Seed Res. 52 (2): 75-84, December, 2024

Review Paper

Harnessing the Power of Machine Learning and Artificial Intelligence in Seed Science and Technology: A Review

AV BARAD, JB PATEL*, YASH J CHAUDHARY AND JYOTI SONDARVA

Junagadh Agricultural University, Junagadh, Gujarat-362001, India
*jbpatelvasai38@gmail.com

(Received June 2024; Revised July 2024; Accepted August 2024)

ABSTRACT: Integrating machine learning (ML) and artificial intelligence (AI) into seed science and technology represents a transformative paradigm in agricultural research. This study explores the potential and applications of ML and AI methodologies to enhance various facets of seed-related processes. From seed viability assessment to crop yield prediction, using advanced algorithms enable a more precise and efficient understanding of seed characteristics. The abstract delves into specific applications such as predictive modeling, image recognition, and data-driven decision-making in seed breeding. By harnessing the power of ML and AI, researchers and practitioners in seed science can revolutionize traditional approaches, fostering sustainable agriculture and ensuring food security in an ever-evolving global landscape.

Keywords: Machine learning (ML), Artificial intelligence (AI), image recognition, advanced algorithms

INTRODUCTION

Artificial Intelligence (AI) is a broad field encompassing simulation of human intelligence processes by computer systems. These processes include learning, reasoning, perception, self-correction, and creative problem-solving. The evolution of AI has been fueled by various technologies, with modern advancements heavily relying on machine learning algorithms and deep neural networks.

Machine learning enables computers to learn from experience without explicit programming. This paradigm shift allows AI systems to improve performance over time, adapting to new information and evolving circumstances. Deep neural networks, inspired by the structure of the human brain, enable the processing of vast amounts of data and the extraction of complex patterns, contributing to the advancement of AI capabilities.

The applications of AI are diverse and impactful. Natural language processing empowers machines to understand, interpret, and generate human language, facilitating communication between humans and computers. Image recognition enables AI systems to analyze and interpret visual information, leading to advancements in various fields, including agriculture.

NEED FOR AI IN SEED SCIENCE AND TECHNOLOGY

Artificial Intelligence (AI) can play a significant role in Seed Science and Technology, offering various benefits and addressing challenges in the industry. Here are some key areas where AI can be applied in seed science and technology:

1. **Genetic improvement:** AI algorithms can analyze vast genomic datasets to predict the performance of different plant varieties. This can accelerate the breeding process by identifying traits associated with desirable characteristics, leading to the development of high-yielding and resilient crops.

2. **Crop monitoring and precision agriculture:** AI-powered image analysis can be used to monitor crop health, identify diseases, and assess environmental conditions. This information can guide farmers in making data-driven decisions, optimizing resource allocation, and improving overall crop yield.

3. **Seed sorting and quality control:** AI-based image recognition can enhance seed sorting processes by automatically identifying and sorting seeds based on quality parameters such as size, shape, and color. This ensures the production of high-quality seeds for planting.

4. **Data-driven decision-making:** AI algorithms can analyze historical and real-time data to predict optimal planting times, irrigation schedules, and harvesting periods. This enables farmers to make informed decisions, maximizing crop yield and resource efficiency.

5. **Disease and pest management:** AI can detect early signs of diseases or pest infestations by analyzing sensor data, images, and other relevant information. This allows for prompt intervention and reduces the risk of crop losses.

6. **Climate Adaptation: Climate modeling:** AI can assist in modeling and simulating the impact of climate change on different crops. This information helps researchers and farmers develop strategies to adapt to changing environmental conditions and ensure sustainable seed production.

7. **Supply chain optimization:** AI algorithms can optimize the supply chain by predicting demand, managing inventory levels, and streamlining distribution processes. This helps ensure a steady supply of seeds to farmers and minimizes waste.

8. **Research and development:** AI-powered Natural Language Processing (NLP): tools can assist researchers in mining and extracting valuable information from scientific literature, facilitating the discovery of new insights, techniques, and innovations in seed science.

APPLICATIONS OF MACHINE LEARNING AND AI IN SEED SCIENCE AND TECHNOLOGY

1. Enhancing seed quality assessment

Seed quality assessment is a critical aspect of seed science, influencing crop establishment, vigor, and yield potential. Traditionally, assessing seed quality involved labor-intensive and subjective methods. However, this process has been revolutionized with the advent of machine learning algorithms. By analyzing vast datasets encompassing genetic information, environmental parameters and phenotypic traits, AI can accurately predict seed quality parameters such as germination rate, vigor and disease resistance [1, 2 and 3]. This data-driven approach in seed science plays a crucial role in modern plant breeding. By using advanced technologies like genomics, bioinformatics, and data analytics, scientists can identify genetic markers linked to desirable traits such as disease resistance, drought tolerance, higher yields, and improved nutritional quality.

