Article

Not peer-reviewed version

Identifying Optimal Wavelengths from Near Infrared Spectroscopy Using Meta-heuristic Algorithms to Assess Peanut Seed Viability

Mohammad Rajabi-Sarkhani , Yousef Abbaspour-Gilandeh * , Abdolmajid Moinfar , Mohammad Tahmasebi , Miriam Martínez-Arroyo * , Mario Hernández-Hernández , José Luis Hernández-Hernández

Posted Date: 13 November 2023

doi: 10.20944/preprints202311.0733.v1

Keywords: Seed viability; spectrometry; variable selection method; machine learning; non-destructive diagnosis; meta-heuristic algorithm

 doi:10.20944/preprints202311.0733.v1

Article

Identifying Optimal Wavelengths from Near Infrared Spectroscopy Using Meta-Heuristic Algorithms to Assess Peanut Seed Viability

Mohammad Rajabi-Sarkhani [1], Yousef Abbaspour-Gilandeh [1,*], Abdolmajid Moinfar [1], Mohammad Tahmasebi [1], Miriam Martínez-Arroyo [2,*], Mario Hernández-Hernández [3] and José Luis Hernández-Hernández [4]

[1] Department of Biosystems Engineering, College of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran; mohammadrajabi587@gmail.com (M.R.-S.), majid.mf93@gmail.com (A.M.); m.tahmasebi@uma.ac.ir (M.T.)
[2] National Technology of Mexico/Acapulco Institute of Technology, Acapulco 39905, Guerrero, Mexico
[3] Faculty of Engineering, Autonomous University of Guerrero, Chilpancingo 39070, Mexico; mhernandezh@uagro.mx
[4] National Technological of Mexico/Chilpancingo Institute of Technology, Chilpancingo 39070, Guerrero, Mexico; joseluis.hernandez@itchilpancingo.edu.mx
* Correspondence: abbaspour@uma.ac.ir (Y.A.-G.); miriam.ma@acapulco.tecnm.mx (M.M.-A.)

Abstract: Peanuts, owing to their composition of complex carbohydrates, plant protein, unsaturated fatty acids, and essential minerals (magnesium, iron, zinc, and potassium), hold significant potential as a vital component of the human diet. Additionally, their low water requirements and nitrogen fixation capacity make them an appropriate choice for cultivation in adverse environmental conditions. The germination ability of seeds profoundly impacts the final yield of the crop, assessing seed viability of extreme importance. Conventional methods for assessing seed viability and germination are both time-consuming and costly. To address these challenges, this study investigated Visible-Near Infrared Spectroscopy (Vis/NIR) in the wavelength range of 500–1030 nm as a non-destructive and rapid method to determine the viability of two varieties of peanut seeds: North Carolina-2 (NC-2) and Spanish flower (Florispan). The study subjected the seeds to three levels of artificial aging through heat treatment, involving incubation in a controlled environment at a relative humidity of 85% and a temperature of 50 °C over 24-h intervals. The absorbance spectra noise was significantly mitigated and corrected to a large extent by combining the Savitzky-Golay (SG) and Multiplicative Scatter Correction (MSC) methods. To identify the optimal wavelengths for seed viability assessment, a range of meta-heuristic algorithms were employed, including world competitive contest (WCC), league championship algorithm (LCA), Genetics (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), imperialist competitive algorithm (ICA), learning automata (LA), heat transfer optimization (HTS), forest optimization (FOA), discrete symbiotic organisms search (DSOS), and cuckoo optimization (CUK). These algorithms offer powerful optimization capabilities for effectively extracting relevant wavelength information from spectral data. Results revealed that all the algorithms demonstrated remarkable accuracy in predicting the allometric coefficient of seeds, achieving correlation coefficients exceeding 0.985 and errors below 0.0036, respectively. In terms of execution time, the ICA (2.3635 s) and LCA (44.9389 s) algorithms exhibited the most and least efficient performance, respectively. Conversely, the FOA and the LCA algorithms excelled in identifying the least number of optimal wavelengths (10 wavelengths). Subsequently, the seeds were classified based on the wavelengths selected by the FOA (10 wavelengths) and (DSOS (16 wavelengths) methods, in conjunction with logistic regression (LR), decision tree (DT), Multilayer Perceptron (MP), support vector machine (SVM), k-nearest neighbor (K-NN), and NaiveBayes (NB) classifiers. The DSOS-DT and FOA-MP methods demonstrated the highest accuracy, yielding values of 0.993 and 0.983, respectively. Conversely, the DSOS-LR and DSOS-KNN methods obtained the lowest accuracy, with values of 0.958 and 0.961, respectively. Overall, our findings demonstrated that Vis/NIR spectroscopy, coupled with variable selection algorithms and learning methods, presents a suitable and non-destructive approach for detecting seed viability.

Keywords: Seed viability; spectrometry; variable selection method; machine learning; non-destructive diagnosis; meta-heuristic algorithm

 doi:10.20944/preprints202311.0733.v1

1. Introduction

Peanuts hold significant economic importance globally, serving as a stable and cost-effective source of complex carbohydrates, plant protein, unsaturated fatty acids, and essential minerals such as magnesium, iron, zinc, and potassium [1]. This plant's ability to thrive with low water requirements makes peanuts a favorable choice for cultivation in arid and semi-arid regions [2]. Moreover, in adverse environmental conditions, it exhibits resistance and is capable of performing adequately in poor soil conditions [3]. Additionally, peanuts' excellent nitrogen fixation properties make them a suitable rotation option with cereal crops in successive planting seasons, contributing to soil fertility improvement [4]. Since peanut cultivation is conducted via seeds, evaluating seed germination vigor and viability is of paramount importance.

Seeds are considered the fundamental elements in agriculture and forestry, as they are directly or indirectly involved in establishing fields for various crops, vegetables, fruits, fodder, and economic forest products [5]. The quality of seeds profoundly impacts crop growth uniformity, yield, and overall crop quality. Furthermore, the safety and quality of seeds and their products directly affect human health [6]. Among the vital parameters associated with seed quality, seed vigor represents a key criterion for assessing seed quality, as it reflects the potential for seed germination, germination in the field, resistance to biotic and abiotic stresses, and the ability to withstand different storage conditions compared to standard germination tests [7]. Furthermore, it is well-established that seeds with desirable viability capabilities, achieved through significant yield performance for farmers and reduced crop diversity, will be profitable for seed industries [8]. A vigorous seed possesses the potential to thrive in environmental conditions that may not be optimal for its species. Such seeds exhibit high and uniform germination rates, quick germination, and produce robust seedlings, ultimately leading to higher field yields [9]. The study of the relationship between the growth rates of different parts of an organism or the creature as a whole is known as allometry. In identical environmental conditions, the growth and functioning of both the root and aerial systems are closely interrelated, and this relationship can be quantified through allometric relationships. Specifically, the allometric coefficient, which is calculated based on the length of the aerial parts and roots, represents the ratio of shoot length to root length [10]. Several research studies have indicated that the allometric coefficient is influenced by seed vigor and can be served as an indicator for diagnosing seed quality [11,12]. (Ebrahimi & Miri, 2016; Mohajeri et al., 2016). Traditionally, various methods, including standard germination tests, electrical conductivity tests, seedling growth tests, accelerated aging tests, cold tests, and tetrazolium tests, have been proposed and employed to evaluate seed germination. However, these methods typically require significant time, are non-automatic and may lead to seed destruction, and often necessitate specialized training and expertise. Consequently, they are not well-suited for large-scale applications or for protecting endangered species. Therefore, non-destructive and high-throughput screening methods are essential for the seed industry to provide high-quality seeds with superior characteristics to ensure the supply of high-germination seeds to farmers before planting [13].

