


Article

Non-Destructive Seed Viability Assessment via Multispectral Imaging and Stacking Ensemble Learning

Ye Rin Chu , Min Su Jo, Ga Eun Kim, Cho Hee Park, Dong Jun Lee, Sang Hoon Che and Chae Sun Na * 

Forest Bioresources Department, Baekdudaegan National Arboretum, Bonghwa 36209, Republic of Korea; yerinchu@koagi.or.kr (Y.R.C.); jms991@koagi.or.kr (M.S.J.); kge120212@koagi.or.kr (G.E.K.); epurerbeau@koagi.or.kr (C.H.P.); senlife@koagi.or.kr (D.J.L.); che1004@koagi.or.kr (S.H.C.)

* Correspondence: chaesun.na@koagi.or.kr

Abstract: The tetrazolium (TZ) test is a reliable but destructive method for identifying viable seeds. In this study, a non-destructive seed viability analysis method for *Allium ulleungense* was developed using multispectral imaging and stacking ensemble learning. Using the Videometerlab 4, multispectral imaging data were collected from 390 *A. ulleungense* seeds subjected to NaCl-accelerated aging treatments with three repetitions per treatment. Spectral values were obtained at 19 wavelengths (365–970 nm), and seed viability was determined using the TZ test. Next, 80% of spectral values were used to train Decision Tree, Random Forest, LightGBM, and XGBoost machine learning models, and 20% were used for testing. The models classified viable and non-viable seeds with an accuracy of 95–91% on the K-Fold value ($n = 5$) and 85–81% on the test data. A stacking ensemble model was developed using a Decision Tree as the meta-model, achieving an AUC of 0.93 and a test accuracy of 90%. Feature importance and SHAP value assessments identified 570, 645, and 940 nm wavelengths as critical for seed viability classification. These results demonstrate that machine learning-based spectral data analysis can be effectively used for seed viability assessment, potentially replacing the TZ test with a non-destructive method.

Keywords: seed viability test; multispectral imaging; stacking ensemble learning; *Allium*



Citation: Chu, Y.R.; Jo, M.S.; Kim, G.E.; Park, C.H.; Lee, D.J.; Che, S.H.; Na, C.S. Non-Destructive Seed Viability Assessment via Multispectral Imaging and Stacking Ensemble Learning. *Agriculture* **2024**, *14*, 1679. <https://doi.org/10.3390/agriculture14101679>

Academic Editors: Wei Lu and Zhao Zhang

Received: 2 September 2024

Revised: 20 September 2024

Accepted: 23 September 2024

Published: 26 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Non-destructive methods of selecting viable seeds represent a promising alternative to traditional methods, such as seed selection, germination tests, and destructive analysis, used in breeding and cultivation processes, which are often time-consuming and labor-intensive. Advancements in spectrometric device capabilities and data processing techniques, including machine learning, have led to significant research interest in the development of non-destructive seed viability assessment methods using spectrometric analysis for applications in the agricultural sector [1,2].

Non-destructive analysis of seed viability is a primary area of research, because most conventional methods are destructive. A prominent example of conventional methods is the tetrazolium (TZ) test, in which seeds are treated with tetrazolium solution, which stains all live tissues red but leaves dead tissues unstained [3]. While the TZ test enables visual assessment of seed viability, it is a subjective and destructive method, requiring the seeds to be destroyed for analysis.

Multispectral imaging (MSI), a non-destructive technique, is being utilized to overcome the drawbacks of conventional methods. MSI combines computer vision and spectroscopy and provides data on the texture, color, shape, size, and chemical composition of seeds [4]. Not only does MSI minimize seed resource consumption and damage through multispectral data analysis, but it also enhances objectivity and reduces subjectivity during evaluation, thereby improving existing seed viability assessment methods.

Videometerlab 4 (Videometer A/S, Hørsholm, Denmark) is an MSI analyzer that precisely evaluates seed quality factors based on LED technology, enabling rapid, non-destructive analysis of large quantities of seeds. Videometerlab 4 combines LED lamps, a color camera, and optical filters [5], which enable seed-surface mold detection, seed identification, maturity determination, varietal differentiation, and pest detection with high accuracy [6].

Recent studies have successfully applied MSI and machine learning techniques to assess seed viability and quality in various crop species. Mihailova et al. [7] demonstrated the use of MSI for distinguishing between Arabica and Robusta coffee beans through discriminant and regression analysis. Zhang et al. [8] achieved 92.9% and 97.8% accuracy in classifying alfalfa seeds based on maturity and harvest year, respectively, by employing MSI and machine learning for morphological and multispectral data collection and multivariate analysis. Olesen et al. [9] determined the viability of *Ricinus communis* L. seeds with 92% accuracy through principal component analysis of multispectral data. These studies highlight the utility of MSI as a non-destructive means of seed classification and viability assessment, but they predominantly focus on crop seeds. However, research on applying MSI and machine learning techniques to wild plant seeds is limited. Wild plant species are crucial for biodiversity conservation and ecological balance, but their seed viability assessment often faces challenges due to limited seed availability and the destructive nature of conventional methods [10]. Applying MSI and machine learning to wild plant seeds could provide a non-destructive, efficient method for assessing seed viability, aiding in conservation efforts and habitat restoration projects.

MSI analyzers, such as Videometerlab 4, enable rapid and accurate analysis of large quantities of seeds, significantly reducing the time and labor involved. The current study advances the field using machine learning techniques for processing and analyzing multispectral data to evaluate seed viability and identify key wavelengths for classification.

While Videometerlab 4 can analyze seeds using 19 spectral wavelengths ranging from 365 nm to 970 nm, not all spectral wavelengths are essential for classification. Instead, the identification of wavelengths that provide significant information for characteristic-based classification of seeds is crucial [11]. This study uses machine learning algorithms to classify seeds as viable or non-viable based on spectral data and conducts feature importance analysis with 19 spectral wavelengths as variables to identify significant wavelengths for classification.

Four machine learning algorithms were employed in this study including Decision Tree, Random Forest, Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost). The Decision Tree algorithm learns explicit rules from the data to classify and predict targets [12]; Random Forest constructs multiple decision trees through data learning to classify targets [13]; XGBoost optimizes performance and prevents overfitting through parallel processing, tree pruning, and regularization [14]; and LightGBM provides accurate learning outcomes using techniques such as Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) [15]. Additionally, ensemble learning was applied to overcome the limitations of single machine learning models, thus enhancing classification accuracy.

Although machine learning and ensemble learning have been applied to MSI data analysis in previous studies, most have focused on crop seeds for maturity, harvest year, and viability classifications. Research on replacing the destructive TZ test with non-destructive seed viability classification using machine learning and ensemble learning for wild plant seeds is still limited.

Therefore, this study aims to (1) develop a classification model for non-destructive viability analysis of viable and non-viable seeds of the wild plant *Allium alleungense* using machine learning and ensemble learning and (2) identify critical wavelengths for classifying viable and non-viable seeds through feature importance analysis of spectral wavelengths used in machine learning classification. By focusing on wild plant species, this research ex-

tends MSI and machine learning applications beyond crop seeds, highlighting the potential for broader applications in biodiversity conservation and ecological studies.

