



## **Prediction of Seed Germination Quality Utilizing Ensemble-Based Precision Forming**

**Mukkoti Marathi Venkata Chalapathi<sup>1</sup>    A Vijaya Krishna<sup>2</sup>    G Gopichand<sup>3\*</sup>  
Kadiyala Ramana<sup>4</sup>    V C Bharathi<sup>1</sup>    Neeraj Sharma<sup>5</sup>**

<sup>1</sup>*School of Computer Science and Engineering, VIT-AP University, Amaravati, India*

<sup>2</sup>*Department of Computer Science and Technology, GNITS, Hyderabad, India*

<sup>3</sup>*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India*

<sup>4</sup>*Department of Artificial Intelligence & Data Science, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India*

<sup>5</sup>*Department of Information Technology, VPPCOE & VA, Mumbai, India*

\* Corresponding author's Email: [gopichand.g@vit.ac.in](mailto:gopichand.g@vit.ac.in)

---

**Abstract:** Seed germination is a primary objective of precision agriculture. Precision agriculture, which makes extensive use of machine learning, has been the subject of recent studies on predictive analytics. These machine learning methods typically employ supervised learning models to make predictions about how successfully seeds will germinate. However, a major challenge that modern models face when attempting to make accurate predictions is the curse of dimensionality in the training corpus. The primary contribution of this manuscript is an ensemble-based method for predicting seed germination quality (EL-GQP) in precision agriculture. The accuracy of predictions can be improved using this ensemble method, which combines the positive aspects of a number of different models while minimising the negative aspects of the individual models. The proposed model is significantly superior to the current model, as demonstrated by experimental results of cross-validation on the benchmark dataset. During the simulation, work is done on the corpus dataset contains 4250 negative records and 6230 positive records.

**Keywords:** Precision farming, Ensemble classification, Germination quality, Machine learning, Predictive analytics.

---

### **1. Introduction**

[1] Asserts that seed quality is an essential component of agricultural production that directly influences yield. Utilizing high-quality seeds and inbreeding the plant lowers the cost of field experiments while increasing the likelihood of finding the best crop variety. As shown in [2] programmes for ensuring the quality of seeds use a range of methods to validate the traits of seed-like vigour and germination tests. These processes have limitations associated with the consumption of time, destructive nature, and subjectivity of measuring seed quality [3-5]. The work [6] presents that there has been a rapidly increasing demand for effective approaches, which might provide reliable, rapid, objective, and non-destructive identification of the

quality of seed [6].

To a human society, seeds are required as the main source of food and service to be significant crops materials. The yield of a crop is heavily influenced by seed quality and environmental factors. According to [7], measuring seed germination is a necessary task for seed researchers in order to evaluate the various seed lots and enhance the efficiency of the food chain. It is important to note that worldwide crop product needs to be doubled to supply the increasing population by 2050 as in [8]. Traditional measures of seed testing, mainly tests of seed-vigor, have not been utilized extensively because of time-intensive and cumbersome protocols as in [9]. Moreover, many seed tests introduced by ISTA (international seed testing association) have been manually assessed by utilizing a standardized

process, which varies for diversified crops [10]. Concerning to the prologue abovementioned, it is necessary to contribute computer aided models those build on machine learning, artificial intelligence, and deep h platforms. In this context, a novel machine learning model, which is an ensemble learning-based seed germination quality check model has portrayed in this manuscript.

EL-GQP, the proposed precision agriculture method, has several advantages over current methods. First, it uses ensemble-based modelling to improve prediction accuracy. This method reduces model drawbacks, improving predictions. Second, EL-GQP addresses the curse of dimensionality, which plagues machine learning models. Predictions are difficult when the training dataset has many variables. EL-GQP solves this problem by using an ensemble of models to handle more variables and produce more accurate results.

Third, EL-GQP accurately predicts seed germination. Precision agriculture aims to achieve this, and the suggested approach outperforms existing models. Cross-validation on the benchmark dataset shows that EL-GQP performs significantly better than existing models.