2. Accelerating breeding programs

Breeding new crop varieties with desired traits is a time-consuming process traditionally reliant on trial and error. However, machine learning and artificial intelligence techniques have expedited this process significantly. By leveraging predictive modeling and genomic analysis, AI algorithms can sift through vast amounts of genetic data to identify candidate genes linked to desirable traits. This accelerates the selection of parental lines and progeny with desired characteristics, ultimately shortening breeding cycles and increasing breeding efficiency. Additionally, AI-driven breeding programs optimize hybridization and genome editing techniques, enhancing traits such as yield, stress tolerance, and nutritional content. By leveraging AI, seed scientists can develop resilient crop varieties adapted to changing climatic conditions and evolving market demands more rapidly than ever before.

3. Precision agriculture and seed placement

Precision agriculture relies on data-driven decision-making to optimize resource use and maximize crop productivity. AI plays a pivotal role by integrating data from various sources, including satellite imagery, drones and IoT sensors [4]. By analyzing spatial variability in soil properties, climate conditions and crop health indicators, AI generates precise seeding maps and recommends optimal seed placement strategies. Additionally, AI-powered machinery enables precise seed placement and planting density adjustments, ensuring uniform crop establishment and maximizing yield potential. This precision enhances resource efficiency, minimizes input costs, and mitigates environmental impact, promoting sustainable agricultural practices.

ADVANTAGES OF AI IN SEED SCIENCE AND TECHNOLOGY

Artificial Intelligence (AI) application in Seed Science and Technology offers numerous advantages, impacting various aspects of seed production, breeding, and agricultural practices.

1. **Accelerated breeding programs:** AI enables faster and more efficient identification of desirable traits in plant genomes. This accelerates the breeding process, reducing the time required to develop new and improved seed varieties.

2. **Precision in trait identification:** Computer vision and AI algorithms enhance the accuracy of identifying phenotypic traits in plants. This precision is crucial for selecting seeds with specific characteristics, contributing to the development of high-yielding and resilient crops.

3. **Optimized crop monitoring:** AI-based technologies, such as remote sensing and data analytics, provide real-time monitoring of crop health, disease detection, and environmental conditions. This allows for timely interventions and optimized crop management.

4. **Resource efficiency:** Precision agriculture guided by AI optimizes the use of resources such as water, fertilizers, and pesticides. This leads to more efficient resource allocation, reducing waste and environmental impact while maximizing crop yields.

5. **Seed sorting and quality control:** AI applications in seed sorting, utilizing computer vision and machine learning, ensure the production and distribution of high-quality seeds. This improves crop uniformity and overall productivity.

6. **Early detection of diseases and pests:** AI algorithms analyze data to detect early signs of diseases and pest infestations. Early intervention helps prevent the spread of diseases, minimizing crop losses and reducing the reliance on chemical treatments.

7. **Climate-resilient seed varieties:** AI assists in modeling and predicting the impact of climate change on crops. This information aids in developing seed varieties that are more resilient to changing climate conditions, ensuring stable yields in diverse environments.

8. **Data-driven decision making:** AI processes large volumes of data to provide actionable insights for farmers and researchers. This data-driven approach enhances decision-making, allowing for more informed and efficient agricultural practices.

9. **Supply chain optimization:** Predictive analytics powered by AI optimizes seed production, distribution, and inventory management. This results in a more streamlined supply chain, reducing costs and ensuring a consistent and reliable supply of seeds to farmers.

10. **Automation in laboratory processes:** AI-driven robotics and automation streamline laboratory processes, reducing manual labor and expediting research efforts. This accelerates the pace of research and development in seed science.

11. **Knowledge discovery and literature analysis:** AI, particularly Natural Language Processing (NLP), helps researchers analyze vast amounts of scientific literature, facilitating knowledge discovery and keeping scientists informed of the latest advancements in seed science.

12. **Improved farm management systems:** Integrated AI platforms offer comprehensive solutions for farm management, providing real-time insights into weather conditions, soil health, and crop performance. This enables farmers to make data-driven decisions for better overall farm management.

REVIEW OF WORK ON AI AND ML IN SEED SCIENCE AND TECHNOLOGY

Soybean

The viability of seeds is a critical factor in determining their quality, and ensuring high germination rates is essential for maximizing agricultural productivity. Traditional methods for assessing seed viability, such as germination tests, are often time-consuming and labor-intensive. As a result, there is a growing demand for rapid, non-destructive techniques that can accurately evaluate seed quality. Fourier Transform Near-Infrared (FT-NIR) spectroscopy has emerged as a powerful tool in this context. This technique measures the absorption of near-infrared light by the chemical bonds within the seed, providing detailed information about its molecular composition. Since viable and non-viable seeds have different chemical profiles, FT-NIR spectroscopy can be used to differentiate between them. Using both viable and artificially aged soybean seeds, Kusumaningrum *et al.* [5] applied partial least-squares discriminant analysis (PLS-DA) along with the variable importance in projection (VIP) method for optimal variable selection. Out of 1557 variables, 146 were identified as crucial by the VIP method. Results showcased the effectiveness of FT-NIR spectral analysis with PLS-DA, achieving prediction accuracies close to 100% for soybean viability. This highlights the potential of FT-NIR techniques, combined with chemo\metric analysis, for rapid and accurate soybean seed viability assessment, offering substantial benefits for agricultural productivity and seed quality control.