2. Materials and Methods

2.1. Plant Material

This study used *Allium ulleungense* as the plant material for seed viability analysis. Seeds were collected on 31 August 2021 from the Nari Basin in Ulleung County, Gyeongsangbuk-do, Republic of Korea and stored at $-20\text{ }^{\circ}\text{C}$ under 40% relative humidity in the Seed Bank of Baekdudaegan National Arboretum. Multispectral and viability data were obtained from 130 seeds subjected to the saturated sodium chloride (NaCl)-accelerated aging test, with three repetitions per treatment. The experiment was conducted over 3 months, from 1 August to 13 October 2023.

2.2. Saturated Sodium Chloride (NaCl)-Accelerated Aging Test

To obtain multispectral and viability data from seeds in various viability states, an accelerated aging (AA) test was conducted. According to the method of Meriaux et al. [16], 500 mL of NaCl solution was prepared by dissolving 380 g of NaCl in 1 L of triple-distilled water and poured into the AA test box. Seeds were placed in glass Petri dishes (26 seeds per Petri dish), ensuring they did not touch the solution, and were exposed to $40\text{ }^{\circ}\text{C}$ and 75% relative humidity for different durations: 1, 2, 3, 5, and 7 days. Each exposure duration was treated as a separate experimental condition, with three repetitions conducted for each day. Thus, multispectral and viability data were collected from a total of 390 *A. ulleungense* seeds (26 seeds \times 5 days \times 3 repetitions).

2.3. MSI

Multispectral data were obtained from *A. ulleungense* seeds using VideometerLab 4 (Videometer A/S, Hørsholm, Denmark). Seeds were placed under the device, and color calibration and light diffusion were performed before capturing the multispectral images. Seeds were exposed to uniform LED light at 19 specific wavelengths (365, 405, 430, 450, 470, 490, 515, 540, 570, 590, 630, 645, 660, 690, 780, 850, 880, 940, and 970 nm) for 5–10 s, and high-resolution multispectral images (2054 \times 2054 pixels) were captured using a 5-megapixel CCD camera. The obtained multispectral images consisted of grayscale images captured at each of the 19 wavelengths and RGB images that combined pixel values of the red, green, and blue channels.

After acquiring multispectral images, seeds were separated from the background using the VideometerLab 3.18 software, and each seed was designated as a Region of Interest (ROI). The average spectral reflectance values of seeds within the ROI were calculated at each of the 19 wavelengths, and the data were saved in Microsoft Excel 2021 (Figure 1). Viable and non-viable seeds were scored as '1' and '0', respectively.

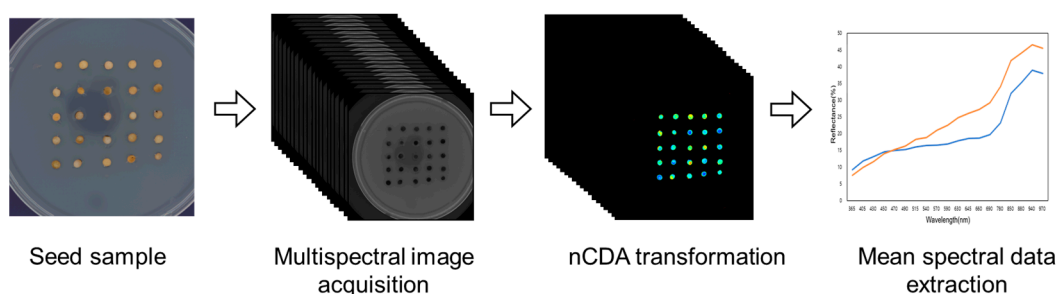


Figure 1. Acquisition process of multispectral imaging and spectral data.

2.4. Seed Viability Test

After obtaining the multispectral data, the TZ test was conducted to determine seed viability. Briefly, *A. ulleungense* seeds were soaked in distilled water for 18 h and then sliced

to expose the embryo. The sliced seeds were then soaked in TZ solution and incubated at 30 °C in the dark. Viability was determined based on the staining of the embryo, according to ISAT guidelines [3]. Digital images of viable and non-viable seeds were obtained using an optical microscope (DVM6, Leica MicroSystems GmbH, Wetzlar, Germany) (Figure 2).



Figure 2. Seed viability test pictures of *A. ulleungense* seeds. (A) A viable seed and (B) a non-viable seed.

2.5. Data Analysis

Based on the TZ test results, seeds were divided into two classes: viable and non-viable. To prevent class imbalance during machine learning training, SMOTE oversampling was applied. Of the collected data, 80% was used for training the machine learning models, and 20% was used for model evaluation. K-Fold cross-validation (Fold = 5) was applied to the training data to prevent overfitting. K-Fold cross-validation, a common validation method, divides the data into N folds and repeats the training and validation processes N times, averaging the results for evaluating model performance [17].

2.6. Machine Learning Models and Ensemble Learning

This study utilized classification-specific machine learning models, including Decision Tree, Random Forest, LightGBM, and XGBoost. The Decision Tree model generates decision rules based on data features, providing intuitive insights into data patterns [18]. Random Forest uses ensemble learning with multiple decision trees to reduce overfitting and improve prediction performance [19]. LightGBM offers fast learning speeds and high efficiency, making it suitable for large datasets [20]. XGBoost enhances learning speed and prediction accuracy through advanced regularization and parallel processing [14]. Each model was trained individually, and several stacking models were constructed using each base model (Decision Tree, Random Forest, LightGBM, XGBoost) as the meta-model. The predictions from the other base models were used as input features for these meta-models. This approach allowed for comparing the performance of different meta-models within the stacking framework. This two-stage process enabled the ensemble models to leverage the complementary strengths of the base models, leading to improved prediction accuracy and overall performance compared to any single model [21]. The training and evaluation of machine learning models were conducted in a virtual environment using Google Colaboratory (Colab).

2.7. Evaluation Metrics and Model Interpretation

Evaluation metrics for the classification models were determined based on the agreement between the actual results and predicted results using a new test dataset rather than the data used for model training. Five metrics, namely accuracy, precision, recall, specificity, and F1-score, were calculated using a confusion matrix comprising the numbers of True

Positive (TP) cases (i.e., when the model correctly predicted a sample as positive), False Positive (FP) cases (i.e., when the model incorrectly predicted a sample as positive), False Negative (FN) cases (i.e., when the model incorrectly predicted a sample as negative), and True Negative (TN) cases (i.e., when the model correctly predicted a sample as negative). The description of each performance metric and the formulae used to calculate these metrics are provided in Figure 3.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	TP	FN	Recall $\frac{TP}{(TP + FN)}$
	Negative	FP	TN	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Figure 3. Model evaluation values and confusion matrix.

2.8. Feature Importance Evaluation

To identify the wavelengths providing significant information for seed viability classification, feature importance analysis was conducted on the 19 spectral wavelengths (365–970 nm) used as variables in the machine learning classification models. Feature importance analysis was performed using the ‘feature_importance_’ attribute within each algorithm.

2.9. SHAP (Shapley Additive Explanations) Analysis

To gain deeper insights into how each wavelength influenced the model’s predictions, SHAP analysis was employed. SHAP values offer a consistent and interpretable means to evaluate the contribution of each feature to the classification outcomes. The SHAP evaluation was conducted using the SHAP library [22] in Python within Google Colaboratory (Colab), which facilitated efficient computation and visualization.

