNN) are also used in germination percent determination apart from classical machine learning models. Genze *et al.* (2020)^[39] developed a model based on CNN with a mean average precision (mAP) of 94-98% to detect the germination stages in maize, rye and pearl millet. 98.14% accuracy was achieved by a corn germination detector developed by combining CIE LAB segmentation with CNNs and k-fold validation. This model provides accurate results even under varying environmental conditions due to its capacity to process thousands of images (Grant *et al.*, 2023)^[40]. A study was conducted to train five machine learning models such as Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Random Forest Classifier (RFC), and Artificial Neural Network (ANN) using temporal and spatial data of 800 viable and non-viable soybean seeds. Among the five models, ANN model showed superior performance achieving an accuracy of 97.65%, sensitivity of 97.67% and specificity of 97.97%. (Thakur *et al.*, 2022)^[41]. Enhanced seed vigour and viability prediction is also obtained by combining deep learning models with multispectral and hyperspectral imaging data. Multispectral imaging and a novel deep learning model called MsiFormer was used to develop a non-destructive model for accurate rice seed vigor testing. MsiFormer achieved a high accuracy of 94.17% in rice seed germination prediction and outperformed other deep learning methods by 2.5% to 18.34% (Qiao *et al.*, 2023)^[42]. Apart from multispectral imaging hyperspectral imaging is also used to assess the viability of aged rice seeds (Jin *et al.*, 2022)^[43]. SeedGerm is a cost-effective platform which uses automated imaging and machine learning for crop seed germination analysis. It combines accessible hardware and open source software to measure germination related traits in different species of crops including tomato, pepper and barley accurately. It was also used as a research tool to identify the genetic factors that influence the germination in Brassica napus (Colmer *et al.*, 2020)^[44]. DiSCount (Digital Striga Counter) is a computer vision tool and it uses deep learning to quantify total number of germinated seeds of Striga from images. This software is able to increase the speed of analysis. When a CPU is used, it only takes 3seconds to analyse an image while a GPU takes only 0.1 second per image. This method is accurate and have an average error of 3.38 percent (Masteling *et al.*, 2020)^[45]. Seed germination in wide range of vegetables such as tomato, pepper, watermelon, cabbage and cauliflower can be predicted by combining deep learning with RGB image analysis (Nehoshtan *et al.*, 2021)^[46]. Viability of sweet corn seeds were successfully assessed non-destructively using hyperspectral imaging combined with deep learning algorithms. The proposed Firefly Algorithm optimized CNN-LSTM and it outperformed traditional machine learning algorithms in classifying seeds according to seed viability. This model achieved the highest

classification accuracy of 97.23 percent (Wang & Song, 2024). On examining wheat seeds it was found that there are distinct spectral differences between germinating and non-germinating seeds. Among the models tested, the SNV-SPA-PLS-DA model which used mixed spectral dataset from both sides of the seeds recorded highest classification accuracy of 88.9 percent confirming that hyperspectral imaging is a reliable technology for non-destructive analysis of seed viability (Zhang *et al.*, 2018) ^[48]. A novel detection algorithm called DDST-CenterNet achieved an accuracy of 96.57 percent and have the capacity to detect over 40 seeds (Peng *et al.*, 2022) ^[49]. The integration of machine learning, deep learning, and advanced imaging technologies has greatly enhanced the precision and efficiency of germination percent determination and seed quality prediction. By automating image acquisition, feature extraction, and predictive modelling, these approaches deliver high accuracy across crops and conditions, far surpassing traditional manual assessments. Together, these innovations highlight the growing role of AI-driven methodologies in establishing fast, reliable, and non-destructive germination testing systems.