The EL-GQP method has several advantages over current precision agriculture methods. Precision agriculture practitioners use it to integrate multiple models, overcome dimensionality, and accurately predict seed germination quality.

This article has been organized into five parts. The first section discusses seed germination and how machine learning (ML) algorithms can be used to predict seed germination quality. The second section contains a comprehensive review of the most recent ML algorithms for predicting seed germination quality. The third section discusses the study's methods and materials, with a focus on an ensemble learning-based seed germination quality check model. Section four presents the results of an experimental study on seed germination quality prediction using the proposed model, as well as performance measures. Finally, the conclusion highlights the study's contribution and summarises the key findings.

## 2. Related work

Generally, variations in internal anatomical characteristics and chemical composition of seeds have been associated with loss of vigor and viability, as in [11]. However, these variations have been unlikely recognized through visual inspection. Moreover, approaches dependent on X-ray imaging and spectrometric strategies have been utilized successfully to gather data on complex traits

associated with seed quality. In this case, FT-NIR (Fourier transform near-infrared) spectroscopy has demonstrated significant potential for identifying seed compounds by acquiring a large number of spectral details, as shown in [12-18]. FT-NIR spectroscopy can acquire a large amount of spectral data and is effective in identifying seed compounds, according to several studies. As shown in [19], the FT-NIR spectroscopy is dependent on electromagnetic radiation absorption at wavelengths ranging from 780 to 2500 nm. This wide range of wavelengths enables direct and simultaneous measurement of multiple constituents in seed samples. As a result, it is versatile for simultaneous and direct measurements of various constituents in seed samples, as demonstrated in [16, 20-23]. Previous research has found that these characteristics make it an effective method for analysing seed samples. On other dimensions, X-ray imaging has been dependent on X-ray attenuation variances in diversified tissues types [24]. Therefore, it might reveal a physical seed state with an internal morphology as in [11]. Even though these strategies have the maximal possibility for classification of seed quality, integrating the datasets might produce novel information regarding the samples of seed or enhance the performance of the classifier as in [25].

Current advancements of ML algorithms have revolutionized agriculture due to their fundamental for building approaches to categorize products, mainly seeds quality attributes. The robust algorithms might capture non-linear and linear associations, and they might attain maximal accuracy of classification. Various algorithms proved as effective to solve the issues in several researching domains like PLS-DA (partial least squares discriminant analysis), LDA (linear discriminant analysis), NB (naïve Bayes), SVM, and many more, as stated in [12, 15, 17, 25, 26]. Nevertheless, distinct algorithms perform differently, and they might have diversified performances as in [26].

Even though models based on optical might produce accurate data on the quality of seed, combining datasets by ML algorithms might enhance the further performance of classification. There were no endeavors in utilizing FT-NIR integrated with X-ray data images for categorizing the quality of seed. Hence, by utilizing U. Brizantha grass seeds as an approach, we have tested whether combined data from the X-ray imaging and FT-NIR with ML algorithms might enhance the vigor and germination of seed predictions.

In order to cut down on the number of manual steps in the error-prone seed-testing process, several researchers have proposed models for automating

this process. Because they are easy to automate and provide more accurate measurements with fewer errors than alternative methods, traditional image analysis strategies are currently used to identify seeds. In contrast, many published algorithms predict the perimeter, roundness, colour, width, and perimeter values of the seed using colour-based thresholds. [32] Defines a "germinator" as a piece of software that, by calculating the variance and area between individual points in an image over time, predicts whether an *Arabidopsis-thaliana* seed will germinate. Different seeds require adjustments to different system components, and the system has probably failed in a variety of partial occlusion or illumination scenarios. Similar to [33], SVIS (Seed vigour imaging system) uses a flat-bed scanner to scan images and digitally process the RGB pixel values to determine the lengths of the seeds. The camera settings have been standardised to enhance performance on other dimensions when scanning in any lighting. In order to measure the seeds using this method, the researcher must be present during the entire germination simulation. KNN, SVM, NB, and ANN were previously compared to determine which ANN approaches had the highest accuracy and performance for recommended seed germination, as stated in [34]. As a result, the researchers manually performed an additional germination test using image processing to extract eleven features. Convolutional neural networks (CNN) are a relatively new technique for image processing, in contrast to deep learning [35]. CNNs use features that are automatically extracted and learned from the original images to solve a variety of image classification issues. One of the factors for success is a low reliance on clearly defined obstructions and illuminations to achieve maximum accuracy in computer vision tasks. Rice seed germination has already been automatically measured using CNNs [36]. This model could only forecast the final percentage of germination because images weren't taken until after the simulation of germination was complete.