3. Results

3.1. Spectroscopic Analysis of *A. ulleungense* Seeds

Before conducting the TZ test, the average spectral reflectance values of 200 viable and 180 non-viable seeds were compared at 19 different wavelengths (365, 405, 430, 450, 470, 490, 515, 540, 570, 590, 630, 645, 660, 690, 780, 850, 880, and 970 nm) using multispectral images captured with VideometerLab 4. The average reflectance values of *A. ulleungense* seeds increased with increasing wavelengths (from 13.719 at 365 nm to 32.448 at 970 nm for viable seeds, and from 9.246 at 365 nm to 20.490 at 970 nm for non-viable seeds), although the average reflectance values of viable seeds were greater than those of non-viable seeds at the different wavelengths (Figure 4). An independent *t*-test showed that differences between the average reflectance values of viable and non-viable seeds were highly significant at all wavelengths ($p < 0.01$), with the differences becoming more pronounced at wavelengths above 850 nm.

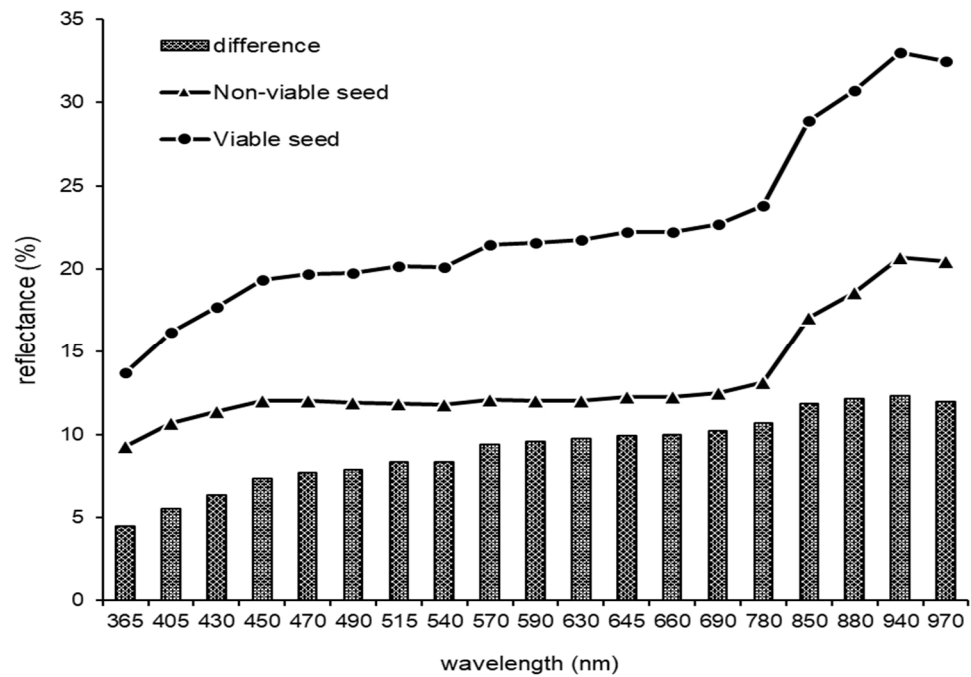


Figure 4. Mean spectral reflectance at 19 wavelengths based on reflection of viable and non-viable seeds of *A. ulleungense*.

Next, Linear Discriminant Analysis (LDA), a dimensionality reduction technique for supervised classification, was used to classify viable and non-viable seeds using the 19 spectral wavelengths as variables. Non-viable seeds were primarily located at negative values along the LDA1 axis, while viable seeds were concentrated at positive values, indicating that certain spectral wavelengths were closely associated with seed viability. The clear separation of viable and non-viable seeds along the LDA1 axis suggested that the two groups could be effectively distinguished based on spectral data. Except for some overlap near the center, most data points were well separated, demonstrating the utility of spectral data for classifying viable and non-viable seeds (Figure 5).

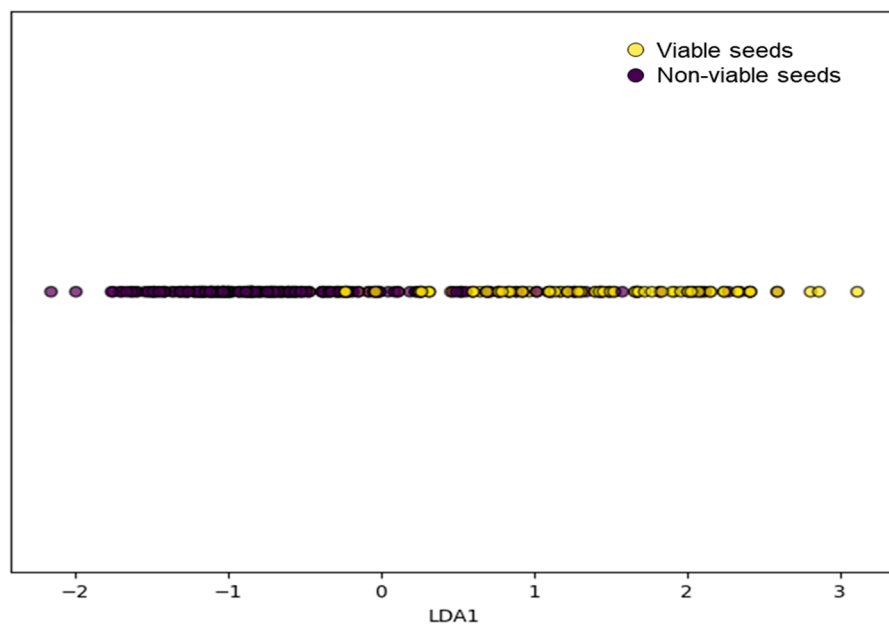


Figure 5. Linear Distribution Analysis (LDA) of viable and non-viable seeds of *A. ulleungense*.

The results of the independent *t*-test and LDA showed that different wavelengths hold different levels of significance for viability assessment. These findings suggest that spectral analysis can serve as a valuable non-destructive method for evaluating seed viability. Based on these results, a classification model was developed to automatically distinguish between viable and non-viable seeds using spectral data collected through machine learning.

3.2. Machine Learning Training and Evaluation

To classify viable and non-viable seeds, 80% of the data were used for model training and 20% for model evaluation. To prevent overfitting on the training data, K-Fold cross-validation ($k = 5$) was applied and the average evaluation results for each fold were obtained. All models were re-trained and evaluated using optimal hyperparameters identified through Grid Search (Table 1).

Table 1. Tuned hyperparameter results of machine learning models.

Model	Tuned Hyperparameter
Decision Tree	'criterion': 'gini', 'max_depth': 20, 'min_samples_split': 5
Random Forest	'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50
LightGBM	'boosting_type': 'dart', 'learning_rate': 0.5, 'num_leaves': 31
XGBoost	'subsample': 0.7, 'n_estimators': 400, 'max_depth': 8, 'learning_rate': 0.2, 'colsample_bytree': 0.9

Four machine learning models with optimal hyperparameters, including Decision Tree, Random Forest, LightGBM, and XGBoost, were comprehensively evaluated using the training dataset ($n = 320$), test dataset ($n = 80$), and K-Fold cross-validation. The evaluation metrics included accuracy, precision, recall, specificity, and F1-score.