Artificial intelligence in seed moisture content measurement

Seed moisture content is an important factor that determines seed quality, storability and germination potential. Traditional methods of seed moisture determination such as oven drying is time consuming and destructive. So, integration of artificial intelligence with modern sensor technologies will enhance the accuracy and also save time. Free space measurement data was used to determine the moisture content of flowing grains using machine learning algorithms. Among the models tested to determine the moisture content of bulgur wheat, Support Vector Regression (SVR) model performed the best and K-Nearest Neighbour (KNN) algorithm recorded most accurate results for durum wheat and corn silage kernel (Yigit & Duysak, 2022) ^[50]. Moisture content in rice was predicted by different machine learning models using Received Signal Strength Indicator (RSSI) data from Wireless Sensor Network (WSN) and Radio Frequency Identification (RFID) systems. Random Forest model recorded the highest accuracy of 87 percent when using RSSI from WSN and 99 percent using RSSI from both WSN and RFID (Azmi *et al.*, 2021) ^[51]. The advancement of deep learning has led to the use of neural network architectures such as multilayer perceptrons (MLPs), stacked autoencoders (SAEs), and convolutional neural networks (CNNs) for direct feature extraction and end-to-end prediction. Deep learning and machine vision were used to predict moisture content in wood chips through RGB images. Image dataset consisting of 1600 RGB images of wood chips was collected and labelled. This was fed into two high performing neural networks, MoistNetLite and MoistNetMax developed using Neural Architecture Search (NAS) and hyperparameter optimization. The results showed that MoistNetLite achieved an accuracy of 87 percent and MoistNetMax obtained exceptional precision with an accuracy of 91 percent in wood chip moisture content determination (Rahman *et al.*, 2024) ^[52]. SG-SAE-LightGBM model, which used hyperspectral imaging to evaluate sunflower seed vitality and moisture content prediction, achieved an accuracy of 98.65 percent for vitality classification. SNV-

XGBoost-LightGBM model recorded better performance for moisture content prediction with a coefficient of determination of 0.9715 (Yuan *et al.*, 2024) ^[53]. A model that measure the grain moisture uses an improved Sparrow Search Algorithm and Support Vector Regression (SSA-SVR) to enhance the precision of moisture detection in long grain rice. The model measures the moisture content of the grain by measuring the electrical resistance of the grains. SSA-SVR model was found to be more accurate in moisture determination than other algorithms such as Ridge Regression, MLP, and Random Forest (Cao *et al.*, 2024) ^[54]. A portable device for measuring moisture content of paddy was developed using microwave microstrip sensors and machine learning among the different machine learning models tested, Random Forest achieved the highest accuracy with a coefficient of determination of 0.99 (Liu *et al.*, 2022) ^[56]. A model which combines computer vision and machine learning was developed to predict the moisture content in fresh guava, guyabano and lagundi leaves. The system consists of a Raspberry Pi 4 model B, a camera module, 40 kg load cell and an Arduino Uno integrated into a stainless steel chamber. It captures the weight and visual features of the leaf samples and process them to predict the moisture content of the samples using Random Forest Regression model. (Gapusan *et al.*, 2024) ^[57]. A smartphone based method was developed to determine the grain moisture content using machine learning which is an important indicator to determine the optimal harvest timing. The models tested were Random Forest, Multilayer Perceptron, Support Vector Regression and Multivariate Linear Regression. Among these models, Support Vector Regression (SVR) was found to be the most suitable and accurate model with a mean absolute error of 1.23 percent (Yang *et al.*, 2021). An image analysis system that can accurately measure the dimensions of bean and corn seeds was developed. An experiment was conducted to confirm the relationship of the seed dimensions to the wet basis moisture content. An artificial neural network (ANN) was successfully used to predict the moisture content of the seeds based on the dimensions with an efficiency of 82 percent (Chen *et al.*, 2024) ^[59]. The integration of artificial intelligence with advanced sensing and imaging technologies has enabled fast, accurate, and non-destructive prediction of moisture content in seeds and agricultural materials. Across sensor-based, spectral, and image-driven approaches, machine learning and deep learning models consistently deliver higher precision than traditional methods, offering reliable real-time assessment. These developments collectively demonstrate the growing importance of AI-powered systems in improving seed quality evaluation and storage management.

Artificial intelligence in seed physical purity detection

Seed physical purity assessment plays a vital role in maintaining the quality and authenticity of agricultural production systems. Traditional manual evaluation methods are labor-intensive, time-consuming, and prone to human error. Recent studies have demonstrated the efficacy of multispectral and hyperspectral imaging for discriminating between pure seeds and contaminants. These methods rely on combining optical imaging modalities with statistical or machine learning classifiers to differentiate seed types. Alfalfa and sweet clover seeds, which are difficult to distinguish by morphology alone was effectively