Machine learning-based germination quality prediction for seeds (GDPQA) [37, 40] is another contemporary model that endeavored to speed up the assessment of seed germination quality prediction. The other contemporary model, -seed quality classification using merger data from spectroscopy and x-ray imaging (ML-SQC) [38, 41] has been aimed to avoid preprocessing of the input data to achieve rapid classification. However, these contemporary models do not address the curse of dimensionality that is considering lowering the prediction accuracy, specificity, and sensitivity.

The purpose of this contribution is to lessen the

consumption of time and human intervention for experiments on seed germination and to enhance the prediction model of germination, which has been (a) independent thresholds based on custom color and hence could be implemented to various illumination settings and seed cultivars and (b) could be utilized for exploring the vigorous germination of seed by predicting not only the ultimate percentage of germination however different indices such as uniformity and rate [44, 45].

According to the existing literature, variations in the internal anatomical characteristics and chemical composition of seeds have been linked to a loss of vigour and viability. X-ray imaging and spectrometric methods have been used successfully to collect data on complex traits related to seed quality. Machine learning algorithms have been successfully used to classify seed quality attributes. However, no attempt has been made to use FT-NIR combined with X-ray data images to categorise seed quality, and the curse of dimensionality limits prediction accuracy, specificity, and sensitivity. The goal of this research is to create an accurate, efficient, and robust model that can predict seed germination and investigate the vigorous germination of seeds using machine learning algorithms and data from FT-NIR and X-ray imaging. The goal is to reduce the time and human intervention required for seed germination experiments, as well as to improve the germination prediction model, which is independent of thresholds based on custom colour and can be applied to various lighting settings and seed cultivars.

### 3. Methods and material

The characteristics of the dataset and their importance for the simulation study and systematic model of the suggested EL-GQP approach are examined in the subsections of this section (ensemble learning based seed germination quality prediction).

#### 3.1 Quantitative phase imaging principle

The idea of quantitative stage imaging has been explained in this section. Seeds incident under a microscope can pass through a particular area and scatter at a particular sample area using light in the 400–1100 nm range, framing a structured sample background and indicating a quantitative image. Additionally, over the course of 48 hours, the light incident procedure on target would be repeated at various seed intervals in order to collect the quantitative images phase of the necessary sample.

Quantitative image phase processing can also be described mathematically as follows:

The intensity of light perceived at the light

incident event on the required sample could be measured under the below inputs, modulation of phase, coordinates of the incident  $[x, y]$ , and phase variance among pass through light and scattered light in the following Eq. 1.

$$I_{\phi}^{[x,y]} = (L_p)^2 + (L_s^{[x,y]})^2 + 2(L_p) \times (L_s^{[x,y]}) \times \cos(\Delta\phi^{[x,y]} + \phi) \quad (1)$$

The variance of quantitative phase among pass through light and scattered light has been measured by utilizing below Eq. 2.

$$\Delta\phi^{[x,y]} = \arctan\left(\frac{I_{\frac{3\pi}{2}}^{[x,y]} - I_{\frac{\pi}{2}}^{[x,y]}}{I_0^{[x,y]} - I_{\pi}^{[x,y]}}\right) \quad (2)$$

Microscopic images of these formulations at various quantitative stages might attain.

$$0, \frac{\pi}{2}, 3\frac{\pi}{4}, \pi \quad (3)$$

There would be three quantitative phase images for each sample in the data corpus described in section 3.1, leading to a total of 1,324 and 2,144 microscopic images for qualified and unqualified seeds, respectively.