The study conducted by Baek *et al.* [6] demonstrates the power of near-infrared hyperspectral imaging (NIR-HSI) combined with advanced analytical methods in rapidly and non-destructively assessing soybean seed viability. Traditional methods of seed viability testing, while accurate, suffer from several limitations, including being time-consuming, labor-intensive, and destructive, which the NIR-HSI approach aims to overcome. Hyperspectral imaging integrates both spectroscopy and imaging, capturing spectral data for each pixel in an image. This allows for a detailed analysis of the chemical composition of the seeds. The NIR-HSI was employed to capture the spectral information of both viable and non-viable soybean seeds across a wide range of wavelengths. Partial Least Squares Discriminant Analysis (PLS-DA) was used as the main statistical method to classify seeds based on their spectral profiles. PLS-DA is effective for handling complex and high-dimensional data, making it suitable for analyzing the large datasets generated by hyperspectral imaging. The method initially provided reasonable classification accuracy, especially for viable seeds, but the pixel-based analysis did not offer high enough precision for non-viable seed detection. To improve efficiency, the study used Variable Importance in Projection (VIP) to identify the most informative wavelengths. VIP helps in selecting the wavelengths that contribute the most to the variance in the dataset, thus reducing the complexity of the model. This selection process resulted in a multispectral model that used only a few optimal wavebands, significantly simplifying the analysis without compromising accuracy. In response to the limitations of pixel-based classification, a kernel image threshold method was introduced. This method relies on creating image-based features, such as seed kernels, which better capture the overall characteristics of seeds rather than individual pixels. This kernel-based approach, combined with the optimal detection-rate strategy, achieved over 95% accuracy in distinguishing viable from non-viable seeds when using only seven wavebands identified by VIP. This study highlights the potential of using hyperspectral imaging for fast and non-invasive seed viability testing. Unlike traditional methods that often require destroying the seed (e.g., tetrazolium tests or germination tests), NIR-HSI preserves the seed for future use. The ability to accurately detect seed viability using only seven optimal wavebands is significant. This reduction in data dimensionality allows for the development of more cost-effective and efficient multispectral imaging systems that could be more readily adopted in practical seed testing environments. The approach outlined in this study could be extended beyond soybean seeds to other crops, improving seed quality testing in agricultural production. Additionally, it could be used for real-time sorting in seed processing facilities, contributing to better seed selection and enhanced crop yields. The approach outlined in this study could be extended beyond soybean seeds to other crops, improving seed quality testing in agricultural production. Additionally, it could be used for real-time sorting in seed processing facilities, contributing to better seed selection and enhanced crop yields. The NIR-HSI technology, combined with PLS-DA and kernel-based classification, presents a promising avenue for automating seed viability testing on a larger scale.

de Medeiros *et al.* [1] proposed an interactive and traditional machine learning approach to classify soybean seeds and seedlings based on appearance and physiological potential. Categorizing 700 seeds from seven lots into seven groups, they considered both physical and physiological characteristics. Seed appearance and physiological quality were classified into seven and three classes, respectively: vigorous seedlings, weak seedlings and non-germinated seeds. Utilizing descriptors from Ilastik, the researchers developed classification models using Linear Discriminant Analysis (LDA), Random Forest (RF) and Support Vector Machine (SVM) algorithms. The dataset was divided into 70% training and 30% validation sets to ensure the robustness of the models. The cost-effective models, utilizing free-access software, exhibited an impressive 0.94 overall accuracy, effectively categorizing soybean seeds and assessing seedling vigor. The observed correlation between seed appearance and physiological performance added significant value to their findings.

Da Silva Andre *et al.* [7] employed machine learning techniques, including Artificial Neural Networks, REPTree, M5P, Random Forest and Linear Regression, to predict soybean seed quality based on temperature (T), packaging (P) and storage time (ST). The research incorporated a randomized three-factor design with variations in temperature, packaging types (raffia bag and polyethylene-coated raffia bags) and evaluation times, conducting three repetitions every three months. The models, encompassing Artificial Neural Network, decision trees (REPTree and M5P), Random Forest, and Linear Regression, utilized inputs such as temperature (T), packaging (P) and storage time (ST), alongside combinations like T+P, T+ST, and T+P+ST. Results highlighted the superior accuracy of temperature and storage time predictions, particularly with REPTree and Random Forest models compared to linear regression. Notably, combinations like T+P+ST, T+ST, P+ST and ST accurately predicted apparent specific mass, with

T+P+ST and T+ST demonstrating superior accuracy in germination results. The study underscores the efficacy of computational intelligence for predicting soybean seed quality, providing valuable insights for decision-making in seed storage processes.