differentiated using multispectral imaging combined with supervised discrimination methods such as Linear Discriminant Analysis (LDA), AdaBoost, and Support Vector Machine (SVM). The LDA model was found to be most effective which used a combination of spectral and morphological data and obtained a classification accuracy of 99.58 percent (Hu *et al.*, 2020) ^[60]. Rice seeds were distinguished from grains using derivative preprocessing and multivariate classifiers such as LDA, QDA, KNN and SVM achieving F1 scores ranging from 0.74 to 0.97 (Boateng *et al.*, 2025) ^[61]. A new Convolutional Neural Network model was studied for the classification of ten types of Brassica seeds and it obtained a high accuracy of 93 percent. The proposed model showed superior performance when compared with pre-trained models such as DenseNet121, InceptionV3 and ResNet152. Among the pre-trained models, DenseNet121 achieved the highest accuracy of 90.03 percent followed by InceptionV3 and ResNet152 with an accuracy 84.71 percent and 73.34 percent respectively (Elfatimi *et al.*, 2025) ^[62]. Purity of rice seeds were identified using a hybrid method that combines deep learning models for feature extraction and machine learning algorithm for classification. The combination of feature extraction from VGG16 with SVM or LR algorithms achieved accuracies above 95 percent upto 99 percent significantly outperforming the methods that only used deep learning or traditional machine learning (Thi-Thu-Hong *et al.*, 2024). A novel Fuzzy Cognitive Map (FCM) model combined with Levenberg-Marquardt (LM) deep learning algorithm was used to predict seed purity, variety and health. FCMLM deep learning model demonstrated superior performance with the highest correlation coefficient of 0.9 when compared to other statistical methods like BP-ANN and Multivariate Regression (Suganthi & Sathiaselvan, 2022) ^[64]. A Convolutional Neural Network (CNN) model was developed for identifying and analysing the purity of different seed types with a maximum accuracy of 89 percent. Initially, the accuracy of the model was low at 60 percent, but it improved as the number of epochs increased (Kalaivani *et al.*, 2022) ^[65]. AIseed software is based on machine vision technology and enables high-throughput handling and phenotyping of various-sized plant seeds. After phenotyping for a total of 54 features, AIseed numbers each seed to facilitate further seed quality assessment or grain detection. In addition, AIseed includes modules for seed quality detection and prediction, such as seed clarity, purity, vigor, and viability testing, which are based on machine learning or deep learning models developed by analyzing the associations between seed quality indicators and the features obtained by the software. Through a series of experiments, this paper has confirmed that AIseed has a high performance in extracting phenotypic features and testing seed quality from images for seeds of several plants of different sizes, with high speed and high precision in separating seeds from the background and then analyzing them (Keling *et al.*, 2023) ^[66]. A two stage method using visible and X-ray imaging is an effective, non-destructive and reliable alternative to conventional methods of physical purity, viability and vigour testing of soybean seeds. This method obtained highest correlation between the imaging test results and the standard germination tests confirming the usability of this technique in seed quality assessment (Mahajan *et al.*, 2018) ^[57]. Paddy seed varietal purity based on colour, morphological, and textural features were

analysed using machine vision combined with multivariate analysis. The Support Vector Machine Classification (SVM-C) model achieved the highest classification accuracy of 93.9 percent for distinguishing between three paddy varieties (BR 11, BRRI dhan 28, and BRRI dhan 29) compared to PLS-DA (83.8 percent) and KNN (87.2 percent) models. It was also noted that morphological features were more significant for classification compared to colour and textural features (Ansari *et al.*, 2021) ^[68]. Overall, recent advancements in multispectral imaging, deep learning, and hybrid machine vision systems have significantly improved the accuracy and efficiency of seed physical purity assessment. These approaches consistently outperform traditional methods by enabling rapid, non-destructive, and highly reliable discrimination of seed types across crops. Collectively, the findings highlight the growing potential of AI-driven imaging technologies to revolutionize seed quality evaluation in modern agriculture.

Conclusion

The integration of machine learning in seed quality analysis provides a more accurate and nondestructive analytical method. Machine learning also makes the assessment of seed characteristics such as germination potential, vigour, viability and purity faster compared to traditional methods. Additionally, the integration of machine learning in seed quality analysis has shown significant promise in enhancing accuracy and efficiency in seed defect detection and quality prediction. Overall the researches show the transformative impact of machine learning on seed quality analysis paving way for more efficient agricultural practices.

Conflict of interest

The authors don't have any conflict of interest in publishing this article

Author contribution

VT prepared the review. AS, BR and NMS have corrected the review. All authors have read the manuscript and approved for publication.