### 3.2 Morphological characteristics

The non-variant moment's characteristics have been considered morphological characteristics, which have been resourceful for exhibiting the variance among seeds' optical phases. Moreover, these characteristics depend on examining the differences in volume and shape of quantitative and healthy phases of infested seed.

The morphometric information has a significant role in exhibiting abnormal seeds detection. In the case of shape, anemia, and size of seeds became imbalanced regarding the everyday environment. There are nine shape features like major axis, minor axis, area, eccentricity, orientation, perimeter, circularity, and seven moments of invariant that have been mined.

### 3.3 Choosing attributes

Let set  $N$  depicts one of the quantitative phase features, and set  $M$  depicts other optical seeds phase features.

Choose optimum features in respect to both seed's optical phases in the following way

- The feature will be optimum if coverage of the

feature is for both optical phases has been covariant. If this property has been challenged through feature, where coverage at both infected and normal seeds would not be divergent, that feature would be eradicated. Also, this feature optimization has been carried out as follows:

- Values perceived for overall considered features corresponding to both optical seed phases have been represented in a matrix format in the corresponding sequence.
- The corresponding normal seeds matrix has been represented as follows:
- Values perceived in each row have been for overall considered texture and morphological features resulting to normal-seed.
- Similarly, the perceived values for each row in the infected seed matrix represent all of the factors that are thought to go into making an infected seed.
- Furthermore, by predicting the z-score among the observed values for the resultant feature corresponding to both optical phases of the seed, the significance of each feature is assessed. The corresponding feature is deemed optimal when the composite variance is significant at a specific threshold probability.
- By estimating the z-score between the values perceived for each feature during the two optical phases of the seed, the significance of each feature was also ascertained. The resulting feature is deemed ideal if the composite difference is significant at a specific probability threshold.

### 3.4 Composite variance

The datasets n-gram like generated attributes in the 2<sup>nd</sup> stage of the projected model that is organized further for attaining required attributes. Electing the required characteristics depends on variance among co-occurrence of 2 optical seed phases to cater to overall attributes.

For estimating the difference of overall features, the optical phase's co-occurrence in order in the recommended model depends on bi-face integrated variance estimation for statistical measurement. The significance of different assessment models has been concerned for the projected model. The Bi-face variance compound estimation has been held for choosing required attributes related to overall records in the learning database. Following is Manifold group values associated to two distinct vectors differential values set associated to two diversified vectors as explored for estimation has been formulated

```

Begin
    |v1|           // for overall elements
    ∇ {ei∃ei     presents in vector
    i = 1
    ∈ v1}
    av1+ = ei     //identifying the
                    accumulative of overall
                    elements present in the
                    vector

    |v2|           //for all elements
    ∇ {ei∃ej     presents in vector
    j = 1
    ∈ v2}
    av2+ = ej     //identifying the
                    accumulative of overall
                    elements present in the
                    vector

    μ(v1) =  $\frac{av_1}{|v_1|}$  //identifying average of
                    overall elements presents
                    in vector

    μ(v2) =  $\frac{av_2}{|v_2|}$  //identifying average of
                    overall elements presents
                    in vector

    dv1+
    = (μ(v1) - ei)2 //Identifying the
                    accumulative of
                    squared-mean-distance
                    of elements present in
                    the vector

    md(v1, v2) //identifying the variance
    = μ(v1) - μ(v2) among mean-values of
                    elements resents in
                    corresponding vectors

    σ(v1) =  $\sqrt{\frac{dv_1}{|v_1|}}$  //Identifying standard-
                    deviation that is the
                    square-root of variance
                    perceived for vector v1

    σ(v2) =  $\sqrt{\frac{dv_2}{|v_1|}}$  //Identifying standard-
                    deviation that is the
                    square-root of variance
                    perceived for vector v2

    cv(v1,v2) =  $\frac{md(v_1,v_2)}{\sigma(v_1)+\sigma(v_2)}$  //Identifying composite
                    difference of vectors that
                    is the ratio of variance
                    among mean-values in
                    opposed to sum of
                    standard deviations
                    value perceived for both
                    vectors.
    
```

End

Moreover, the degree of composite variance probability has been lower than the preset threshold level; then, it depicts that specified vector distribution has diversified. Hence, the attribute depicting the values of the resulting vector has been regarded to be the required attribute.