In recent years, there have been significant advancements in electronic technologies and equipment, leading to notable improvements in the resolution and accuracy of light and image-based systems. These systems are now capable of determining qualitative indicators of the chemical components of materials, either in a static setting or online in production lines. This progress has enabled the fast and precise classification of materials with reduced labor requirements [14]. Light and image-based detection systems have been successfully employed in assessing the quality of agricultural food products, offering reliable and accurate results. By minimizing the influence of human intervention, these systems have become a preferred approach due to their consistency and stability [15–17]. Among the non-invasive methods used for identifying the chemical components of agricultural products, Near Infrared Spectroscopy (NIR) has gained widespread popularity in recent years. NIR operates based on the absorption of electromagnetic radiation within the wavelengths of 780 to 2500 nm [18]. When agricultural products are exposed to this radiation, their spectral response varies depending on the wavelength due to scattering and absorption processes. The tissue structures of these products, consisting of cells and intracellular/extracellular environments, are responsible for

 doi:10.20944/preprints202311.0733.v1

radiation scattering. Additionally, the absorption of electromagnetic rays is mainly influenced by C\\H, O\\H, and N\\H bonds present in major compounds such as water, sugars, chlorophylls, carotenoids, and so on. The NIR spectrum comprises broad wavebands resulting from the overlapping of absorption bands, which are closely associated with the colors and combinations of these chemical bonds. As a result, organic and biological substances can be effectively detected using NIR spectroscopy [19]. The investigation of artificially aged soybean seeds in comparison to healthy seeds revealed that changes in radiation absorption within the wavelength range of 1000–25000 nm can effectively distinguish between healthy and old seeds [20]. Similarly, differentiating between viable and non-viable soybean seeds, which underwent accelerated aging through heat treatment, was accomplished using NIR reflectance spectra in the wavelength range of 400–2500 nm. Partial least square discriminant analysis (PLS-DA) was employed in this research to classify viable and non-viable seeds, with the best model achieving an accuracy of 95% in the short-wave infrared (SWIR) region of 750–2500 nm [21]. In the case of tomato seeds subjected to accelerated aging, NIR spectroscopy in absorption mode, within the wavelength range of 911- 2258 nm, was utilized to classify viable and non-viable seeds. Both PLS-DA and interval partial least squares discriminant analysis (iPLS-DA) were employed to construct corresponding models. Specific spectral regions (1160–1170, 1383–1397, 1647–1666, 1884–1860, and 1915–1940 nm) were identified by iPLS-DA for the classification of viable and non-viable tomato seeds, resulting in a classification accuracy of 94% [22]. In a study focusing on spinach seeds, NIR spectroscopy within the wavelength range of 833–1667 nm was used to differentiate between viable and non-viable seeds. The optimal wavelengths were selected using successive projections algorithms (SPA), and classification models created with these 10 selected wavelengths demonstrated satisfactory accuracy in distinguishing viable seeds from non-viable ones [23]. Although previous studies revealed that the use of NIR spectroscopy for diagnosing seed viability had acceptable accuracy, the practical and commercial feasibility of using the entire wavelength range is not economical. This can be explained by the high cost associated with producing spectroscopic instruments based on the full spectrum [24]. Therefore, there is a need to explore methods for identifying optimal wavelengths that would enable the production of industrial-commercial tools at a lower cost. The application of chemometrics techniques in analyzing spectroscopic data poses a fundamental challenge due to the high dimensionality of the data set. This refers to situations where the number of features greatly exceeds the size of the data set itself [25]. For instance, in spectroscopic applications with a large number of wavelengths, the classification parameters also increase, leading to a significant decrease in the performance of the classification tool [26]. When obtaining a substantial number of training data becomes impractical, reducing the size of the feature subset becomes crucial as it helps in reducing the number of required training data and, in turn, enhances the performance of the classification algorithm [27]. Dimension reduction serves as a common method to address this challenge by removing noise and unnecessary features. It proves to be an efficient approach for improving accuracy, reducing computational complexity, building more generalized models, and reducing storage space requirements [28]. The main idea behind feature selection is to select a subset of features by eliminating those with little or no informative value and removing highly correlated features [29]. Generally, feature selection methods aim to optimize two conflicting objectives: maximizing the association with the target class and minimizing redundancy (correlation) among the selected features [30].

The current research aimed to develop an intelligent model based on NIR spectroscopy data and deep learning analysis to assess the viability of peanut seeds. To achieve this, two peanut cultivars, North Carolina-2 (NC-2) and Florispan, were selected, that were exposed to three levels of artificial aging. The NIR spectroscopy data of the samples were collected, and deep-learning approaches were employed to create predictive models for germination indicators. Moreover, meta-heuristic variable selection methods such as world competitive contest (WCC), League Championship Algorithm (LCA), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), imperialist competitive algorithm (ICA), learning automata (LA), heat transfer optimization algorithm (HTS), forest optimization algorithm (FOA), discrete symbiotic organism search (DSOS),

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

and cuckoo optimization (CUK) were used to select optimal wavelengths based on seed age and qualitative classification of peanuts.

2. Materials and Methods

2.1. Seed selection and aging treatment

Two common peanut seed cultivars for cultivation in Iran, namely North Carolina 2 (NC-2) and Florispan, were chosen for the experiments. Seeds from the last crop year were selected to ensure optimal conditions for survival and germination vigor. The seeds with similar mass and size were selected to minimize the impact of unfavorable factors on the experiment results. To induce artificial senescence, 300 seeds of each cultivar were subjected to accelerated aging treatment in three-time intervals, with 24 h between each interval. The seeds were placed in a single layer of aluminum nets positioned above water containers in an incubator. Before placing the seeds, the containers were thoroughly cleaned with a 15% sodium hypochlorite solution to prevent fungal contamination. The incubator was set to maintain a relative humidity of 85% and a temperature of 50 °C. At 24-h intervals, one-third of the samples were removed from the incubator, resulting in three different aging periods for the seeds [31].

2.2. Preparation of Vis/NIR Spectra from Samples

A PS-100 model spectroradiometer (Apogee Instruments, INC., Logan, UT, USA) was utilized to acquire the spectra of peanut seeds. This spectroradiometer is compact, lightweight, and portable, equipped with a sputtering-type monochromator with a resolution of 1 nm and a linear silicon CCD array detector containing 2048 pixels, covering the spectral range of 250–1150 nm (Vis/NIR). Furthermore, the spectroradiometer PS-100 can be connected to a computer via an optical fiber, and the acquired spectra are displayed and stored in the Spectra Wiz software through a USB port. For obtaining the absorption spectra of the samples, a probe-detector sensor was employed, which is designed in a way that the light source is positioned at a 45 ° angle to the detector, allowing for the acquisition of internal diffuse reflection rather than using mirror reflection acquisition from the sample. This internal diffuse reflection provides information about the internal contents of the sample. Figure 1 illustrates the process of acquiring the spectrum using the probe-detector sensor.

Figure 1. The spectrum acquisition process along with the probe-detector sensor.

2.3. Standard germination test

For the standard germination test, 100 samples were selected for each treatment. The germination test was conducted using the method of placing seeds between wet papers. The samples were placed in a germinator with a constant temperature of 25 °C and kept in these conditions for 10 days to facilitate germination. Before conducting the test, the containers used were disinfected with a 15% hypochlorite solution, and the peanut seeds were treated with 1% mercury chloride [32]. The identification and counting of normal and abnormal seedlings were performed was done according to the guidelines of the International Seed Testing Association (ISTA) from the fifth day up to the tenth day. On the tenth day, the seedlings were placed in a dryer at a temperature of 60 °C for 24 h [10]. The mass and length of the seedlings were measured using a scale with an accuracy of 0.0001 g and a caliper with an accuracy of 0.01 mm, respectively. Based on the counts and measurements, various indicators related to seed germination were calculated for each treatment group. These indicators include Germination Energy (GE), Mean Daily Germination (MDG), Germination Value (GV), Daily Germination Speed (DGS), and Germination Vigor (GVI). Additionally, the allometric coefficient (AC) was calculated for each seed. The relationships used to calculate the seed germination indices were presented in Table 1.

Table 1. Calculation relationships of the studied indicators.