The Random Forest model achieved the highest performance on the training dataset, with an accuracy of 99.7%, recall of 100%, precision of 99.4%, specificity of 93.7%, and F1-score of 99.7%. These results indicate that the Random Forest model could effectively distinguish between positive and negative classes, demonstrating high predictive power despite a tendency towards overfitting.

To further validate the models and prevent overfitting, K-Fold cross-validation ($k = 5$) was performed. The Random Forest model exhibited the most stable performance, with an average accuracy of 95.17%, followed by XGBoost (94.86%), Decision Tree (92.6%), and LightGBM (91.3%) (Table 2).

Table 2. Analysis of model classification performance on training and test data with K-Fold validation by calculation of accuracy, precision, recall, specificity, and F1-score based on spectral features.

	Prediction	Decision Tree	Random Forest	LightGBM	XGBoost
Training ($n = 320$)	Accuracy	0.959	0.997	0.959	0.950
	Precision	0.940	0.994	0.941	0.929
	Recall	0.981	1.000	0.981	0.975
	Specificity	0.937	0.994	0.937	0.925
	F1-score	0.936	0.997	0.961	0.952
Test ($n = 80$)	Accuracy	0.850	0.825	0.850	0.813
	Precision	0.846	0.791	0.805	0.811
	Recall	0.846	0.872	0.892	0.790
	Specificity	0.846	0.781	0.814	0.833
	F1-score	0.854	0.829	0.847	0.800
K-Fold ($n = 320$)	Accuracy	0.9260 ± 0.0363	0.9517 ± 0.0339	0.9131 ± 0.0364	0.9486 ± 0.0342
	Precision	0.8660 ± 0.0722	0.9123 ± 0.0490	0.8335 ± 0.0666	0.9126 ± 0.0537
	Recall	0.9389 ± 0.0467	0.9563 ± 0.0499	0.9355 ± 0.0618	0.9480 ± 0.0515
	Specificity	0.9207 ± 0.0415	0.9499 ± 0.0275	0.9013 ± 0.0252	0.9500 ± 0.0318
	F1-score	0.8993 ± 0.0483	0.9335 ± 0.0463	0.8804 ± 0.0580	0.9293 ± 0.0461

On the test dataset, both the LightGBM and Decision Tree models achieved the highest accuracy (85%). LightGBM demonstrated high performance in identifying positive samples, with a recall of 89.2%. All four models showed a slight decrease in performance on the test dataset compared with the training and validation datasets (Figure 6A,B). To better understand the variation in performance across different datasets, the classification ratios of positive and negative classes were analyzed using confusion matrices.

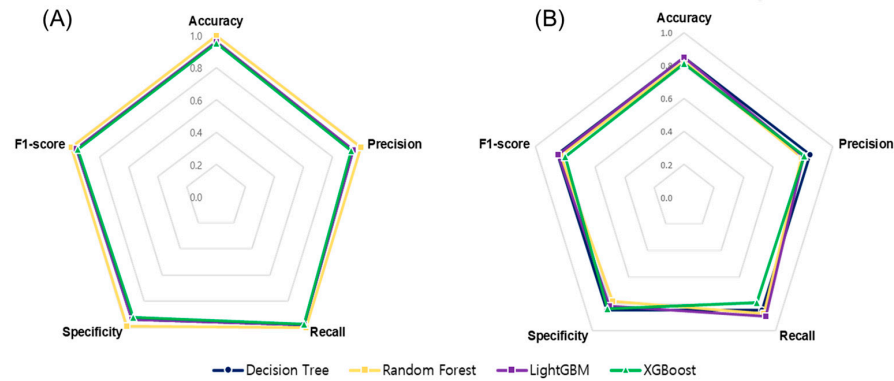


Figure 6. Star plot of model classification performances on test data with overall accuracy, precision, specificity, sensitivity, and F1-score based on spectral features. (A) Training data and (B) test data.

The confusion matrix for the Decision Tree model showed TP 35, FP 6, TN 33, and FN 6 (Figure 7A), indicating that this model consistently and evenly classified viable and non-viable seeds. The Random Forest model showed TP 32, FP 9, TN 34, and FN 5 (Figure 7B), demonstrating high classification accuracy for viable seeds and the best predictive power for positive samples among the four models. The LightGBM model showed TP 35, FP 8, TN 33, and FN 4 (Figure 7C), indicating strong predictive power for classifying non-viable seeds, similar to the Decision Tree model. The XGBoost model showed TP 35, FP 7, TN 30, and FN 8 (Figure 7D), demonstrating high efficacy for classifying non-viable seeds but the lowest performance for classifying viable seeds among the four models.

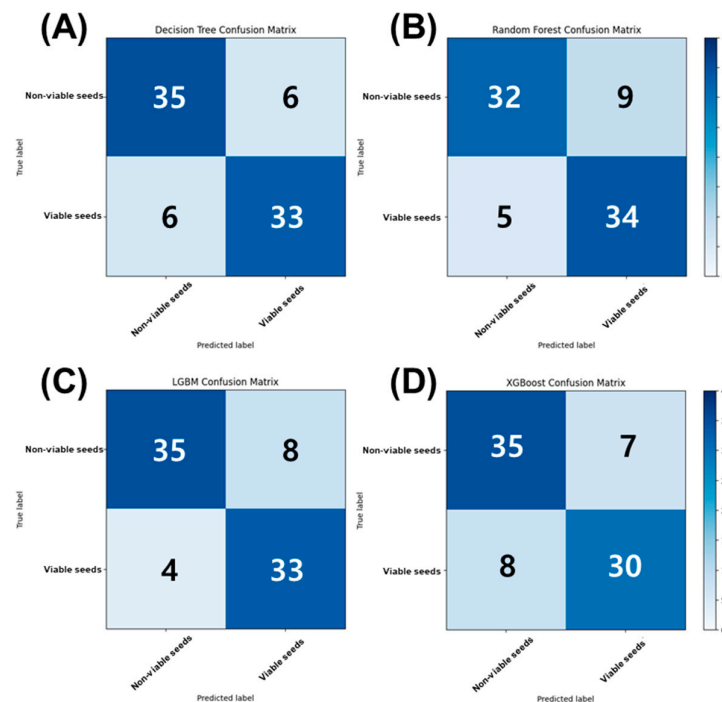


Figure 7. Confusion matrix and metrics for classification of viable and non-viable seeds of *A. ulleungense*. (A) Decision Tree, (B) Random Forest, (C) LightGBM, and (D) XGBoost.

Overall, single models based on spectral wavelengths showed higher performance in classifying non-viable seeds than viable seeds. The K-Fold cross-validation ($k = 5$) used to prevent overfitting on the training data showed high accuracy for all four models (91.3–95.1%), but accuracy decreased to 81–85% on the test data. This approximately 10% decrease in accuracy indicates that the models were overfitting to the training data and had reduced generalization ability for new data like the test dataset. To address this overfitting issue, “Stacking”, an ensemble technique that combines the predictions of individual models to generate a final prediction, was applied.