### 3.5 Classification procedure

This segment details the classifier utilized in this proposal, the model of the training stage, and the objective function that has been utilized in the classification procedure.

#### 3.5.1 Classifier

The projected classifier has been designed on an adaptive boosting strategy. The classifier has been designed to integrate manifold Boolean-classifiers often signified to be weak-classifiers, which were built by utilizing Cuckoo-search that has been explored in one-time contribution. Every weak-classifier has been built by utilizing optimum features chosen from diversified quantitative stages. These unreliable classifiers divide the given test data according to whether the condition is true or false. Bipartite negatives, which can contain both false positives and false negatives, would be produced by a second weak classifier. Until the overall weak-classifier was thought to have finished the job, this process was repeated. The combined results of all of these Weak-classifiers would also be included in the classification process' final outcomes.

The projected model has been combined with each weak-classifier in this study to determine the most effective ways to extract quantitative seed features for binary classification. The classification process has also been repeated after each weak-classifier iteration; the next iteration of the classifier, also known as boosting, would use the portion of the corpus that was incorrectly classified. Weight classification also recommends employing a weak classifier for each iteration. Records from each of these weak-classifiers would be accurately classified as a result of the iterative completion of weak-classifiers. According to the suggested strategy, each AdaBoost algorithm weak-classifier correctly identifies a particular n-gram.

Additionally, the output of weak classifiers can be used to establish the polarity of the given record. Adaboost classifier has been a resourceful solution to enhance the desired outcome of DT (decision trees) compared to other associated binary classification

Table 1. Corpus dataset of paddy-seed

Total Positives	7920
Total negatives	6140
Positives for training	5607
Negatives for training	5526
Positives for testing	2313
Negatives for testing	614

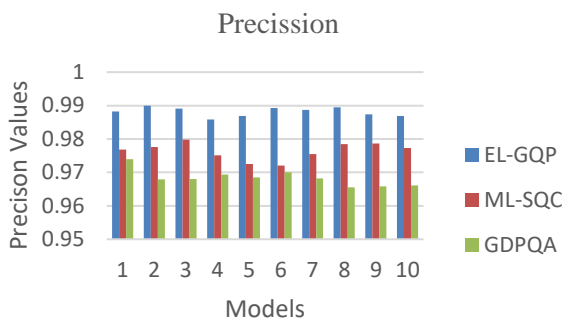


Figure. 1 Precision observed for proposed method EL-GQP, contemporary models ML-SQC, and GDPQA

issues. It might be extensively utilized in augmenting the outcome of distinct algorithms related to ML.

#### 4. Experimental study

##### 4.1 The data

The records corpus labeled negative and positive depicts the germination quality of paddy seed as stated in [39, 42] has been utilized in this simulation. The corpus dataset comprises of negative records 4250, and the number of positive records is 6230. In table 1, the input corpus statistics have been stated. The condition's performance depends on precision, TPR, TNR, and accuracy as significant optimum features elected depending on diversified thresholds set. The simulations have been conducted in cross-validation of 10 folds' format on both the projected EL-GQP model and existing approaches like ML-SQC (machine learning for seed quality classification) [38, 41] and GDPQA (Accurate machine learning-based germination detection, prediction, and quality assessment) [37, 40]. Ensemble algorithms form the [43] authors' are helpful to this research. Moreover, comparative analysis has been representing that the projected EL-GQP model is a possible ensemble supervised learning approach under crucially measured conditions while compared with the performance of existing approaches like GDPQA and ML-SQC.