The studied index	Equation	References
Germination Energy	$GE = \frac{MCGP}{N} \times 100$	[10]
Germination Value	$GV = MDG \times PV$	[33]
Germination Vigour	$GVI = GP \times \frac{Mean(PL + PR)}{100}$	[34]
Allometric Coefficien	$AC = \frac{PL}{PR}$	[35]
Daily Germination Speed	$DGS = \frac{1}{MDG}$	[36]
Mean Daily Germination	$MDG = \frac{GP}{T}$	[37]

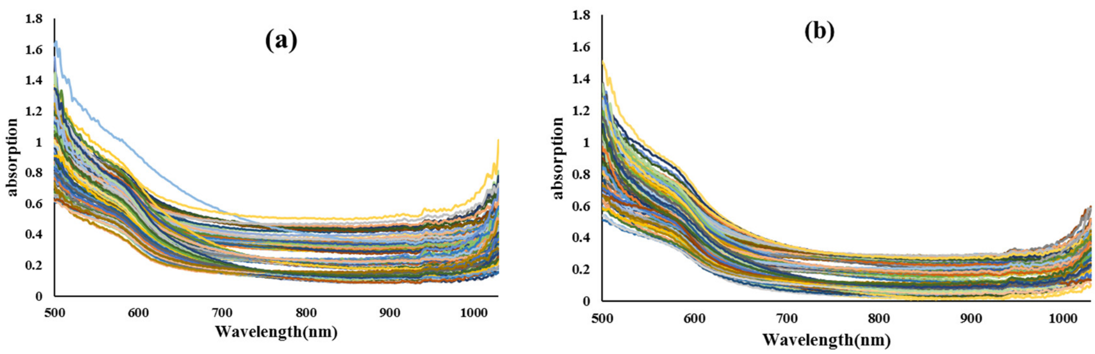
Where MCGP is the maximum percentage of cumulative germination, N is total number of seeds sown, ti is the number of days after the start of germination, GP is percentage of germination final yield, T is length of germination period (days), SFW is seedling wet weight (grams), SDW is seedling dry weight (grams), PLis seedling length (centimeters), PR: Root length (centimeters).

2.4. Preprocessing of Vis/NIR spectra

After acquiring the spectra and transferring them to the computer using Excel software, a single spectrum was created by averaging the two acquired spectra from the sides of the peanut, representing the index spectrum of each sample. Spectral data may contain irrelevant information and noise, such as fluorescence background, stray light, detector noise, cosmic rays, instrument noise, laser power fluctuations, and so on. To extract accurate information and enhance subtle differences between different samples, spectral pre-processing is a crucial step in spectral data analysis [38].

In this study, the first step of spectral pre-processing involved using the Savitzky-Golay (SG) method to smooth the curves and remove random noise. This method effectively smooths out slight fluctuations caused by noise in the curve while enhancing spectral peaks related to changes in the sample components [39,40]. Subsequently, multiplicative scatter correction (MSC) was employed to eliminate the noise caused by light scattering. MSC removes baseline translations and displacements caused by scattering effects between samples, thereby improving the signal-to-noise ratio of the original spectrum [41].

The Unscrambler 10.4 software was used for spectral data preprocessing. Figure 2 depicts the main spectrum curves of the samples and their pre-processed curves associated with each cultivar.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

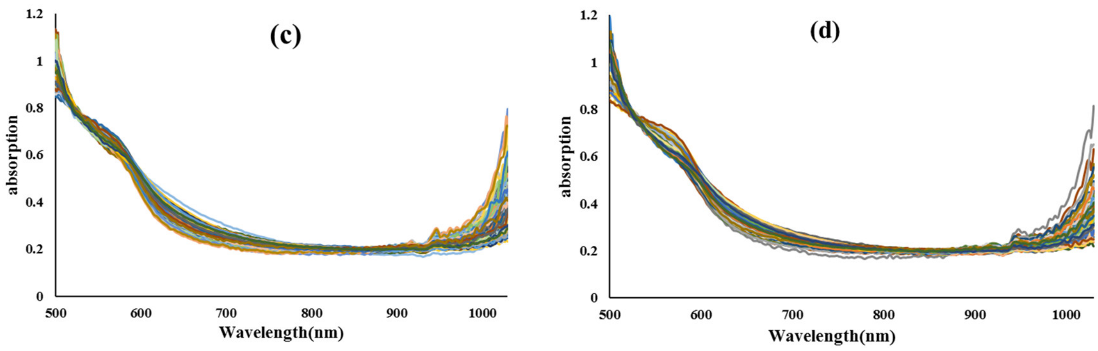


Figure 2. Raw and pre-processed VIS/NIR spectra of peanut seeds, a) NC-2 raw spectrum, b) Florispan raw spectrum, c) NC-2 pre-processing, and (d) Florispan pre-processing.

2.5. Methods of choosing the optimal wavelength

To address complex optimization problems with numerous variables, meta-heuristic algorithms are employed as suitable and efficient approaches. These algorithms rapidly provide approximate solutions, avoiding the need for time-consuming optimal solutions [42]. A key advantage of meta-heuristic algorithms is their ability to escape local optima, ensuring a more comprehensive search for optimal solutions [43]. The working principle of these algorithms involves introducing a set of initial solutions randomly. A fitness function is then calculated to assess the optimality of each solution in the initial population. If the statistical criteria for optimization quality are not met, the algorithm produces a new generation of solutions, repeating the cycle until the desired optimization criteria are satisfied [44]. Meta-heuristic approaches are typically categorized into two main groups: evolutionary algorithms (EA) and swarm intelligence (SI) [45]. EA draws inspiration from biological evolution mechanisms such as reproduction, mutation, recombination, and selection. In the context of optimization problems, the introduced solutions represent individuals within a population, and the fitness function evaluates the quality and accuracy of these solutions. Through iterative steps, the evolutionary algorithm facilitates the evolution of the initial population toward overall optimization [46]. In contrast, swarm intelligence optimization methods involve a collection of simple with no complexity of artificial agents. The concept behind SI algorithms is inspired by natural systems, where each agent performs a basic task. However, the interaction, cooperation, and somewhat random responses of these agents lead to emergent intelligent behavior that is not achievable by any individual agent alone [47]. SI-based feature selection methods have been utilized in previous research, and their operational principles have been thoroughly described in the literature [26]. In this study, the process of selecting optimal wavelengths was carried out using variable selection methods based on various meta-heuristic algorithms, including (WCC) [48], LCA [49], GA [50], PSO [51], ACO [52], ICA [53], LA [54], HTS [55], FOA [56]. The optimization of optimal wavelength selection was carried out using the FeatureSelect software package within MATLAB 2017 [57].

2.6. Modeling methods to predict seed viability

Traditional artificial intelligence methods typically employed a mathematical representation approach to describe optimization problems and discover optimal solutions under specific constraints based on logical mathematical principles. Such an approach is commonly referred to as knowledge-based [58]. Nevertheless, in most natural phenomena where predicting trends or classifying an occurrence should be conducted, logical-mathematical laws may not be able to describe these phenomena adequately. This limitation arises from the fact that these phenomena are often abstract and not easily captured by mathematical formulations [59]. To overcome the limitations of the knowledge-based approach, an alternative strategy inspired by human processes has been developed. Humans are capable of learning from repeated tasks, receiving feedback, and adjusting their decisions or actions accordingly to achieve favorable outcomes [60]. This approach, based on iterative learning from experience, is referred to as a learning-based approach, in contrast to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

knowledge-based approach [59]. Similarly, we can develop machines to perform specific tasks using a learning-based approach, known as Machine Learning (ML). In this current study, machine learning methods such as LR, DT, MP, SVM, k-NN, and NB have been employed in the WEKA 3.8.6 software package for detecting and classifying the vitality of seeds.

2.7. Evaluation criteria of optimal wavelength selection algorithms and machine learning models

To assess the effectiveness and performance of the optimal wavelength selection algorithms, two statistical measures were utilized: the root mean square error (RMSE) and the coefficients of determinate (R^2) [61,62]. These measures were calculated using the equations (1 and 2) [62–65]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \acute{y}_i)^2} \tag{1}$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \acute{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

Where y_i, $\acute{y}_i$ and $\bar{y}$ are predicted, actual, and mean values respectively.