References

1. Ajaz RH, Hussain L. Seed classification using machine learning techniques. *Seed*. 2015;2(5):1098-1102.
2. Ali A, Mashwani WK. A supervised machine learning algorithms: Applications, challenges, and recommendations. *Proc Pak Acad Sci A Phys Comput Sci*. 2023;60(4):1-12.
3. Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: An overview. *J Phys Conf Ser*. 2018;1142:012012.
4. Ayodele TO. Machine learning overview. *New Adv Mach Learn*. 2010;2(9-18):16.
5. Basheer IA, Hajmeer M. Artificial neural networks: Fundamentals, computing, design, and application. *J Microbiol Methods*. 2000;43(1):3-31.
6. Chaugule A. Survey of seed classification techniques. *Turk J Comput Math Educ*. 2021;12(13):1236-1260.
7. Cutler S, Eckhardt J, Subramanian V, Xing Z, Vaidya A. Robotic imaging and machine learning analysis of seed germination: Dissecting the influence of ABA and DOG1 on germination uniformity. *bioRxiv*. 2024;2024-05.

8. El Misilmani HM, Naous T. Machine learning in antenna design: An overview on machine learning concept and algorithms. In: 2019 International Conference on High Performance Computing and Simulation (HPCS). IEEE; 2019. p. 600-607.
9. Geetha MCS. Forecasting the crop yield production in Trichy district using fuzzy C means algorithm and multilayer perceptron (MLP). *Int J Knowl Syst Sci*. 2020;11(3):83-98.
10. Genze N, Bharti R, Grieb M, Schultheiss SJ, Grimm DG. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods*. 2020;16:157.
11. Ghamari S. Classification of chickpea seeds using supervised and unsupervised artificial neural networks. *Afr J Agric Res*. 2012;7(21):3193-3201.
12. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press; 2016. 802 p.
13. Granitto PM, Verdes PF, Ceccatto HA. Large-scale investigation of weed seed identification by machine vision. *Comput Electron Agric*. 2005;47(1):15-24.
14. Grant S, Castaneda J, Samouei H, Tabei F, Parker D, Askarian B. Novel machine learning and image processing method for detecting seed germination. *ASABE Annu Int Meet. Am Soc Agric Biol Eng*. 2023;1 p.
15. Hecht-Nielsen R. *Neurocomputing*. Boston: Addison-Wesley Longman Publishing Co.; 1989. 433 p.
16. Kalaivani N, Dharshini SD, Divya S, Sundari KSJ. Identification and purity analysis of seed using convolutional neural network. *NeuroQuantology*. 2022;20(9):3296-3301.
17. Kiratiratanapruk K, Sinthupinyo W. Color and texture for corn seed classification by machine vision. In: *International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*. IEEE; 2011. p. 1-5.
18. Kohonen T. *Self-organizing feature maps*. Springer Berlin Heidelberg; 1989. p. 119-157.
19. Ławrynowicz A, Tresp V. *Introducing machine learning. Perspectives on Ontology Learning*. 2014;18:35-50.
20. Li H, Zou C, Duan X. Research progress of crop seed quality detection based on spectral imaging technology. *Proc SPIE*. 2023;12715:127151U.
21. Liu Y, Su J, Shen L, Lu N, Fang Y, Liu F, Su B. Development of a mobile application for identification of grapevine (*Vitis vinifera* L.) cultivars via deep learning. *Int J Agric Biol Eng*. 2021;14(5):172-179.
22. Maleki F, Ovens K, Najafian K, Forghani B, Reinhold C, Forghani R. Overview of machine learning part 1: Fundamentals and classic approaches. *Neuroimaging Clin*. 2020;30(4):e17-e32.