In the article, performance metrics like precision, accuracy, and specificity are used to assess how well

the ensemble-based precision forming method for predicting seed germination quality performs.

- The percentage of correctly predicted positive samples (true positives) among all positively predicted positive samples (true positives plus false positives) is known as precision. When predicting seed germination, precision would be defined as the percentage of correctly predicted viable seeds among all seeds predicted to be viable.
- The percentage of correctly predicted samples (true positives plus true negatives) among all samples is known as accuracy. The percentage of seeds that were correctly classified as viable or non-viable in a seed germination prediction would be considered accuracy.
- The percentage of samples that were correctly predicted to be negative (true negatives) out of all samples that were predicted to be negative (true negatives plus false positives) is known as specificity. In the context of predicting seed germination, specificity would be defined as the percentage of correctly predicted non-viable seeds among all seeds predicted to be non-viable.

The comparison has been carried out among projected model EL-GQP and existing models ML-SQC and GDPQA by plotting a graph among ten folds on the x-axis and metric precision on the y-axis, as shown in Fig. 1. The metric precision is also called a positive predictive value. From the statistics, the average standard deviation of projected model EL-GQP is  $0.98818 \pm 0.001303$  and contemporary models ML-SQC and GDPQA are  $0.97638 \pm 0.002464$  and  $0.96798 \pm 0.001635$  in respective order. It has been determined that our form carry out enhanced than the ones currently in use.

In Fig. 2, the graph has been plotted among ten folds on the x-axis and metric specificity on the y-axis for the projected EL-GQP and compared with other contemporary models like ML-SQC and GDPQA. From the statistics, the average standard deviation of specificity over the projected model EL-GQP is  $0.95689 \pm 0.004955$ , and contemporary models ML-SQC and GDPQA are  $0.91576 \pm 0.009092$  and  $0.88891 \pm 0.005879$  respectively. It has been determined that our form carry out enhanced than the ones currently in use.

The comparison between projected model EL-GQP and existing models ML-SQC and GDPQA by plotting graph among ten folds on the x-axis and metric accuracy on the y-axis as shown in Fig. 3. From the statistics, the average standard deviation of

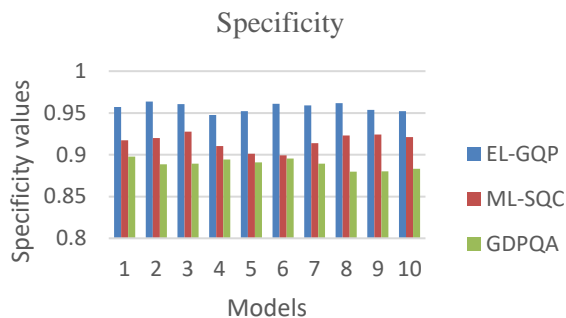


Figure. 2 Specificity observed for proposed method EL-GQP, contemporary models ML-SQC, and GDPQA

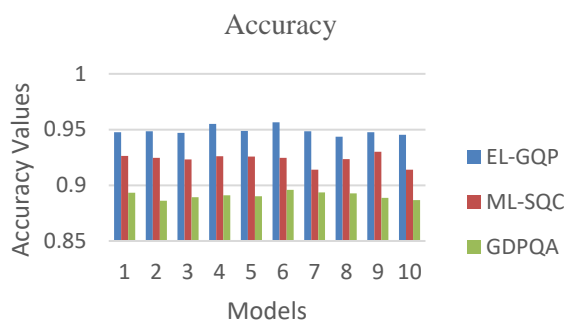


Figure. 3 Accuracy observed for proposed method EL-GQP, contemporary models MK-SQC, and GDPQ

accuracy over the projected model EL-GQP is  $0.94887 \pm 0.003834$  and contemporary models ML-SQC and GDPQA are  $0.9232 \pm 0.004999$  and  $0.89076 \pm 0.002999$  in respective order. It has been determined that our form carry out enhanced than the ones currently in use.