To evaluate the accuracy of the classification models for peanut seed viability, several statistical criteria were employed, including accuracy, precision, sensitivity, specificity, and receiver operating characteristic (ROC). These criteria were calculated using the following equations: Accuracy refers to how close the measured value is to the true value. Precision indicates the closeness of successive measurements to each other (i.e., the consistency of errors in the various measurements). Sensitivity represents the fraction of positive cases correctly identified. Specificity denotes the fraction of negative cases correctly identified. These metrics were calculated using equations (3) to (6) [63].

$$\mathrm{Accuracy} = \frac{(\mathrm{True\ Positive} + \mathrm{True\ Negative})}{(\mathrm{True\ Positive} + \mathrm{False\ Negative}) + (\mathrm{False\ Negative} + \mathrm{True\ Negative})} \tag{3}$$

$$\mathrm{Precision} = \frac{\mathrm{True\ Positive}}{(\mathrm{True\ Positive} + \mathrm{False\ Positive})} \tag{4}$$

$$\mathrm{Sensitivity} = \frac{\mathrm{True\ Positive}}{(\mathrm{True\ Positive} + \mathrm{False\ Negative})} \tag{5}$$

$$\mathrm{Specificity} = \frac{\mathrm{True\ Negative}}{(\mathrm{False\ Positive} + \mathrm{True\ Negative})} \tag{6}$$

where in: True Positive is the number of samples in the i-th category that the algorithm correctly recognized. False Negative is the number of samples in the i-th category that the algorithm has misdiagnosed. False Positive is the number of samples outside the i-th category that the algorithm placed in the i-th category. True Negative is the number of samples outside the i-th category that the algorithm did not place in the i-th category.

3. Results and Discussion

3.1. Examination of seed viability indices

Table 2 presents the results related to seed viability indices for each treatment. Accordingly, it is evident that the accelerated aging test significantly impacted the seed groups, and there were noticeable differences in seed viability indices among the treatments. With an increase in the aging period, the germination percentage for both seed varieties significantly decreased, and in the third period, almost half of the seeds failed to germinate.

Table 2. Results of seed viability indices for peanut seeds.

Variety	Period	Number	Germination percentage	Germination Energy	Mean Daily Germination	Germination Value	Daily Germination Speed	Simple vitality index
NC-2	1	100	91	86	11.3750	250.2500	0.0879	7.0752
NC-2	2	100	69	61	8.6250	146.6250	0.1159	3.7826
NC-2	3	100	47	34	5.8750	64.6250	0.1702	0.9795
Florispan	1	100	94	92	11.7500	329.0000	0.0851	9.5836
Florispan	2	100	72	63	9.000	153.0000	0.1111	4.0761
Florispan	3	100	52	65	6.5000	84.5000	0.1538	2.0072

3.2. Comparison of the efficiency of optimal wavelength selection methods

In this study, to carry out the process of selecting the seed allometric coefficient as the continuous output variable (target) and wavelengths as the input variable (independent variable), regression was considered. The higher the seed allometric coefficient, the higher the likelihood of germination [10]. The selection of variables was considered a regression problem; thus, optimization algorithms were employed to search for wavelengths that create regression models with the highest correlation between actual values and predicted values. Table 3 provides descriptive statistical measures of algorithm accuracy and the number of optimally selected wavelengths by each algorithm.

Table 3. The results obtained from the variable selection algorithms for the regression problem.

AL	NOF	Wavelengths	ET	RMSE	CR(R^2)
WCC	14	704, 694, 775, 835,1025, 991, 906, 824, 852, 738, 795, 699, 963, 767	23.3603	0.0028	0.9870
LCA	10	748, 915, 783, 967, 887, 869, 801, 696, 744, 883	44.9389	0.0025	0.9872
GA	16	870, 799, 783, 636, 846, 785, 734, 737, 762, 954, 827, 913, 714, 810, 904, 725	3.4135	0.0028	0.9868
PSO	15	911, 784, 992, 713, 839, 726, 928, 840, 691, 791, 963, 832, 775, 737, 817	2.5181	0.0026	0.9870
ACO	16	780, 759, 868, 777, 814, 704, 804, 982, 952, 775, 1017, 934, 685, 905, 800, 657	31.1699	0.0027	0.9870
ICA	16	791, 844, 786, 731, 929, 1003, 798, 675, 1022, 774, 710, 888, 777, 978, 901, 697	2.3635	0.0027	0.9867
LA	15	935, 790, 768, 796, 776, 955, 732, 818, 883,694, 866, 1027, 783, 722, 824	33.8388	0.0025	0.9876
HTS	16	920, 828, 762, 804, 811, 503, 862, 837, 785, 779, 698, 846, 845, 957, 854, 633	14.3798	0.0028	0.9866
FOA	10	754, 825, 731, 778, 962, 902, 794, 738, 707, 856	17.1878	0.0025	0.9874
DSOS	16	915, 681, 672, 977, 815, 994, 956, 798, 939, 581, 522, 819, 690, 793, 760, 806	23.8145	0.0033	0.9854
CUK	12	874, 734, 706, 878, 1018, 775, 972, 742, 791, 843, 967, 723	33.3829	0.0027	0.9870

Where Al is Algorithm, NOF is The number of features, ET is Elapsed time, RMSE is the Root mean squared error, and CR(R2) is the Squared correlation coefficient.

Based on Table 3, it is evident that all variable selection algorithms exhibit a high level of accuracy (CR > 0.98) and low error (RMSE < 0.003) in predicting the allometric coefficient. Therefore, it appears that the most logical and appropriate criterion for selecting the optimal algorithm is based on its execution time. Algorithms that require less computational time are more practical for commercial-scale implementation [61,64]. The ranking of variable selection algorithms based on execution time is as follows: ICA < PSO < GA < HTS < FOA < WCC < DSOS < ACO < CUK < LA < LCA. Previous research has also indicated the high popularity and practicality of ICA [66], PSO [67], and GA [68], algorithms due to their low execution time.

On the other hand, the number of selected wavelengths is also a crucial criterion in industrial-scale applications because the cost of producing spectroscopic tools for practical purposes is directly dependent on the number of detectable wavelengths by the instrument. As the number of wavelengths decreases, the production cost decreases accordingly [14]. Additionally, instruments capable of detecting fewer wavelengths can provide higher accuracy and resolution in their measurements [69]. Therefore, the variable selection algorithms are ranked based on the number of optimal wavelengths as follows: LCA < FOA < CUK < WCC < PSO < GA < DSOS < ACO < HTS < LA < ICA. The LCA and FOA algorithms perform superior that others by identifying 10 optimal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

wavelengths. Considering that the FOA algorithm's execution time is twice as fast, it can be considered the optimal method for wavelength selection.

In Figure 3, the algorithms' performance is compared in terms of correlation and RMSE in each round of algorithm execution. The LA algorithm's results showed a strong correlation between the number of executions and its performance, achieving lower error and higher correlation with each round of execution. Conversely, the DSOS algorithm is relatively unaffected by more executions, as re-executing the algorithm does not result in significant changes in its error rate and correlation. These findings align with the results presented by Masoudi-Sobhanzadeh et al. [57], who implemented and compared the algorithms using different datasets.

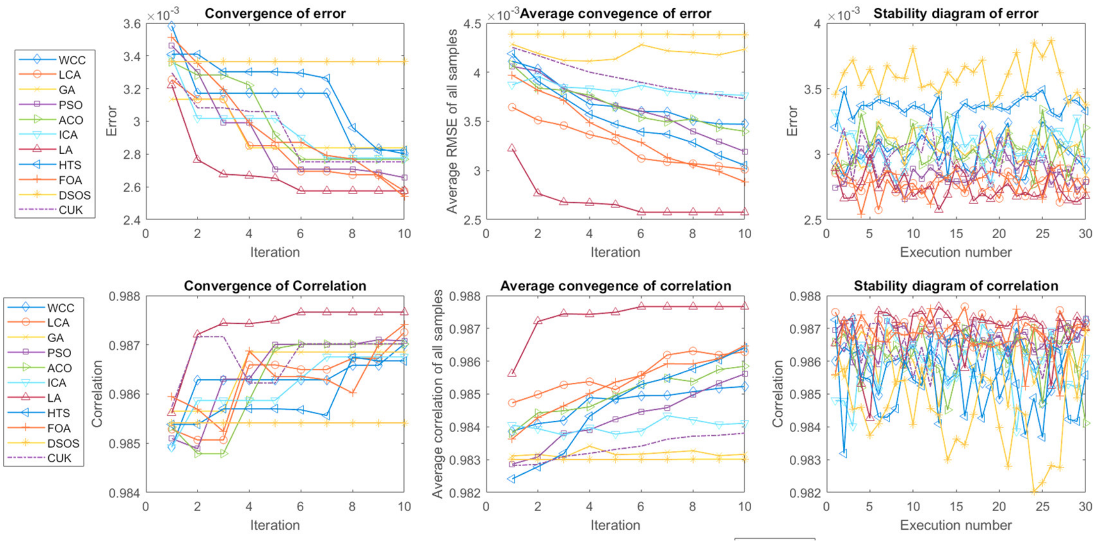


Figure 3. Comparison of correlation and RMSE of variable selection algorithms based on convergence, mean convergence, and stability.