23. McDonald MB. Seed quality assessment. *Seed Sci Res*. 1998;8(2):265-276.
24. Mekala RL, Surapaneni RK, Medisetty V. Automated seed quality assessment and classification using watershed algorithm and ensemble learning. 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST). 2024; p. 1-6.
25. Mogilicharla S, Mummadi UK. Grain quality analysis from the image through the approaches of segmentation. *AIP Conf Proc*. 2024;3007:070001.
26. Mokry FB, Higa RH, Mudadu MA, Oliveira de Lima A, Meirelles SLC, Barbosa da Silva MVG, Correia de Almeida Regitano L. Genome-wide association study for backfat thickness in Canchim beef cattle using random forest approach. *BMC Genet*. 2013;14:1-11.
27. Namias R, Gallo C, Craviotto RM, Arango MR, Granitto PM. Automatic grading of green intensity in soybean seeds. *Proc XIII Argentine Symp Artif Intell*. 2012;96-104.
28. Nehoshan Y, Carmon E, Yaniv O, Ayal S, Rotem O. Robust seed germination prediction using deep learning and RGB image data. *Sci Rep*. 2021;11(1):22030.
29. Patrício DI, Rieder R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput Electron Agric*. 2018;153:69-81.
30. Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci*. 2016;21(2):110-124.
31. Škrubej U, Rozman Č, Stajanko D. Assessment of germination rate of tomato seeds using image processing and machine learning. *Eur J Hortic Sci*. 2015;80(2):68-75.
32. Srinivasiah R, Channegowda RH, Jankatti SK. Analysis and prediction of seed quality using machine learning. *Int J Electr Comput Eng*. 2023;13(5):5770-5781.
33. Taheri-Garavand A, Nasiri A, Fanourakis D, Fatahi S, Omid M, Nikoloudakis N. Automated in situ seed variety identification via deep learning: A case study in chickpea. *Plants*. 2021;10:1406.
34. Uzal LC, Grinblat GL, Namías R, Larese MG, Bianchi JS, Morandi EN, *et al*. Seed-per-pod estimation for plant breeding using deep learning. *Comput Electr Agric*. 2018;150:196-204.
35. Vapnik V. *The nature of statistical learning theory*. New York: Springer Science & Business Media; 2013. 314 p.
36. Vieira S, Pinaya WHL, Mechelli A. Main concepts in machine learning. In: *Machine Learning*. Academic Press; 2020. p. 21-44.
37. Zhao L, Haque SM, Wang R. Automated seed identification with computer vision: challenges and opportunities. *Seed Sci Technol*. 2022;50(2):75-102.
38. Silva CS, Kamantha UDB, Niruni RMC, Thennakoon TMTNB, Jayaprada NVT. Prediction of germination ability of tomato seeds based on phenotypic traits using machine learning. *MERCon*. 2024. 2024. <https://doi.org/10.1109/mercon63886.2024.10688645>
39. Genze N, Bharti R, Grieb M, Schultheiss SJ, Grimm DG. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods*. 2020. <https://doi.org/10.1186/S13007-020-00699-X>
40. Grant S, Castañeda JP, Samouei H, Tabei F, Parker D, Askarian B. Novel machine learning and image processing method for detecting seed germination. *American Society of Agricultural and Biological Engineers*. 2023. <https://doi.org/10.13031/aim.202300869>
41. Thakur PS, Kumar A, Tiwari B, Gedam B, Bhatia V, Rana S, *et al*. Machine learning-based biospeckle technique for identification of seed viability using