3.3. Stacking Ensemble Learning Training and Evaluation

The stacking model is an ensemble method, similar to voting, bagging, and boosting, that improves predictive performance using the predictions from several base models as inputs for a meta-model, which makes the final prediction. By combining the strengths of various models, the stacking model achieves higher accuracy than a single model and effectively reduces overfitting to specific training data. The four previously trained models, namely Decision Tree, Random Forest, LightGBM, and XGBoost, were integrated into a stacking model, and their performances were compared by selecting each one as the meta-model.

The stacking model showed the highest performance when the Decision Tree was chosen as the meta-model. Therefore, the stacking model was built with the Decision Tree as the meta-model and Random Forest, LightGBM, and XGBoost as the base models, and its performance was then evaluated on the test data. The stacking model with the Decision Tree as the meta-model achieved an accuracy of 0.9, precision of 0.9268, recall of 0.8837, specificity of 0.927, and F1-score of 0.9045 on the test data, demonstrating high performance (Figure 8A). Thus, the values of evaluation metrics obtained using the stacking model were higher than those obtained using a single model, indicating that the stacking model effectively prevents overfitting and shows better performance than a single machine learning model in classifying viable and non-viable seeds based on spectral data.

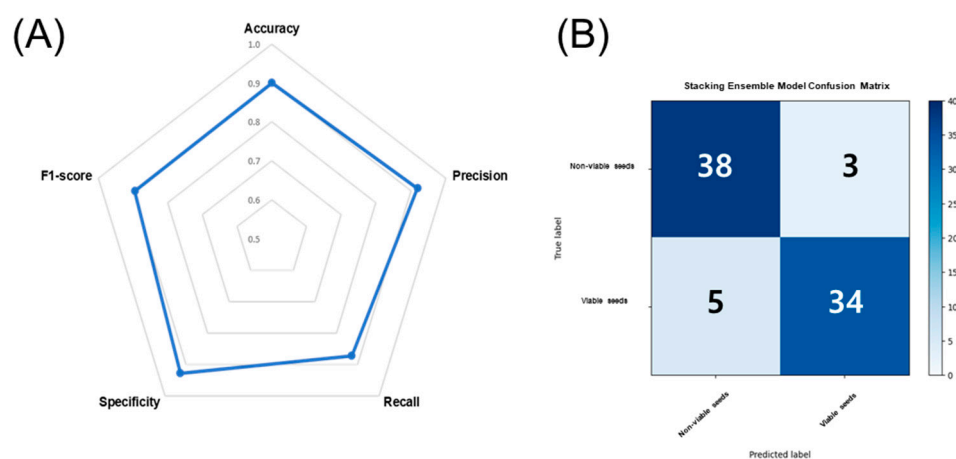


Figure 8. Star plot (A) and confusion matrix (B) of the SEL algorithm using Decision Tree as a meta-model.

The confusion matrix for the stacking model using test data showed TP, FP, TN, and FN values of 38, 3, 34, and 5, respectively, indicating that the stacking model improved prediction and classification abilities for non-viable seeds and maintained high performance for viable seeds compared with single models (Figure 8B).

Additionally, the ROC curve and Area Under the Curve (AUC) were evaluated as metrics to assess the predictive performance of the model across different thresholds. The ROC curve is a graphical representation of the relationship between the true positive rate (TPR, along the Y-axis) and false positive rate (FPR, along the X-axis); the closer the ROC

curve is to the top left corner, the better the performance, while a curve close to the diagonal indicates poor classification performance. The AUC summarizes the performance of a model as a value between 0.5 and 1; a value closer to 1 indicates better performance, while a value near 0.5 indicates poor and inconsistent classification performance.

Analysis of the ROC curve and AUC revealed that the stacking model with the Decision Tree as the meta-model achieved high recall and low FPR, effectively distinguishing between positive and negative samples with minimal misclassification. The proximity of the ROC curve to the top left corner indicated excellent model performance, and the AUC value of 0.93 suggested consistently high performance across various threshold settings (Figure 9). Together, these two metrics indicated that the stacking model provides reliable predictions.

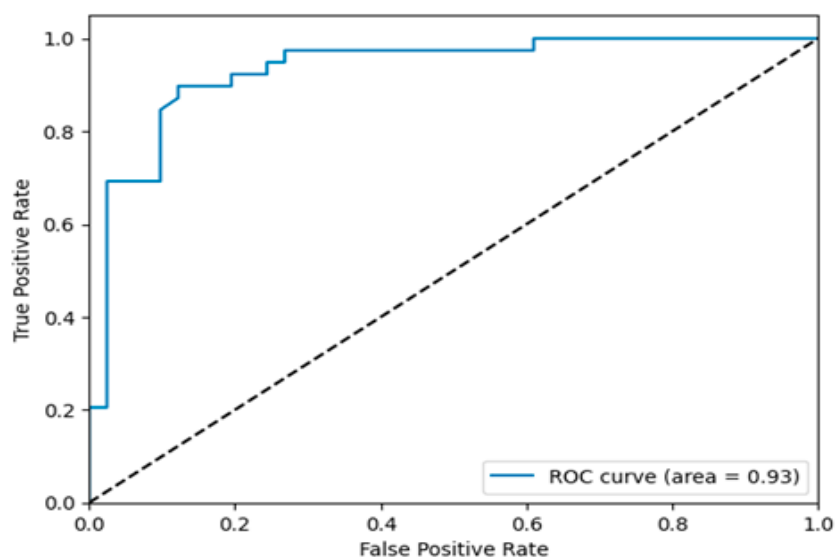


Figure 9. ROC curve and AUC of the SEL algorithm using Decision Tree as a meta-model.

3.4. Machine Learning Feature Importance Analysis

A feature importance evaluation was conducted to understand the contribution of 19 spectral wavelengths to the predictive power of the four machine learning models for classifying viable and non-viable seeds. This evaluation identifies the variables that have the most significant impact on model predictions, enhancing the understanding of how the models operate. Analysis of the importance of features allows the determination of which predictors the models rely on during training. Additionally, understanding each feature's relative importance helps improve model performance by reducing the risk of overfitting via removing less important features.

In this analysis, feature importance was evaluated using all four models, i.e., Decision Tree, Random Forest, LightGBM, and XGBoost, to gain crucial insights into how feature importance varies across different algorithms and to identify whether certain features consistently appear important across multiple models or if their importance is specific to particular models. Feature importance was measured through each algorithm's 'feature_importance_' attribute, showing the relative weights assigned to each feature and their influence on model predictions. The importance of each feature was depicted using bar graphs, where longer bars indicated higher importance, while shorter bars indicated lower importance.

In the Decision Tree model, the 970 nm wavelength showed the highest importance, with a value above 0.10, indicating significant influence on the decision-making process of the model. The next most important features were 880 nm and 940 nm, with importance values of approximately 0.08 and 0.07, respectively, while the 430 nm wavelength showed the lowest importance (Figure 10A).