Forage grass (*Urochloa brizantha*)

De Medeiros *et al.* [2] utilized FT-NIR spectroscopy and X-ray imaging coupled with machine learning to classify seed quality for *Urochloa brizantha*. For data acquisition, FT-NIR spectroscopy data and X-ray images were collected individually from 200 seeds. Various machine learning algorithms, including Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA) and Random Forest (RF), were employed. Notably, individual models achieved an 82 per cent accuracy for germination (FT-NIR) and 90 per cent accuracy (X-ray) while for seed vigor; the accuracy was 61 per cent (FT-NIR) and 68 per cent (X-ray). Upon combining the datasets, the integrated model demonstrated 85 per cent accuracy for germination and 62 per cent for seed vigor. This comprehensive approach effectively assessed seed germination, with X-ray data and LDA exhibiting promise for the quality classification of *U. brizantha* seeds.

Wheat

Fan *et al.* [3] utilized NIR spectroscopy and machine learning for wheat seed vigor detection, analyzing 1152 samples with artificial ageing. They created four machine learning models (SVM, ELM, RF, AdaBoost) and applied PCA and SPA for spectral dimensional reduction. Of 1152 samples, 2/3 were used for calibration and 1/3 for prediction. Data preprocessing involved SG second derivative-method and SNV. The study focused on classifying wheat kernels into three categories based on vigor. The eight resulting models achieved high accuracies, exceeding 84.0 per cent, with PCA-ELM and SPA-RF reaching the highest at 88.9 per cent and 88.5 per cent, respectively. This research advances the development of a rapid, non-destructive NIR-based sorting system for assessing wheat kernel vigor, which holds implications for plant breeders, wheat quality inspectors, and processors.

Khatri *et al.* [8] addressed the critical need for recognizing and authenticating wheat varieties in the grain supply chain. Manual methods for seed inspection are traditionally employed, but the researchers explored machine learning and computer vision for automatic categorization. The dataset comprises 210 occurrences of wheat kernels from three varieties: Kama, Rosa and Canadian. Machine learning algorithms, including K-nearest neighbor, classification and regression tree and Gaussian Naïve Bayes, were applied based on seven physical features. Comparative analysis reveals accuracies of 92%, 94% and 92% for KNN, decision and Naïve Bayes, respectively. The ensemble classifier, combining these algorithms through hard voting, achieves the highest accuracy of 95%. This approach offers a fast and high-throughput solution for wheat variety classification in seed inspection.

Alfalfa

Yang *et al.* [9] explored a novel method for the rapid and accurate discrimination of alfalfa cultivars using multispectral imaging combined with object-wise multivariate image analysis. The conventional approaches to identifying alfalfa cultivars are typically labor-intensive and time-consuming. In this study, the authors aimed to streamline the process by focusing on the seed level, employing three multivariate analysis techniques: principal component analysis (PCA), linear discrimination analysis (LDA), and support vector machines (SVM). Their approach integrated both morphological and spectral traits of alfalfa seeds, demonstrating that while seed morphological features alone were insufficient for cultivar classification, combining them with spectral data provided highly accurate results. Specifically, the study achieved classification accuracies of 91.53% using LDA and 93.47% with SVM for testing sets. This finding suggests that multispectral imaging paired with multivariate analysis offers a simple, robust, and nondestructive method to effectively distinguish between alfalfa seed cultivars. This method holds promise for producers, consumers, and market regulators seeking more efficient and accurate seed identification.

Wang *et al.* [10] developed a rapid, nondestructive method for detecting seed aging and predicting seed viability in alfalfa, addressing the limitations of traditional, destructive techniques. The study employed multispectral imaging to capture both morphological and spectral characteristics of alfalfa seeds stored for various durations. Using five multivariate analysis methods—principal component analysis (PCA), linear discrimination analysis (LDA), support vector machines (SVM), random forest (RF), and normalized canonical discriminant analysis (nCDA)—they assessed the ability to distinguish between aged and viable seeds. Their findings showed significant differences in light reflectance between 450 and 690 nm for non-aged and aged seeds. Among the models, LDA demonstrated the highest accuracy (99.8–100.0%) in differentiating aged seeds from non-aged ones, outperforming SVM (87.4–99.3%) and RF (84.6–99.3%). Additionally, LDA accurately distinguished dead seeds from aged seeds with an accuracy of 97.6%, compared to lower accuracies for RF (69.7%) and SVM (72.0%). For predicting the germination of aged seeds, nCDA achieved accuracies between 75.0% and 100.0%. Overall, this study presents a nondestructive, rapid, and high-throughput method for screening aged alfalfa seeds with varying levels of viability, offering significant potential for seed production and research.