3.3. Examination of averages of Vis/NIR absorption spectra and evaluation of the location of the selected wavelength

Figure 4 displayed the mean Vis/NIR absorption spectra for three seed aging periods along with the locations of the optimum selected wavelengths. Accordingly, as increasing seeds' age increases, the absorption levels decrease along the entire curve. This phenomenon can be attributed to the reduction of water in the sample due to the aging process, because the amount of radiation absorption largely depends on the amount of water in the chemical components [40]. Similar trends have been reported by other researchers [45,69,70]. The absorption changes in the spectral range of 500–550 nm can be attributed to carotenoids and anthocyanins [71]. Furthermore, changes in the range of 650–700 nm are related to the presence of chlorophyll [72], whereas changes in 680–760 nm were associated with the presence of amino acids in the seed [70]. On the other hand, changes in the region of 750–850 nm were attributed to the water content in the sample [73]. The alterations in the 860–910 nm indicate the presence of CH and OH bonds in carbohydrates [74]. Furthermore, changes in the 930–1030 nm were explained by the presence of protein compounds in the sample [75]. The presence of each of these substances in the seed composition increases the probability of germination, highlighting the importance of their detection [71]. Hence, it can be concluded that the variable selection algorithm provides an optimal mode that has selected at least one wavelength in all the mentioned ranges.

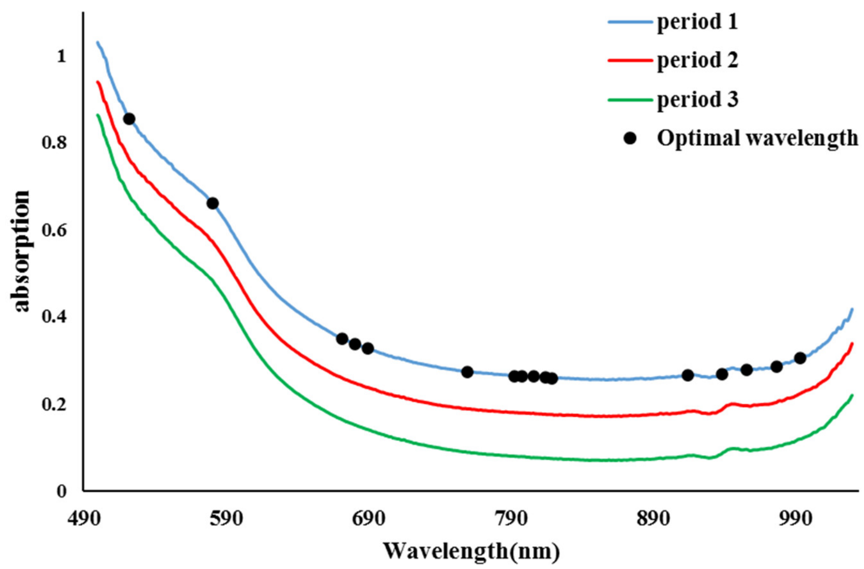


Figure 4. Averages of Vis/NIR absorption spectra and the location of the selected wavelengths.

3.4. Results of seed viability classification modeling based on selected wavelengths by machine learning methods

Table 4 presented the results of the seed viability classification models based on the selected wavelengths using the FOA algorithm (lowest number of wavelengths) and DSOS algorithm (highest number of wavelengths). As shown in Table 4, all classifications exhibited good performance in determining the viability of the seeds. Overall, it can be concluded that the seed classification using selected wavelengths by the FOA algorithm performed better. However, to more precisely check and compare the classifiers' performance, the results are graphically presented in Figure 5.

Table 4. Classifications results to determine seeds viability.

Algorithm	classifiers	Accuracy	Precision	Sensitivity	Specificity	ROC Area
FOA	Logistic Regression	0.9750	0.9260	0.9250	0.9850	0.9960
	Naive Bayes	0.9770	0.9370	0.9330	0.9860	0.9970
	Decision trees	0.9820	0.9460	0.9460	0.9890	0.9820
	k-Nearest Neighbors	0.9790	0.9380	0.9380	0.9870	0.9630
	Support Vector Machines	0.9770	0.9290	0.9210	0.9840	0.9860
	Multilayer Perceptron	0.9830	0.9500	0.9500	0.9900	0.9980
DSOS	Logistic Regression	0.9580	0.8750	0.8750	0.9750	0.9880
	Naive Bayes	0.9710	0.9160	0.9130	0.9820	0.9930
	Decision trees	0.9930	0.9800	0.9790	0.9960	0.9890
	k-Nearest Neighbors	0.9610	0.8840	0.8830	0.9770	0.9300
	Support Vector Machines	0.9620	0.8990	0.8880	0.9770	0.9790
	Multilayer Perceptron	0.9830	0.9500	0.9500	0.9900	0.9980

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

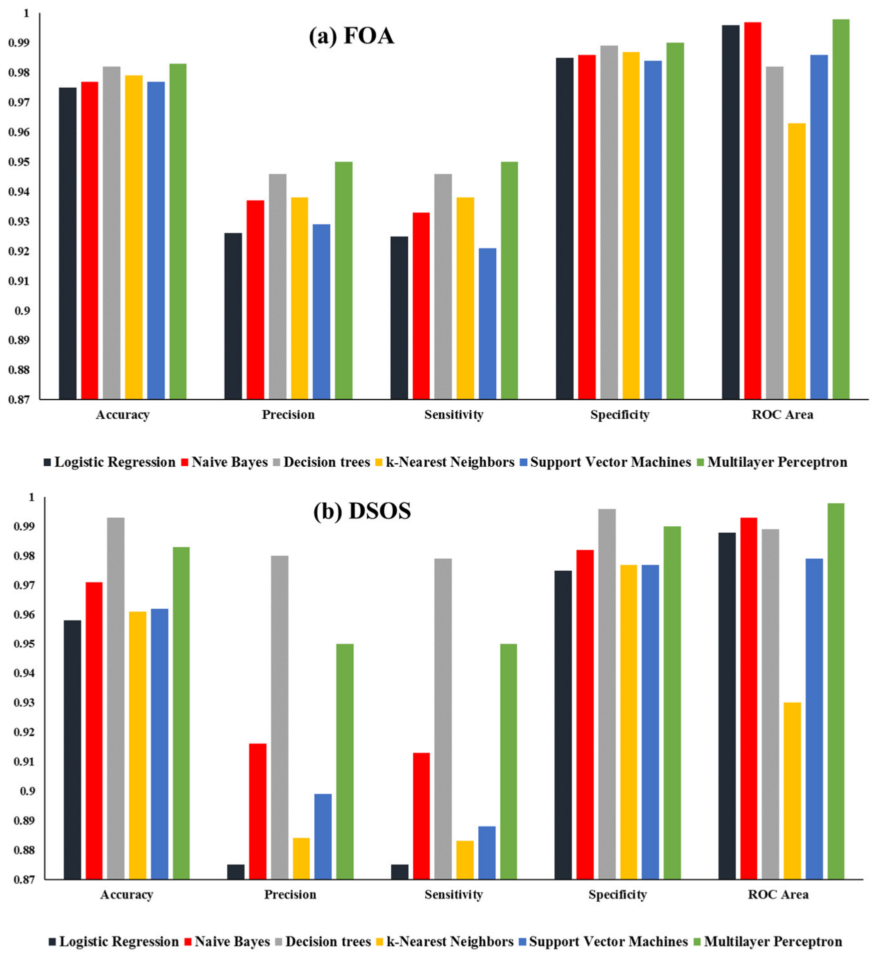


Figure 5. Results of classification for (a) seed viability identification, b) FOA algorithm selected wavelength, and c) DSOS algorithm selected wavelength.