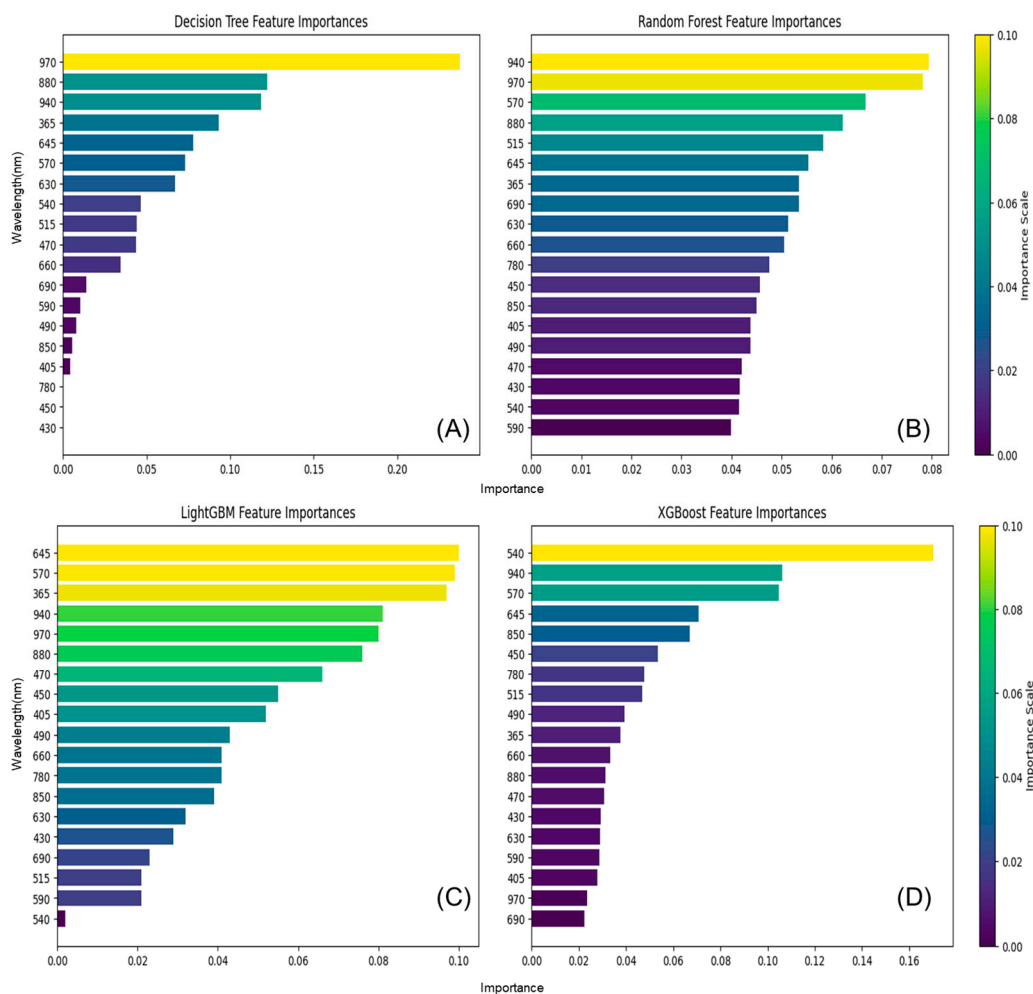


Figure 10. Importance scores of spectral features for (A) Decision Tree, (B) Random Forest, (C) LightGBM, and (D) XGBoost.

In the Random Forest model, the 940 nm wavelength had the highest importance (approximately 0.08), followed by 970 nm at 0.075 and 570 nm (0.067), with the 590 nm wavelength being the least important. The Random Forest model also relied heavily on the top features for predictions, but the importance of the less critical features was more evenly distributed in the Random Forest model than in the Decision Tree model (Figure 10B).

In the LightGBM model, the 645 nm wavelength showed the highest importance (0.09), indicating significant impact on model predictions. The importance values of 570 and 365 nm wavelengths (0.06 and 0.055, respectively) followed those of the 645 nm wavelength, while the 540 nm wavelength showed the lowest importance. These results suggest that LightGBM heavily depends on the top features (Figure 10C).

Lastly, in the XGBoost model, the 540 nm wavelength showed the highest importance (0.16), followed by 940 nm and 570 nm (0.13 and 0.11, respectively), with the 690 nm wavelength being the least important. Similar to the Decision Tree model, XGBoost assigned significant weight to certain features (Figure 10D).

Thus, feature importance analysis across all four primary machine learning models allowed us to understand how each model responds to different features in the data. Common features with the top 50% importance across all models were the 570, 645, and 940 nm wavelengths. Although some features showed varying importance across different models, this variation was likely due to differences in their learning algorithms and decision mechanisms.

Together, the above results indicate that comparison of the impact of specific features on model predictions can enhance model performance by highlighting key features and removing unnecessary ones, leading to more efficient models and improving their interpretability and transparency. However, feature importance analysis alone could not clearly identify universally important variables across all models. Therefore, to gain a deeper understanding of the model’s decision-making process, SHAP analysis was conducted.

3.5. Machine Learning SHAP Analysis

Through SHAP analysis, the mean absolute SHAP values were calculated, and a bar graph was constructed to indicate the average impact of each feature (wavelength) on the model’s predictions, with the length of the bars representing the magnitude of the SHAP values. This approach allowed for a quantitative assessment of each feature’s influence on the model’s predictive power.

In the Decision Tree model, the 970 nm wavelength had a SHAP value of 0.223217, showing a dominant influence over other wavelengths. The 365, 570, and 645 nm wavelengths were among the top 50% of important features (Figure 11A).

In the Random Forest model, the 645 nm wavelength was the most important (SHAP value = 0.0416), followed by the 570 nm wavelength (SHAP value = 0.0327). Other relatively important wavelengths included 365, 880, and 940 nm (Figure 11B).

In the LightGBM model, the 645 nm wavelength was identified as the most important (SHAP value = 1.1090). The 570 nm wavelength also showed high importance (SHAP value = 0.8600). The 365 and 880 nm wavelengths were among the top 50% of important features (Figure 11C).

In the XGBoost model, the 645 and 940 nm wavelengths showed the highest SHAP values of 1.8190 and 1.5268, respectively. Other important features included the wavelengths of 365, 570, and 880 nm (Figure 11D).