Zhang *et al.* [11] utilized multispectral imaging and various multivariate analysis methods, including LDA, SVM, RF, and nCDA, to assess seed vigor in alfalfa seeds with different maturity levels and harvest years. Alfalfa seeds from varied harvest years (2004, 2008 and 2019) and maturity stages (green ripe, yellow ripening and full ripe) were harvested in 2021. Morphological and spectral information from 19 wavelengths (ranging from 365 nm to 970 nm) was collected through multispectral imaging. Five multivariate analysis methods (principal component analysis (PCA), linear discriminant analysis (LDA), support vector machine (SVM), random forest (RF) and normalized canonical discriminant analysis (nCDA)) were utilized to analyze the morphological and spectral traits of alfalfa seeds. The LDA model showcased exceptional effectiveness, achieving 92.9% accuracy for maturity and 97.8% for harvest years, demonstrating high sensitivity, specificity and precision. Through the integration of multispectral imaging and these analyses, the study successfully predicted seed viability and germination percentages, presenting a rapid and non-destructive approach to evaluating seed quality.

Maize

NIR hyperspectral imaging was evaluated by Williams and Kucheryavskiy [12] to classify maize kernels into three hardness categories: hard, medium and soft. Two approaches, pixel-wise and object-wise, were tested for their effectiveness in grouping kernels by hardness. In pixel-wise classification, each pixel within the maize kernels was assigned a hardness class. However, this method led to significant misclassification. To improve accuracy, a predefined threshold was introduced to classify entire kernels based on the number of correctly predicted pixels. This approach yielded improved results, with a sensitivity of 0.75 and a specificity of 0.97. Object-wise classification classified whole kernels using two feature extraction techniques *viz*., Score histograms and Mean spectra. Score histograms technique was particularly effective for classifying hard kernels, achieving a sensitivity of 0.93 and a specificity of 0.97. Mean Spectra method performed better for medium kernels, with a sensitivity of 0.95 and a specificity of 0.93. Both the score histogram and mean spectra methods were recommended for classifying maize kernels on a production scale due to their high accuracy in categorizing kernel hardness.

Jatropha

Bianchini *et al.* [13] introduced a novel method for automatic seed quality characterization by combining multispectral and X-ray imaging technologies, reducing human interference and enhancing objectivity. Utilizing X-ray images to examine internal tissues, they applied the normalized canonical discriminant analyses (nCDA) algorithm on various seed lots of *Jatropha curcas*, a globally significant oilseed plant. The developed classification models, employing linear discriminant analysis (LDA) on reflectance data and X-ray classes, demonstrated high accuracy (>0.96). Reflectance at 940 nm and X-ray data effectively predicted quality traits such as normal seedlings, abnormal seedlings and dead seeds. The strong correlation between multispectral and X-ray imaging proved valuable in assessing seed physiological performance. This approach offers a rapid, efficient, sustainable and non-destructive means of characterizing seed quality, overcoming the inherent subjectivity in conventional analyses. It presents a promising avenue for more objective seed quality assessments in the future.

Maize

Agelet *et al.* [14] evaluated the use of near-infrared spectroscopy (NIRS), a fast and non-destructive analytical method, to discriminate between maize kernels with heat and frost damage, as well as to distinguish viable and non-viable seeds. The current U.S. corn grading system relies on time-consuming and inaccurate visual inspections to assess kernel damage, making NIRS a potentially valuable alternative. Four classification algorithms (partial least squares discriminant analysis (PLS-DA), soft independent modeling of class analogy (SIMCA), k-nearest neighbors (K-NN) and least-squares support vector machines (LS-SVM)) were utilized to test the feasibility of NIRS. The application of NIRS could be highly valuable for seed breeders and germplasm-preservation managers, as it offers a non-destructive alternative to traditional viability tests, which require seeds to be germinated. Heat-damaged maize kernels were best discriminated using PLS-DA, achieving 99% accuracy. Frost-damaged corn kernels could not be reliably discriminated by NIRS. Distinguishing of non-viable seeds from viable also proved unsuccessful. Despite the promising results for heat-damaged kernels, the inability to detect frost damage and seed viability contradicts some prior studies, suggesting that further analysis is required to determine the threshold of seed damage necessary for NIRS detection. Overall, NIRS shows potential in classifying seed damage based on chemical and structural changes, although it may not directly correlate with seed germination potential.

Ambrose *et al.* [15] investigated the potential of hyperspectral imaging (HSI) to differentiate viable and nonviable maize seeds. Utilizing optimized agronomic practices and technological interventions, corn's quantity and quality were emphasized. Seed germination and vigor were identified as crucial factors for high yield. Seed viability, prone to loss during storage or processing, necessitates testing for the prevention of losses. The study applied HSI to corn samples, including a heat-treated group and an untreated control group. Hyperspectral images captured between 400 and 2500 nm were analyzed using Partial Least Squares Discriminant Analysis (PLS-DA). The model achieved a peak classification accuracy of 97.6% (calibration) and 95.6% (prediction) in the short-wave infrared (SWIR) region. PLS-DA and binary images visually conveyed treated and untreated corn seeds. In conclusion, the study affirms the accuracy of HSI for non-destructive classification of viable and nonviable seeds.