As shown in Figure 5, except for the DT and MP methods, the effectiveness of other methods decreased with an increase in the number of variables. The MP method provides consistent results with both the number of variables (10 and 16 variables), and the DT method exhibits the best performance with increasing variables. LR, SVM, and K-NN methods were strongly affected by the increased number of variables, leading to a sharp drop in their performance. Previous research has indicated that two methods, MP and DT, have favorable capabilities in data mining of high-scale data with many variables. This can be explained by the fact that LR and SVM methods are not affected by the collinearity of variable problems [76]. However, the DT method, with its node-branch expansion capability, can check different features in various branches and overcome the problem of collinearity of variables, which is essential for achieving good performance [59,63]. On the other hand, the MP method features a layered structure comprising several neurons in each layer. Each neuron in this method models a set of features, thus providing a solution to the problem of non-linearity among variables [77]. However, the LR approach, when executed with a limited number of variables, possesses an advantage over other methods, as it can explicitly describe the relationship between each wavelength and the response variable and rank the importance of wavelengths in the target classification [60]. Hence, considering the high accuracy of all algorithms in seed classification, no particular algorithm can be considered superior to others. Consequently, the appropriate algorithm can be chosen based on specific research or operational goals. Various classification methods have been employed in different research to classify seed viability. For instance, corn seed viability was classified using SVM, KNN, random forest, and a deep convolutional neural network (CNN), with the best accuracy achieved using CNN [31]. Peanut seed viability was classified using SVM, DT, and LDA, and the best accuracy was obtained using the DT classification [31]. Hyperspectral images were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

used to identify damaged rice seeds with SVM, KNN, DT, and deep forest classifiers, with the new DF classifier, developed based on the DT classifier, providing higher accuracy than other classifiers [78]. The germination vigur of sugar beet seeds was predicted using hyperspectral images and KNN, SVM, and RF classifiers, with the SVM classifier providing the best performance [70].

Figures 6–9 displayed the confusion matrices for DT-DSOS, MP-FOA, LR-DSOS, and KNN-DSOS classifications, respectively, enabling a clear comparison between the best and worst classifications in each treatment. Accordingly, the most accurate diagnosis was related to the third senescence period for both seed varieties, and seeds with poor viability were correctly identified and distinguished from healthy seeds. Additionally, the correct identification of healthy seeds was acceptable, with most misclassified seeds belonging to the second senescence period. As a result, there was no significant difference in the accuracy of the correct diagnosis based on the seed variety.

Treatment	Number	a	b	c	d	e	f	True positive rate
a: NC-2- Period 1	40	38	2	0	0	0	0	95%
b: Florispan- Period 1	40	0	40	0	0	0	0	100%
c: NC-2- Period 2	40	0	1	38	1	0	0	95%
d: Florispan- Period 2	40	0	0	1	39	0	0	97.5%
e: NC-2- Period 3	40	0	0	0	0	40	0	100%
f: Florispan- Period 3	40	0	0	0	0	0	40	100%
False positive rate		0%	1.5%	0.5%	0.5%	0%	0%	

Figure 6. Disturbance matrix for DT-DSOS classification.

Treatment	Number	a	b	c	d	e	f	True positive rate
a: NC-2- Period 1	40	37	3	0	0	0	0	92.5%
b: Florispan- Period 1	40	5	34	1	0	0	0	85%
c: NC-2- Period 2	40	0	1	38	1	0	0	95%
d: Florispan- Period 2	40	0	0	1	39	0	0	97.5%
e: NC-2- Period 3	40	0	0	0	0	40	0	100%
f: Florispan- Period 3	40	0	0	0	0	0	40	100%
False positive rate		2.5%	2%	1%	0.5%	0%	0%	

Figure 7. Perturbation matrix for MP-FOA application.

Treatment	Number	a	b	c	d	e	f	True positive rate
a: NC-2- Period 1	40	38	2	0	0	0	0	95%
b: Florispan- Period 1	40	2	35	3	0	0	0	87.5%
c: NC-2- Period 2	40	0	4	31	5	0	0	77.5%
d: Florispan- Period 2	40	0	0	5	32	3	0	80%
e: NC-2- Period 3	40	0	0	0	3	36	1	90%
f: Florispan- Period 3	40	0	0	0	0	2	38	95%
False positive rate		1%	3%	4%	4%	2.5%	0.5%	

Figure 8. Perturbation matrix for LR-DSOS application.

 doi:10.20944/preprints202311.0733.v1

Treatment	Number	a	b	c	d	e	f	True positive rate
a: NC-2- Period 1	40	34	6	0	0	0	0	85%
b: Florispan- Period 1	40	6	33	1	0	0	0	82.5%
c: NC-2- Period 2	40	0	3	36	1	0	0	90%
d: Florispan- Period 2	40	0	0	2	35	3	0	87.5%
e: NC-2- Period 3	40	0	0	0	3	35	2	87.5%
f: Florispan- Period 3	40	0	0	0	0	1	39	97.5%
False positive rate		3%	4.5%	1.5%	2%	2%	1%	

Figure 9. Perturbation matrix for KNN-DSOS classification.

4. Conclusions

In this research, Vis/NIR spectroscopic technology in absorption mode was utilized to detect aging and classify two varieties of peanut seeds. The results revealed significant differences between the spectral curves of the aging treatments, with healthy (young) seeds displaying higher absorption compared to unhealthy (artificially aged) seeds. Notably, the differences in absorption properties due to aging were so distinctive that the two seed varieties did not hinder the identification of healthy seeds from unhealthy ones. Based on the absorption spectrum curves of different treatments, it can be concluded that seed aging induces changes in its chemical components, which can be effectively detected through Vis/NIR spectroscopy.

This study employed meta-heuristic optimization algorithms for the optimal wavelength selection process. The results demonstrated that all these algorithms exhibited a high capability in identifying the optimal wavelengths and achieved excellent accuracy in modeling the seed's allometric coefficient. The algorithms achieved correlation coefficients higher than 0.985 and errors lower than 0.0036, respectively. However, the primary difference in the performance of the algorithms depended on the user's specific objectives. ICA < PSO < GA algorithms demonstrated much shorter execution times than other algorithms, with execution times of 2.3635 < 2.5181 < 3.4135, respectively. However, these methods selected 60% more optimal wavelengths. Therefore, these algorithms appear more suitable for online and fast applications. On the other hand, LCA < FOA < CUK algorithms introduced the least number of optimal wavelengths. However, their execution times were significantly longer than other algorithms, with values of 33.3829, 17.1878, and 44.9389, respectively. This limitation may be a significant drawback in online applications.

In this study, LR, SVM, KNN, DT, MP, and NaiveBayes classifiers were utilized for seed classification. The highest accuracies were achieved by the DSOS-DT and FOA-MP methods, with accuracies of 0.993 and 0.983, respectively. Conversely, the lowest accuracies were obtained by the DSOS-LR and DSOS-KNN methods, with values of 0.958 and 0.961, respectively. The results of this study indicated that the DT and MP algorithms exhibit superior performance in seed classification, and their classification performance remains unaffected by an increase in the number of wavelengths. However, other classifiers appear to be sensitive to the number of wavelengths, leading to relatively weak classification performance. In conclusion, it can be concluded that the combination of Vis/NIR spectroscopic technology with chemometrics techniques can provide promising prospects for practical applications in seed viability detection tools.