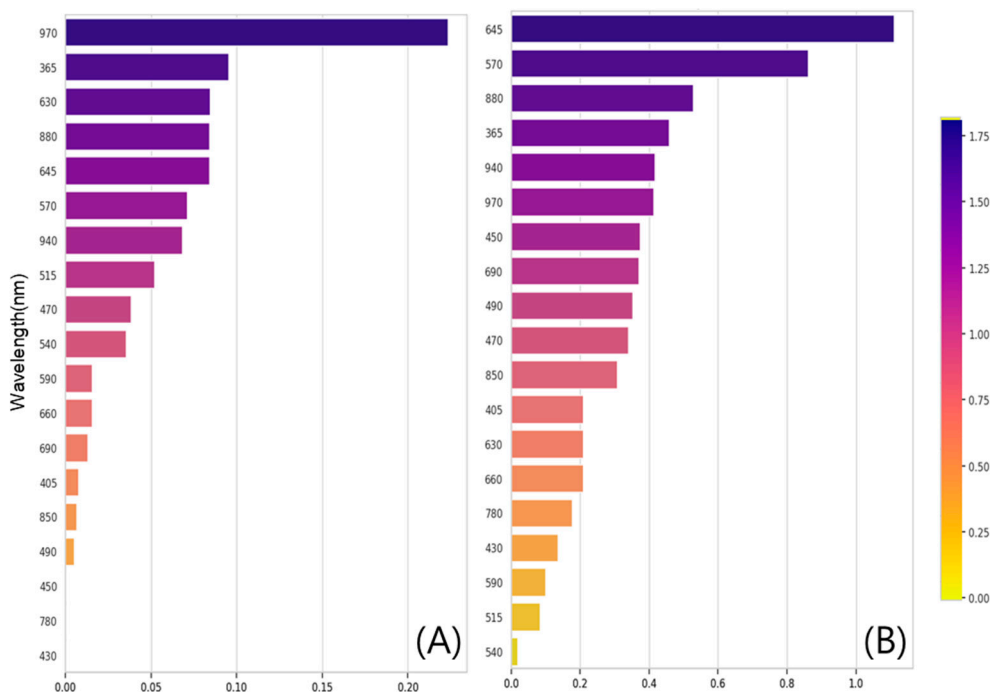


Figure 11. Cont.

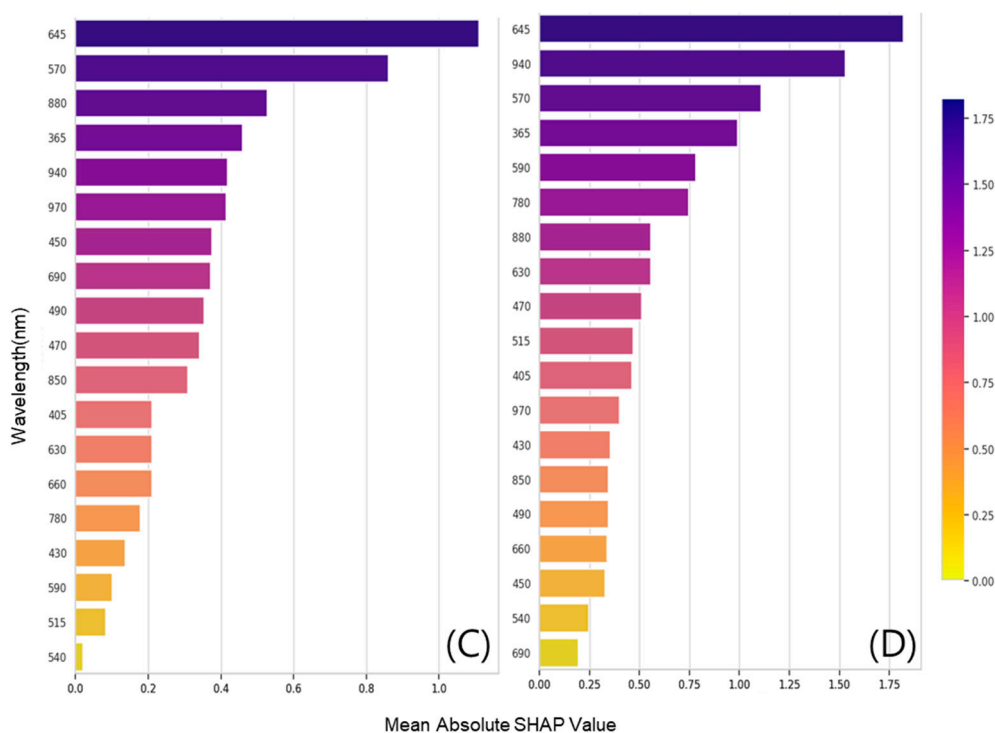


Figure 11. SHAP value of spectral features for (A) Decision Tree, (B) Random Forest, (C) LightGBM, and (D) XGBoost.

Overall, SHAP analysis of the four primary machine learning models allowed us to understand how each model responds to different features in the data. Common features with the top 50% importance across all models were the 365, 570, 645, 880, and 940 nm wavelengths. Feature importance analysis also highlighted 570, 645, and 940 nm as important wavelengths across all models. Combining these results, it is evident that 570, 645, and 940 nm wavelengths consistently showed high importance across all models.

3.6. Establishing Classification Criteria Based on Key Wavelengths Using the ROC Curve

Using the ROC curve to evaluate the performance of the ensemble model, the optimal threshold c (where the FPR and TPR were maximized) was determined to be 0.823. Thus, the ensemble model classified seeds as viable if the predicted probability was >0.823 and non-viable if the predicted probability was <0.823 (Figure 9).

For data points above the optimal threshold, the average spectral values of the 570, 645, and 940 nm wavelengths were analyzed because these were identified as highly important through feature importance and SHAP analyses. The average spectral values of 570, 645, and 940 nm wavelengths were 17.779, 18.327, and 28.305, respectively. Using these spectral values as thresholds and classifying seeds with values above the threshold as viable and those below the threshold as non-viable, an overall classification accuracy of 0.85 was achieved. The model demonstrated an accuracy of 85%, indicating that the spectral values at 570, 645, and 940 nm are reliable indicators for seed viability classification.

4. Discussion

The current study developed a classification model using machine learning to differentiate between viable and non-viable *A. ulleungense* seeds based on spectral data. The most influential spectral wavelengths contributing to model performance were identified through feature importance and SHAP analyses. While individual machine learning models achieved an accuracy of 95–99% on the training data, their accuracy on the test data was lower (81–85%). Additionally, all models showed relatively lower performance in classifying viable seeds than non-viable seeds (Figure 7).

A stacking ensemble model was developed to enhance model performance and tackle overfitting in the classification of viable seeds, with a Decision Tree as a meta-model. This stacking model displayed improved performance on the training data, achieving an accuracy of 0.9, precision of 0.9268, recall of 0.8837, specificity of 0.927, and an F1-score of 0.9045. Compared with individual models, the stacking model also reduced classification errors for viable and non-viable seeds (Figure 8).

These results suggest that the machine learning-based stacking model can serve as an effective tool for seed viability classification, with an accuracy of approximately 90%. This performance is similar to the 92% accuracy in seed viability classification achieved by Olesen et al. [7] using principal component analysis of spectral wavelengths. Furthermore, the optimal threshold analysis of the ensemble model confirmed that the critical spectral wavelengths identified by feature importance and SHAP analyses, namely 570 nm, 645 nm, and 940 nm, are effective for seed viability classification.

Among the three critical wavelengths, the 570 nm wavelength falls within the green spectrum and is known to influence the inhibition of seed dormancy release [23]. The 645 nm wavelength falls within the red spectrum, critical for promoting seed germination and breaking dormancy [24,25]. The 940 nm wavelength falls within the near-infrared (NIR) spectrum, which may not directly impact seed germination and dormancy but is widely used for seed quality evaluation [26] and varietal classification [27]. Further detailed analysis of NIR wavelengths is required to elucidate their roles in plants.