The Fourier transform near-infrared (FT-NIR) and Raman spectroscopy techniques were used by Ambrose *et al.* [16] for evaluating seed viability to investigate their comparative advantages with regard to the maize viability test and classification. In the study, white, yellow, and purple corn seeds were analyzed using Fourier Transform Near-Infrared (FT-NIR) and Raman spectroscopy to assess seed viability. A total of 300 corn seeds from each category were divided into two groups: one group was heat-treated (via microwaving), and the other served as the control. Spectral data from both treated and untreated seeds were collected using an FT-NIR spectrometer (wavelength range of 1000–2500 nm) and a Raman spectrometer (wavelength range of 170–3200 cm{ ^{1}). The samples were split into training (70%) and testing (30%) sets for calibration and validation purposes. Principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) were applied to assess the spectral data from FT-NIR and Raman spectroscopy. FT-NIR spectroscopy demonstrated superior performance, achieving 100% accuracy in correctly classifying viable and non-viable seeds across all corn categories, with a predictive ability of more than 95%. Raman spectroscopy also performed well with PLS-DA, though some seed samples overlapped when using PCA, indicating lower classification accuracy compared to FT-NIR. Analysis of Variance showed that the difference between treated and untreated seeds was not statistically significant ($P < 0.05$), suggesting that the heat treatment did not produce significant spectral variations detectable by either method. Overall, the study highlighted FT-NIR spectroscopy as the more effective technique for evaluating corn seed viability compared to Raman spectroscopy.

Muskmelon

Kandpal *et al.* [17] employed a near-infrared (NIR) hyperspectral imaging (HSI) system to predict viability and vigor (germination periods) in muskmelon seeds. Hyperspectral images, covering the spectral range of 948-2494 nm, were acquired and subjected to a germination test. Partial least-squares discriminant analysis (PLS-DA) was utilized for classification in the study, focusing on effective wavelengths selected through three key techniques, Variable Importance in Projection (VIP), Selectivity Ratio (SR) and significance Multivariate Correlation (sMC). These methods selected 23, 18 and 19 optimal variables, respectively, from a set of 208 variables. The PLSDA-SR method achieved the highest classification accuracy (94.6%) for the validation set, indicating its effectiveness in determining the viability and vigor of muskmelon seeds. The selected wavelengths primarily represented chemical components related to germination ability.

Castor

Olesen *et al.* [18] applied multispectral imaging to assess seed quality in castor seeds. They categorized 120 seeds into three visually distinct classes: yellow, grey, and black. Images were captured at 19 different wavelengths ranging from 375 to 970 nm, and the mean intensity for each seed was extracted from these images. A significant difference between the three color classes was observed, with the best separation occurring in the near-infrared wavelengths. The study employed normalized canonical discriminant analysis (nCDA), particularly focusing on a feature termed “Region MSI mean,” which was crucial for differentiating viable from dead seeds. This model achieved 92% accuracy in distinguishing viable seeds from dead ones. A validation set was also used, consisting of seeds divided into two groups based on their germination ability: 241 seeds were predicted to be viable, while 59 were predicted to be non-viable. The model accurately classified 96% of the seeds in the validation set. These results demonstrate how multispectral imaging can be an effective tool for predicting seed viability in castor seeds based on seed coat color, providing a non-destructive and efficient method for seed quality testing.

Tomato

Shrestha *et al.* [19] explored the potential of near-infrared (NIR) spectroscopy for non-destructive classification of viable and non-viable tomato seeds from two cultivars using chemometrics. Principal component analysis (PCA) revealed clustering of viable and non-viable seeds in each cultivar and the pooled samples. However, PCA did not exhibit a pattern of separation among early, normal and late germinated seeds. Through partial least squares-discriminant analysis (PLS-DA) and interval PLS-DA (iPLS-DA), specific NIR spectral regions (1160–1170, 1383–1397, 1647–1666, 1860–1884 and 1915–1940 nm) were identified as crucial for classification. The iPLS-DA model demonstrated higher sensitivity (0.94) and specificity (0.94) compared to the PLS-DA model on original spectra. The iPLS-DA model achieved a lower classification error rate (6.29%) than the PLS-DA model (13.10%). The study suggested positive relationships between NIR regions related to protein-bound water, protein, and carbohydrates with seed viability. Overall, NIR spectroscopy shows promise for non-destructive discrimination of viable and non-viable tomato seeds using spectral information.

Cowpea

ElMasry *et al.* [20] explored the efficacy of computer vision and multispectral imaging, supported by multivariate analysis, for high-throughput classification of cowpea (*Vigna unguiculata*) seeds. Utilizing an automated germination system, continuous monitoring allows the identification of seed categories based on ageing, viability, seedling condition, and germination speed. Linear discriminant analysis (LDA) models demonstrate significant accuracy, ranging from 62% to 97%. The approach introduces a non-destructive, efficient, and high-throughput method with potential applications in the seed industry.