- spatio-temporal analysis. IEEE WRAP. 2022. <https://doi.org/10.1109/WRAP54064.2022.9758219>
42. Qiao J, Liao Y, Yin C, Yang X, Tú HM, Wang W, *et al.* Vigour testing for the rice seed with computer vision-based techniques. *Front Plant Sci.* 2023;14:1194701. <https://doi.org/10.3389/fpls.2023.1194701>
 43. Jin B, *et al.* Determination of viability and vigor of naturally-aged rice seeds using hyperspectral imaging with machine learning. *Infrared Phys Technol.* 2022;127:104097. <https://doi.org/10.1016/j.infrared.2022.104097>
 44. Colmer J, *et al.* SeedGerm: A cost-effective phenotyping platform for automated seed imaging and machine-learning-based phenotypic analysis of crop seed germination. *New Phytol.* 2020. <https://doi.org/10.1111/NPH.16736>
 45. Masteling R, Voorhoeve L, IJsselmuiden J, Dini-Andreote F, de Boer W, Raaijmakers JM. DiSCount: Computer vision for automated quantification of *Striga* seed germination. *Plant Methods.* 2020. <https://doi.org/10.1186/S13007-020-00602-8>
 46. Nehoshtan Y, Carmon E, Yaniv O, Ayal S, Rotem O. Robust seed germination prediction using deep learning and RGB image data. *Sci Rep.* 2021;11:17126. <https://doi.org/10.1038/S41598-021-01712-6>
 47. Wang Y, Song S. Detection of sweet corn seed viability based on hyperspectral imaging combined with firefly algorithm optimized deep learning. *Front Plant Sci.* 2024;15:1361309. <https://doi.org/10.3389/fpls.2024.1361309>
 48. Zhang T, Wei W, Zhao B, Wang R, Li M, Yang L, *et al.* A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. *Sensors.* 2018;18(3):813.
 49. Peng Q, Tu L, Wu Y, Yu Z, Tang G, Song W. Automatic monitoring system for seed germination test based on deep learning. *J Electr Comput Eng.* 2022;2022:4678316. <https://doi.org/10.1155/2022/4678316>
 50. Yigit E, Duysak H. Determination of flowing grain moisture contents by machine learning algorithms using free space measurement data. *IEEE Trans Instrum Meas.* 2022. <https://doi.org/10.1109/TIM.2022.3165740>
 51. Azmi N, Kamarudin LM, Zakaria A, Ndzi DL, Rahiman MHF, Zakaria SMMS, Mohamed L. RF-based moisture content determination in rice using machine learning techniques. *Sensors.* 2021;21(5):1875. <https://doi.org/10.3390/s21051875>
 52. Rahman A, *et al.* MoistNet: Machine vision-based deep learning models for wood chip moisture content measurement. *Expert Syst Appl.* 2024. <https://doi.org/10.1016/j.eswa.2024.125363>
 53. Yuan J, Xiao F, Zhao Y. Nondestructive detection of sunflower seed vigor and moisture content based on hyperspectral imaging and chemometrics. *Foods.* 2024;13(9):1320. <https://doi.org/10.3390/foods13091320>
 54. Cao W, Song H, Quan B. Research on grain moisture model based on improved SSASVR algorithm. *Appl Sci.* 2024;14(8):3171. <https://doi.org/10.3390/app14083171>
 55. Song H. Rapid nondestructive testing of wheat moisture content based on dielectric characteristics and machine learning. *IEEE Int Conf Emerg Technol.* 2024. <https://doi.org/10.1109/ICET61945.2024.10673046>
 56. Liu J, Qiu S, Wei Z. Real-time measurement of moisture content of paddy rice based on microstrip microwave sensor assisted by machine learning strategies. *Chemosensors.* 2022;10(10):376. <https://doi.org/10.3390/chemosensors10100376>
 57. Gapusan NA, *et al.* Design and development of a nondestructive feature-based multi-commodity moisture content sensor system using machine learning and computer vision. *Proc ICONAT.* 2024. <https://doi.org/10.1109/ICONAT61936.2024.10774723>
 58. Yang MD, *et al.* Assessment of grain harvest moisture content using machine learning on smartphone images for optimal harvest timing. *Sensors.* 2021;21(17):5875. <https://doi.org/10.3390/S21175875>
 59. Chen X, Li Z, Zhang H. AI-based hyperspectral prediction of corn seed moisture content. *Sci Rep.* 2024;14(3):60852.
 60. Hu X, Yang L, Zhang Z, Wang Y. Differentiation of alfalfa and sweet clover seeds via multispectral imaging. *Seed Sci Technol.* 2020;48(1):83-99. <https://doi.org/10.15258/SST.2020.48.1.11>
 61. Boateng R, Opoku-Ansah J, Amuah CLY, Adueming POW, Huzortey A, Taah K, *et al.* Laser-induced fluorescence spectroscopy combined with multivariate analysis for rice seeds and grains discrimination. *Appl Opt.* 2025;64:C148-C158. <https://doi.org/10.1364/AO.546627>
 62. Elfatimi E, Eryigit R, Elfatimi L. A novel convolutional neural network-based framework for complex multiclass *Brassica* seed classification. *arXiv preprint arXiv:2505.21558v1.* 2025.
 63. Thi-Thu-Hong P, Quoc-Trinh V, Huu-Du N. A novel method for identifying rice seed purity based on hybrid machine learning algorithms. *arXiv preprint arXiv:2406.07581v1.* 2024.
 64. Suganthi M, Sathiaselvan JGR. Prediction of seed purity and variety identification using image mining techniques. *ICTACT J Soft Comput.* 2022;12(4).
 65. Kalaivani N, Deva Dharshini S, Divya S, Jeya Sundari KS. Identification and purity analysis of seed using convolutional neural network. *Neuro Quantology.* 2022;20(9):3296-3301. <https://doi.org/10.14704/2022.20.9.NQ44380>
 66. Keling T, Weifeng W, Ying C, Han Z, Yanan X, Xuehui D, *et al.* ASeed: An automated image analysis software for high-throughput phenotyping and quality non-destructive testing of individual plant seeds. *Comput Electr Agric.* 2023;207:107740. <https://doi.org/10.1016/j.compag.2023.107740>
 67. Mahajan S, Mittal SK, Das A. Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (*Glycine max*). *J Food Sci Technol.* 2018;55(10):3949-3959.
 68. Ansari N, Ratri SS, Jahan A, Ashik-E-Rabbani M, Rahman A. Inspection of paddy seed varietal purity using machine vision and multivariate analysis. *J Agric Food Res.* 2021;3:100109. <https://doi.org/10.1016/j.jafr.2021.100109>