5. Conclusions

This study demonstrates that a stacking model using spectral data can effectively and non-destructively distinguish between viable and non-viable seeds, with an accuracy of approximately 90%. Identifying critical wavelengths (570, 645, and 940 nm) provides a reliable basis for assessing seed viability. The results of this study align with previous research, confirming the potential of machine learning-based approaches in this field.

Unlike previous studies primarily focused on crop seeds, our research indicates that spectral-based non-destructive seed viability analysis models apply to wild plant seeds. Based on these findings, we plan to expand the model's applicability by including three additional wild plant species within the *Allium* genus, thereby increasing data volume and enhancing model robustness. This expansion aims to improve the model's performance and validate its utility across wild plant seeds.

In conclusion, while this study presents a promising approach for non-destructive seed viability assessment, further validation of the model across different species and larger datasets will be critical to ensure its broader applicability. Future research will incorporate autofluorescence spectral images collected with VideometerLab 4 to improve classification accuracy further and expand the model's capabilities. This approach will increase the number of wavelengths from 19 to 50, allowing a more detailed exploration of spectral data. Addressing the uncertainties observed in the model's performance, particularly regarding the classification of viable seeds, will be a crucial focus of future work.

Author Contributions: Conceptualization, Y.R.C. and C.S.N.; methodology, Y.R.C., M.S.J. and C.S.N.; investigation: Y.R.C., M.S.J., G.E.K., C.H.P., D.J.L. and S.H.C.; formal analysis, Y.R.C.; writing—original draft preparation, C.S.N. and S.H.C.; writing—review and editing, Y.R.C. and M.S.J.; visualization, C.S.N.; supervision, C.S.N.; project administration, C.S.N.; funding acquisition, C.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out with the support of the R&D Program for Forest Science Technology (Project No. 2021400B10-2425-CA02) provided by the Korea Forest Service (Korea Forestry Promotion Institute).

Data Availability Statement: The dataset is available on request from the corresponding author. We are currently extending this method to the *Allium* genus, and the data presented in this paper will serve as foundational data for this ongoing research; therefore, it is difficult to make them publicly available at this time.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, T.; Wei, W.; Zhao, B.; Wang, R.; Li, M.; Yang, L.; Wang, J.; Sun, Q. A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. *Sensors* **2018**, *18*, 813. [[CrossRef](#)] [[PubMed](#)]
2. Baek, I.; Kusumaningrum, D.; Kandpal, L.M.; Lohumi, S.; Mo, C.; Kim, M.S.; Cho, B.K. Rapid measurement of soybean seed viability using kernel-based multispectral image analysis. *Sensors* **2019**, *19*, 271. [[CrossRef](#)]
3. Leist, N.; Krämer, S.; Jonitz, A. *ISTA Working Sheets on Tetrazolium Testing Volumes I and II*; International Seed Testing Association: Wallisellen, Switzerland, 2003.
4. ElMasry, G.; Mandour, N.; Al-Rejaie, S.; Belin, E.; Rousseau, D. Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview. *Sensors* **2019**, *19*, 1090. [[CrossRef](#)] [[PubMed](#)]
5. ElMasry, G.; Mandour, N.; Ejeez, Y.; Demilly, D.; Al-Rejaie, S.; Verdier, J.; Rousseau, D. Multichannel imaging for monitoring chemical composition and germination capacity of cowpea (*Vigna unguiculata*) seeds during development and maturation. *Crop J.* **2022**, *10*, 1399–1411. [[CrossRef](#)]
6. Boelt, B.; Shrestha, S.; Salimi, Z.; Jørgensen, J.R.; Nicolaisen, M.; Carstensen, J.M. Multispectral imaging—a new tool in seed quality assessment? *Seed Sci. Res.* **2018**, *28*, 222–228. [[CrossRef](#)]
7. Mihailova, A.; Liebisch, B.; Islam, M.D.; Carstensen, J.M.; Cannavan, A.; Kelly, S.D. The use of multispectral imaging for the discrimination of Arabica and Robusta coffee beans. *Food Chem. X* **2022**, *14*, 100325. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, S.; Zeng, H.; Ji, W.; Yi, K.; Yang, S.; Mao, P.; Li, M. Non-destructive testing of alfalfa seed vigor based on multispectral imaging technology. *Sensors* **2022**, *22*, 2760. [[CrossRef](#)] [[PubMed](#)]
9. Olesen, M.H.; Nikneshan, P.; Shrestha, S.; Tadayyon, A.; Deleuran, L.C.; Boelt, B.; Gislum, R. Viability prediction of *Ricinus communis* L. seeds using multispectral imaging. *Sensors* **2015**, *15*, 4592–4604. [[CrossRef](#)] [[PubMed](#)]
10. Merritt, D.J.; Dixon, K.W. Restoration seed banks—A matter of scale. *Science* **2011**, *332*, 424–425. [[CrossRef](#)] [[PubMed](#)]
11. Du, H.; Qi, H.; Wang, X.; Ramanath, R.; Snyder, W.E. Band selection using independent component analysis for hyperspectral image processing. In Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, 15–17 October 2003; pp. 93–98.
12. Goel, P.K.; Prasher, S.O.; Patel, R.M.; Landry, J.A.; Bonnell, R.B.; Viau, A.A. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Comput. Electron. Agric.* **2003**, *39*, 67–93. [[CrossRef](#)]
13. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–15 August 1995; Volume 1, pp. 278–282.
14. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
15. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 52.
16. Meriaux, B.; Wagner, M.H.; Ducournau, S.; Ladonne, F.; Fougereux, J.A. Using sodium chloride saturated solution to standardize accelerated aging test for wheat seeds. *Seed Sci. Technol.* **2007**, *35*, 722–732. [[CrossRef](#)]
17. Stone, M. Cross-validated choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–133. [[CrossRef](#)]
18. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
20. Thai, H.T. Machine learning for structural engineering: A state-of-the-art review. In *Structures*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 448–491.
21. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms*; CRC Press: Boca Raton, FL, USA, 2012.
22. Lundberg, S. A unified approach to interpreting model predictions. *arXiv* **2017**, arXiv:1705.07874.
23. Goggin, D.E.; Steadman, K.J. Blue and green are frequently seen: Responses of seeds to short- and mid-wavelength light. *Seed Sci. Res.* **2012**, *22*, 27–35. [[CrossRef](#)]
24. Bewley, J.D.; Black, M.; Negbi, M. Immediate action of phytochrome in light-stimulated lettuce seeds. *Nature* **1967**, *215*, 648–649. [[CrossRef](#)]
25. Mohammed, Q.; Tillberg, E. Rapid effects of red light on the isopentenyladenosine content in Scots pine seeds. *Plant Physiol.* **1989**, *91*, 5–8.
26. Reddy, P.; Guthridge, K.M.; Panozzo, J.; Ludlow, E.; Spangenberg, G.; Rochfort, S. Near-Infrared Hyperspectral Imaging pipelines for pasture seed quality evaluation: An overview. *Sensors* **2022**, *22*, 1981. [[CrossRef](#)] [[PubMed](#)]
27. Sharma, A.; Singh, T.; Garg, N. Combining near-infrared hyperspectral imaging and ANN for varietal classification of wheat seeds. In Proceedings of the 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies, Kannur, India, 11–12 August 2022; pp. 1103–1108.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.