Author Contributions: Conceptualization, M.R.-S., A.M. and M.T.; methodology, M.R.-S., Y.A.-G., A.M. and M.T.; software, M.R.-S., A.M. and M.T.; validation, M.R.-S., Y.A.-G.; A.M.; formal analysis Y.A.-G.; M.K.; investigation, Y.A.-G.; M.R.-S., M.K.; resources, M.R.-S., M.K.; data curation, M.R.-S., Y.A.-G., writing—original draft preparation, M.R.-S.; A.M. and M.T.; writing—review and editing, Y.A.-G., M.M.-A., M.H.-H., J.L.H.-H.; visualization, M.R.-S.; supervision, Y.A.-G.; project administration, Y.A.-G.; funding acquisition, Y.A.-G.; M.M.-A., M.H.-H., J.L.H.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

References

1. Zou, S., Tseng, Y. C., Zare, A., Rowland, D. L., Tillman, B. L., & Yoon, S. C. (2019). Peanut maturity classification using hyperspectral imagery. *biosystems engineering*, *188*, 165-177.
2. Deepa, R., Anandhi, A., Bailey, N. O., Grace III, J. M., Betiku, O. C., & Muchovej, J. J. (2022). Potential Environmental Impacts of Peanut Using Water Footprint Assessment: A Case Study in Georgia. *Agronomy*, *12*(4), 930.
3. Wang, X., Chen, C. Y., Dang, P., Carter, J., Zhao, S., Lamb, M. C., ... & Feng, Y. (2023). Variabilities in symbiotic nitrogen fixation and carbon isotope discrimination among peanut (Arachis hypogaea L.) genotypes under drought stress. *Journal of Agronomy and Crop Science*, *209*(2), 228-241.
4. Tan, X. L., Azam-Ali, S., Goh, E. V., Mustafa, M., Chai, H. H., Ho, W. K., ... & Massawe, F. (2020). Bambara groundnut: An underutilized leguminous crop for global food security and nutrition. *Frontiers in Nutrition*, *7*, 601496.
5. Rahman, A., & Cho, B. K. (2016). Assessment of seed quality using non-destructive measurement techniques: a review. *Seed Science Research*, *26*(4), 285-305.
6. Kaur, N., Erickson, T. E., Ball, A. S., & Ryan, M. H. (2017). A review of germination and early growth as a proxy for plant fitness under petrogenic contamination—knowledge gaps and recommendations. *Science of The Total Environment*, *603*, 728-744.
7. Bastos, L. L. D. S., Calvi, G. P., Lima Júnior, M. D. J. V., & Ferraz, I. D. K. (2021). Degree of seed desiccation sensitivity of the Amazonian palm Oenocarpus bacaba depends on the criterion for germination. *Acta Amazonica*, *51*, 85-90.
8. Carrera-Castaño, G., Calleja-Cabrera, J., Pernas, M., Gómez, L., & Oñate-Sánchez, L. (2020). An updated overview on the regulation of seed germination. *Plants*, *9*(6), 703.
9. Marcos Filho, J. (2015). Seed vigor testing: an overview of the past, present and future perspective. *Scientia agricola*, *72*, 363-374.
10. Moghaddam, S. S., Rahimi, A., Noorhosseini, S. A., Heydarzadeh, S., & Mirzapour, M. (2018). Effect of seed priming with salicylic acid on germinability and seedling vigor fenugreek (Trigonella Foenum-Graecum). *Yuzuncu Yıl University Journal of Agricultural Sciences*, *28*(2), 192-199.
11. Ebrahimi, M., & Miri, E. (2016). Effect of humic acid on seed germination and seedling growth of Borago officinalis and Cichorium intybus. *Ecopersia*, *4*(1), 1239-1249.
12. Mohajeri, F., Taghvaei, M., Ramrudi, M., & Galavi, M. (2016). Effect of priming duration and concentration on germination behaviors of (Phaseolus vulgaris L.) seeds. *Int. J. Ecol. Environ. Conserv*, *22*, 603-609.
13. Xia, Y., Xu, Y., Li, J., Zhang, C., & Fan, S. (2019a). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, *1*, 35-47.
14. Li, A., Yao, C., Xia, J., Wang, H., Cheng, Q., Penty, R., ... & Pan, S. (2022). Advances in cost-effective integrated spectrometers. *Light: Science & Applications*, *11*(1), 174.
15. Abasi, S., Minaei, S., Jamshidi, B., & Fathi, D. (2018). Dedicated non-destructive devices for food quality measurement: A review. *Trends in Food Science & Technology*, *78*, 197-205.
16. El-Mesery, H. S., Mao, H., & Abomohra, A. E. F. (2019). Applications of non-destructive technologies for agricultural and food products quality inspection. *Sensors*, *19*(4), 846.
17. Ali, M. M., & Hashim, N. (2022). Non-destructive methods for detection of food quality. In *Future Foods* (pp. 645-667). Academic Press.
18. Wei, X., Xu, N., Wu, D., & He, Y. (2014). Determination of branched-amino acid content in fermented Cordyceps sinensis mycelium by using FT-NIR spectroscopy technique. *Food and Bioprocess Technology*, *7*, 184-190.
19. Xia, Y., Huang, W., Fan, S., Li, J., & Chen, L. (2019b). Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Physics & Technology*, *97*, 467-477.
20. Kusumaningrum, D., Lee, H., Lohumi, S., Mo, C., Kim, M. S., & Cho, B. K. (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.
21. Ambrose, A., Kandpal, L. M., Kim, M. S., Lee, W. H., & Cho, B. K. (2016). High speed measurement of corn seed viability using hyperspectral imaging. *Infrared Physics & Technology*, *75*, 173-179.
22. Shrestha, S., Deleuran, L. C., & Gislum, R. (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.

23. Lakshmanan, M. K., Boelt, B., & Gislum, R. (2023). A chemometric method for the viability analysis of spinach seeds by near infrared spectroscopy with variable selection using successive projections algorithm. *Journal of Near Infrared Spectroscopy*, *31*(1), 24-32.
24. Li, L., Chen, S., Deng, M., & Gao, Z. (2022). Optical techniques in non-destructive detection of wheat quality: A review. *Grain & Oil Science and Technology*, *5*(1), 44-57.
25. Rostami, M., Berahmand, K., & Forouzandeh, S. (2020). A novel method of constrained feature selection by the measurement of pairwise constraints uncertainty. *Journal of Big Data*, *7*(1), 1-21.
26. Rostami, M., Berahmand, K., Nasiri, E., & Forouzandeh, S. (2021). Review of swarm intelligence-based feature selection methods. *Engineering Applications of Artificial Intelligence*, *100*, 104210.
27. Gokalp, O., Tasci, E., & Ugur, A. (2020). A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Systems with Applications*, *146*, 113176.
28. Tang, X., Dai, Y., & Xiang, Y. (2019). Feature selection based on feature interactions with application to text categorization. *Expert Systems with Applications*, *120*, 207-216.
29. Liu, Y., Nie, F., Gao, Q., Gao, X., Han, J., & Shao, L. (2019a). Flexible unsupervised feature extraction for image classification. *Neural Networks*, *115*, 65-71.
30. Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 52.
31. Zou, Z., Chen, J., Zhou, M., Zhao, Y., Long, T., Wu, Q., & Xu, L. (2022). Prediction of peanut seed vigor based on hyperspectral images. *Food Science and Technology*, *42*.
32. Hampton, J. G., & Tekrony, D. M. (1995). *Handbook of vigour test methods*. The International Seed Testing Association, Zurich (Switzerland).
33. ISTA. 2009. International rules for seed testing. Annexes. Seed Science and Technology Journal. 49, 86-41.
34. Panwar, P., & Bhardwaj, S. D. (2005). *Handbook of practical forestry*. Agrobios (India).
35. ISTA. 1979. The germination test. InternationalSeed Testing Association. Seed Science and Technology, 4, 23-28.
36. Maguire, J. D. (1962). Speed of germination-aid in selection and evaluation for seedling emergence and vigor. *Crop Sci.*, *2*, 176-177.
37. Hunter, E. A., Glasbey, C. A., & Naylor, R. E. L. (1984). The analysis of data from germination tests. *The Journal of Agricultural Science*, *102*(1), 207-213.
38. Xu, Y., Zhong, P., Jiang, A., Shen, X., Li, X., Xu, Z., ... & Lei, H. (2020). Raman spectroscopy coupled with chemometrics for food authentication: A review. *TrAC Trends in Analytical Chemistry*, *131*, 116017.
39. Tie-cheng, B., Tao, W., You-qi, C., & Mercatoris, B. (2019). Comparison of near-infrared spectrum pretreatment methods for Jujube leaf moisture content detection in the sand and dust area of Southern Xinjiang. *Spectroscopy and Spectral Analysis*, *39*(4), 1323-1328.
40. Pang, L., Wang, J., Men, S., Yan, L., & Xiao, J. (2021). Hyperspectral imaging coupled with multivariate methods for seed vitality estimation and forecast for Quercus variabilis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *245*, 118888.
41. Chu, Y. W., Tang, S. S., Ma, S. X., Ma, Y. Y., Hao, Z. Q., Guo, Y. M., ... & Zeng, X. Y. (2018). Accuracy and stability improvement for meat species identification using multiplicative scatter correction and laser-induced breakdown spectroscopy. *Optics Express*, *26*(8), 10119-10127.
42. Zhang, S., Lee, C. K., Chan, H. K., Choy, K. L., & Wu, Z. (2015). Swarm intelligence applied in green logistics: A literature review. *Engineering Applications of Artificial Intelligence*, *37*, 154-169.
43. Hu, Y., Zheng, J., Zou, J., Yang, S., Ou, J., & Wang, R. (2020). A dynamic multi-objective evolutionary algorithm based on intensity of environmental change. *Information Sciences*, *523*, 49-62.
44. Wang, C., Pan, H., & Su, Y. (2020). A many-objective evolutionary algorithm with diversity-first based environmental selection. *Swarm and evolutionary computation*, *53*, 100641.
45. Zhang, L., Sun, H., Rao, Z., & Ji, H. (2020). Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds. *Spectrochimica acta part A: molecular and biomolecular spectroscopy*, *229*, 117973.
46. Gong, D., Xu, B., Zhang, Y., Guo, Y., & Yang, S. (2019). A similarity-based cooperative co-evolutionary algorithm for dynamic interval multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, *24*(1), 142-156.
47. Yong, Z., Dun-wei, G., & Wan-qiu, Z. (2016). Feature selection of unreliable data using an improved multi-objective PSO algorithm. *Neurocomputing*, *171*, 1281-1290.
48. Masoudi-Sobhanzadeh, Y., & Motieghader, H. (2016). World Competitive Contests (WCC) algorithm: A novel intelligent optimization algorithm for biological and non-biological problems. *Informatics in Medicine Unlocked*, *3*, 15-28.
49. Kashan, A. H. (2014). League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships. *Applied Soft Computing*, *16*, 171-200.
50. McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of computational and Applied Mathematics*, *184*(1), 205-222.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0733.v1