Chickpea

The ability of Artificial Neural Networks (ANN) in classification of chickpea varieties was considered based on morphological properties of seeds by Ghamari [21]. Experimentally, the seven morphological features of 400 seeds (including four varieties; Kaka, Piroz, Ilc and Jam) were obtained. Using a combination of input variables, a database consisting of 400 patterns was created for the development of Artificial Neural Network (ANN) models. Two types of ANN approaches viz., back propagation algorithm (BP) and self-organizing map (SOM) were compared for classification: supervised and unsupervised learning. The results of this study showed that unsupervised artificial neural network has a better performance (with 79% accuracy and R2 = 0.8455) in classification of chickpea varieties rather than supervised artificial neural networks (with 73% accuracy and R2 = 0.8236).

Watermelon

Lohumi *et al.* [22] explored the use of Fourier transform near-infrared (FT-NIR) spectroscopy to differentiate between viable and nonviable watermelon seeds. The FT-NIR reflectance spectra of seeds were recorded within the wavelength range of 4,000 – 10,000 cm{ ^{1} (1,000 – 2,500 nm). To classify the seeds, the researchers developed a partial least squares discriminant analysis (PLS-DA) model. The PLS-DA model demonstrated 100% accuracy in classifying both viable and nonviable seeds for both the calibration and validation sets. The beta coefficient from the PLS-DA model indicated that differences in chemical components, particularly lipids and proteins within the seed membrane, played a critical role in distinguishing seed viability. This study suggests that FT-NIR spectroscopy can serve as an effective, non-destructive tool for seed viability assessment. It opens up possibilities for developing online sorting techniques that could enhance seed production processes by rapidly and accurately identifying viable seeds based on their chemical characteristics.

CHALLENGES

1. Data accessibility and quality

Ensuring the availability of high-quality, diverse datasets is imperative for training accurate machine learning (ML) and artificial intelligence (AI) models in seed science and technology. Challenges may emerge concerning data collection, standardization and privacy. Collaborative efforts among stakeholders are necessary to address these challenges. By establishing data-sharing protocols, standardizing data formats and implementing privacy safeguards, the seed science community can overcome data-related obstacles and unlock the full potential of AI and ML.

2. Regulatory considerations

Ethical and regulatory frameworks must be developed to ensure the responsible adoption of AI and ML technologies in seed science. Issues such as data privacy, intellectual property rights and algorithmic bias require careful consideration and transparent governance. Policymakers, researchers and industry leaders must collaborate to establish guidelines that protect data privacy, promote fair use of intellectual property and mitigate the risks of algorithmic bias. By adhering to ethical standards and regulatory guidelines, the seed science community can foster trust and confidence in AI and ML technologies.

3. Interdisciplinary collaboration

Successful integration of AI and ML into seed science and technology hinges on interdisciplinary collaboration. Researchers, breeders, technologists, and policymakers must work together to leverage diverse expertise and perspectives. Interdisciplinary teams can address complex challenges more effectively, drawing on insights from fields such as genetics, agronomy, computer science, and ethics. By fostering a culture of collaboration and knowledge sharing, the seed science community can harness the synergies of AI and ML to drive innovation and address global food security challenges.

FUTURE PROSPECTS

Continued research and innovation in AI and ML promise further advancements in seed science and technology. As technology evolves, new opportunities will emerge to enhance seed selection, breeding programs and crop management strategies. Embracing sustainable practices and addressing ethical considerations will be critical in shaping the future of agriculture. By prioritizing sustainability, equity and ethical governance, the seed science community can harness the transformative power of AI and ML to create a more resilient and food-secure future for generations to arrive.

CONCLUSION

To sum it up, artificial intelligence holds considerable potential in seed science and technology. Multispectral imaging analysis was successfully performed to predict aged seeds and their germination. The findings suggest that combining spectral data with morphological characteristics through advanced multivariate analysis techniques like LDA and nCDA can significantly enhance the ability to classify seeds based on age and viability. This approach not only improves the accuracy of seed assessments but also provides valuable insights for seed producers and researchers in optimizing seed quality evaluation methods. In brief, review clearly shows that multispectral imaging, together with multivariate analysis, is a promising technique in predicting and nondestructive testing of aged and viable seeds.