51. Shi, Y. (2001, May). Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)* (Vol. 1, pp. 81-86). IEEE.
52. Dorigo, M., Bonabeau, E., & Theraulaz, G. (2000). Ant algorithms and stigmergy. *Future generation computer systems*, *16*(8), 851-871.
53. Atashpaz-Gargari, E., & Lucas, C. (2007, September). Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In *2007 IEEE congress on evolutionary computation* (pp. 4661-4667). Ieee.
54. Beigy, H., & Meybodi, M. R. (2009). Cellular learning automata with multiple learning automata in each cell and its applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *40*(1), 54-65.
55. Patel, V. K., & Savsani, V. J. (2015). Heat transfer search (HTS): a novel optimization algorithm. *Information sciences*, *324*, 217-246.
56. Ghaemi, M., & Feizi-Derakhshi, M. R. (2016). Feature selection using forest optimization algorithm. *Pattern Recognition*, *60*, 121-129.
57. Masoudi-Sobhanzadeh, Y., Motieghader, H., & Masoudi-Nejad, A. (2019). FeatureSelect: a software for feature selection based on machine learning approaches. *BMC bioinformatics*, *20*(1), 1-17.
58. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; pp. 1–3.
59. Park, H., & Son, J. H. (2021). Machine learning techniques for THz imaging and time-domain spectroscopy. *Sensors*, *21*(4), 1186.
60. Meza Ramirez, C. A., Greenop, M., Ashton, L., & Rehman, I. U. (2021). Applications of machine learning in spectroscopy. *Applied Spectroscopy Reviews*, *56*(8-10), 733-763.
61. Kumar, A., Alsadoon, A., Prasad, P. W. C., Abdullah, S., Rashid, T. A., Pham, D. T. H., & Nguyen, T. Q. V. (2022). Generative adversarial network (GAN) and enhanced root mean square error (ERMSE): deep learning for stock price movement prediction. *Multimedia Tools and Applications*, 1-19.
62. Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609-629.
63. Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, *54*(1), 17-24.
64. Dhakate, P. P., Patil, S., Rajeswari, K., & Abin, D. (2014). Preprocessing and Classification in WEKA using different classifiers. *Inter J Eng Res Appl*, *4*(8), 91-3.
65. Ferreira, A. J., & Figueiredo, M. A. (2012). Efficient feature selection filters for high-dimensional data. *Pattern recognition letters*, *33*(13), 1794-1804.
66. Kumar, C. A., Sooraj, M. P., & Ramakrishnan, S. (2017). A comparative performance evaluation of supervised feature selection algorithms on microarray datasets. *Procedia computer science*, *115*, 209-217.
67. Majd, A., Sahebi, G., Daneshtalab, M., Plosila, J., Lotfi, S., & Tenhunen, H. (2018). Parallel imperialist competitive algorithms. *Concurrency and Computation: Practice and Experience*, *30*(7), e4393.
68. Atashpendar, A., Dorronsoro, B., Danoy, G., & Bouvry, P. (2018). A scalable parallel cooperative coevolutionary PSO algorithm for multi-objective optimization. *Journal of Parallel and Distributed Computing*, *112*, 111-125.
69. Ramdania, D. R., Irfan, M., Alfarisi, F., & Nuraiman, D. (2019, December). Comparison of genetic algorithms and Particle Swarm Optimization (PSO) algorithms in course scheduling. In *Journal of Physics: Conference Series* (Vol. 1402, No. 2, p. 022079). IOP Publishing.
70. Fu, J., Yu, H. D., Chen, Z., & Yun, Y. H. (2022). A review on hybrid strategy-based wavelength selection methods in analysis of near-infrared spectral data. *Infrared Physics & Technology*, 104231.
71. Yang, J., Sun, L., Xing, W., Feng, G., Bai, H., & Wang, J. (2021). Hyperspectral prediction of sugarbeet seed germination based on gauss kernel SVM. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *253*, 119585.
72. Saputri, D. A. S., Pahlawan, M. F. R., Murti, B. M., & Masithoh, R. E. (2022, June). Vis/NIR spectroscopy for non-destructive method in detecting soybean seeds viability. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1038, No. 1, p. 012043). IOP Publishing.
73. Pahlawan, M. F. R., Wati, R. K., & Masithoh, R. E. (2021). Development of a low-cost modular VIS/NIR spectroscopy for predicting soluble solid content of banana. In *IOP Conference Series: Earth and Environmental Science* (Vol. 644, No. 1, p. 012047). IOP Publishing.
74. Wati, R. K., Pahlawan, M. F. R., & Masithoh, R. E. (2021, March). Development of calibration model for pH content of intact tomatoes using a low-cost Vis/NIR spectroscopy. In *IOP Conference Series: Earth and Environmental Science* (Vol. 686, No. 1, p. 012049). IOP Publishing.
75. Savi, A., De Aguiar, L. M., Tonial, L. M. S., Lafay, C. B. B., Assmann, T. S., & De Bortolli, M. A. (2019). Fast and Non-Destructive Determination of N, P, and K in Sorghum, Oat, and Corn Residue Using Near-Infrared Spectroscopy. *J. Agric. Sci*.
76. Yan, Y. L., Chen, B., & Zhu, D. (2013). Near infrared spectroscopy-principles, technologies and applications. *Beijing, China*., pp. 21–27.

 doi:10.20944/preprints202311.0733.v1

77. Arora, R. (2012). Comparative analysis of classification algorithms on different datasets using WEKA. *International Journal of Computer Applications*, *54*(13).
78. Neo, E. R. K., Yeo, Z., Low, J. S. C., Goodship, V., & Debattista, K. (2022). A review on chemometric techniques with infrared, Raman and laser-induced breakdown spectroscopy for sorting plastic waste in the recycling industry. *Resources, Conservation and Recycling*, *180*, 106217.