REFERENCES

1. DE MEDEIROS AD, NP CAPOBIANGO, JM DA SILVA, LJ DA SILVA, CB DA SILVA AND DCF DOS SANTOS DIAS (2020a). Interactive machine learning for soybean seed and seedling quality classification. *Scientific reports*, **10**(1): 11267 (https://doi.org/10.1038% 2Fs41598- 020-68273-y).
2. DE MEDEIROS AD, LJ DA SILVA, JPO RIBEIRO, KC FERREIRA, JTF ROSAS, AA SANTOS AND CB DA SILVA (2020b). Machine learning for seed quality classification: An advanced approach using merger data from FT-NIR spectroscopy and X-ray imaging. *Sensors*, **20**(15): 4319 (https://doi.org/10.3390/s20154319).
3. FAN Y, S MA AND T WU (2020). Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Physics and Technology*, **105**: 103213 (http://dx.doi.org/10.1016/j.infrared.2020.103213).
4. AKHTER R AND SA SOFI (2022). Precision agriculture using IoT data analytics and machine learning. *Journal of King Saud University-Computer and Information Sciences*, 34(8): 5602-5618.
5. KUSUMANINGRUM D, H LEE, S LOHUMI, C MO, MS KIM AND BK CHO (2018). Non destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT NIR spectroscopy. *Journal of the Science of Food and Agriculture*, **98**(5): 1734-1742.
6. BAEK I, D KUSUMANINGRUM, LM KANDPAL, S LOHUMI, C MO, MS KIM AND BK CHO (2019). Rapid measurement of soybean seed viability using Kernel-based multispectral image analysis. *Sensors*, **19**(2): 271 (http://dx.doi.org/10.3390/s19020271).

7. DA SILVA ANDRE G, PC CORADI, LPR TEODORO AND PE TEODORO (2022). Predicting the quality of soybean seeds stored in different environments and packaging using machine learning. *Scientific Reports*, **12**(1): 8793 (https://doi.org/10.1038/s41598-022-12863-5).

8. KHATRI A, S AGRAWAL AND JM CHATTERJEE (2022). Wheat seed classification: utilizing ensemble machine learning approach. *Scientific Programming*, 2022: Article ID 2626868, 9 pages (https://doi.org/10.1155/2022/ 2626868).

9. YANG L, Z ZHANG AND X HU. (2020). Cultivar discrimination of single alfalfa (*Medicago sativa* L.) seed via multispectral imaging combined with multivariate analysis. *Sensors*, **20**(22): 6575 (http://dx.doi.org/10.3390/s20226575).

10. WANG X, H ZHANG, R SONG, X HE, P MAO AND S JIA (2021). Non-destructive identification of naturally aged alfalfa seeds via multispectral imaging analysis. *Sensors*, **21**(17): 5804 (https://doi.org/10.3390/s21175804).

11. ZHANG S, H ZENG, W JI, K YI, S YANG, P MAO, Z WANG, H YU AND M LI (2022). Non-destructive testing of alfalfa seed vigor based on multispectral imaging technology. *Sensors*, **22**(7): 2760 (https://doi.org/10.3390/s22072760)

12. WILLIAMS PJ AND S KUCHERYAVSKIY (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**:131-138.

13. BIANCHINI VDJM, GM MASCARIN, LCAS SILVA, V ARTHUR, JM CARSTENSEN, B BOELT AND C BARBOZA DA SILVA (2021). Multispectral and X-ray images for characterization of *Jatropha curcas* L. seed quality. *Plant Methods*, **17**(1): 9 (doi: 10.1186/s13007-021-00709-6. PMID: 33499879; PMCID: PMC7836195.).

14. AGELET, LE, DD ELLIS, S DUVICK, AS GOGGI, CR HURBURGH AND CA GARDNER (2012). Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels. *Journal of Cereal Science*, **55**(2): 160-165.

15. AMBROSE A, LM KANDPAL, MS KIM, WH LEE AND BK CHO (2016a). High speed measurement of corn seed viability using hyperspectral imaging. *Infrared Physics and Technology*, **75**: 173-179.

16. AMBROSE A, S LOHUMI, WH LEE AND BK CHO (2016b). Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. *Sensors and Actuators B: Chemical*, **224**: 500-506.

17. KANDPAL LM, S LOHUMI, MS KIM, JS KANG AND BK CHO (2016). Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. *Sensors and Actuators B: Chemical*, **229**: 534-544.

18. OLESEN MH, P NIKNESHAN, S SHRESTHA, A TADAYYON, LC DELEURAN, B BOELT AND R GISLUM (2015). Viability prediction of *Ricinus cummunis* L. seeds using multispectral imaging. *Sensors*, **15**(2): 4592-4604.

19. SHRESTHA S, LC DELEURAN AND R GISLUM (2017). Separation of viable and non-viable tomato (*Solanum lycopersicum* L.) seeds using single seed near-infrared spectroscopy. *Computers and Electronics in Agriculture*, **142**: 348-355.

20. ELMASRY G, N MANDOUR, MH WAGNER, D DEMILLY, J VERDIER, E BELIN AND D ROUSSEAU (2019). Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant methods*, **15**: 24 (https://doi.org/10.1186/s13007-019-0411-2).

21. GHAMARI S (2012). Classification of chickpea seeds using supervised and unsupervised artificial neural networks. *African Journal of Agricultural Research*, **7**(21): 3193-3201.

22. LOHUMI S, C MO, JS KANG, SJ HONG AND BK CHO (2013). Nondestructive evaluation for the viability of watermelon (*citrullus lanatus*) seeds using fourier transform near infrared spectroscopy. *Journal of Biosystems Engineering*, **38**(4): 312-317.