## 5. Conclusion

The contribution of this paper endeavoured to deal with the constraints of the predictive analysis through classification to depict the seed germination quality. Unlike the contemporary models, the proposal is extracting values for the features from quantitative phases. It is observed that the contemporary models downgrade the prediction accuracy if the given microscopic images of seed at their premature level of infection. Predictive analysis based on classification is used in the healthcare industry. Since these microscopic images of seeds are provided, the proposed model, dubbed "Ensemble learning-based germination quality prediction" (EL-GQP), is meant to be trained using the values of the features observed at various quantitative phases. The performance analysis metrics of the EL-represented GQP outcomes were compared to those of existing models in the experimental study. The experimental

study shows that the EL-GQP outperforms other contemporary models in terms of accuracy, sensitivity, and specificity. The fall-out and miss-rate of the proposal are also scaled low compared to contemporary models. But the ratio of fall-out and miss-rate observed from the proposed model EL-GQP are considerable constraints, which leads to future research scope.

## Conflicts of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## Authors' contributions

Mukkoti Maruthi Venkata Chalapathi conceptualised the study, curated the data, performed a formal analysis, devised the methodology, contributed to the coding, A Vijaya Krishna contributed in coding edited the article, G Gopichand devised the methodology, contributed to the coding, validated the content, Kadiyala Ramana supervised the study, validated the content, V C Bharathi helped with reviewed the software, Neeraj Sharma supervised the study.

## References

- [1] W. E. E. F. Savage and G. W. W. Bassel, "Seed vigour and crop establishment: Extending performance beyond adaptation", *J. Exp. Bot.*, Vol. 67, pp. 567–591, 2016.
- [2] G. E. Masry, N. Mandour, S. A. Rejaie, E. Belin, and D. Rousseau, "Recent Applications of Multispectral Imaging in Seed Phenotyping and Quality Monitoring—An Overview", *Sensors*, Vol. 19, p. 1090, 2019.
- [3] Y. Xia, Y. Xu, J. Li, C. Zhang, and S. Fan, "Recent advances in emerging techniques for non-destructive detection of seed viability: A review", *Artif. Intell. Agric.*, Vol. 1, pp. 35–47, 2019.
- [4] C. Wakholi, L. M. Kandpal, H. Lee, H. Bae, E. Park, M. S. Kim, C. Mo, W. H. H. Lee, and B. K. K. Cho, "Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics", *Sens. Actuators B. Chem.*, Vol. 255, pp. 498–507, 2018.
- [5] M. R. Ahmed, J. Yasmin, W. Collins, and B. K. Cho, "X-ray CT image analysis for morphology of muskmelon seed in relation to germination", *Biosyst. Eng.*, Vol. 175, pp. 183–193, 2018.
- [6] A. D. D. Medeiros, D. T. Pinheiro, W. A. Xavier, L. J. D. Silva, and D. C. F. D. Dias, "Quality

- classification of *Jatropha curcas* seeds using radiographic images and machine learning", *Ind. Crops Prod.*, Vol. 146, p. 112162, 2020.
- [7] T. King, M. Cole, J. M. Farber, G. Eisenbrand, D. Zabarar, E. M. Fox, and J. P. Hill, "Food safety for food security: Relationship between global megatrends and developments in food safety", *Trends Food Sci. Technol.*, Vol. 68, pp. 160–175, 2017.
- [8] D. K. Ray, N. D. Mueller, P. C. West, and J. A. Foley, "Yield trends are insufficient to double global crop production by 2050", *PLoS ONE*, 2013.
- [9] J. M. Filho and F. J. Marcos, "Seed vigor testing: an overview of the past, present and future perspective", *Sci Agric Scientia Agricola*, Vol. 72, pp. 363–374, 2015.
- [10] ISTA, "The germination test", *Int Rules Seed Test*, 2015. [Online]. Available: <https://doi.org/10.15258/istarules>. 2015. 05.
- [11] M. R. Ahmed, J. Yasmin, W. Collins, and B. K. Cho, "X-ray CT image analysis for morphology of muskmelon seed in relation to germination", *Biosyst. Eng.*, Vol. 175, pp. 183–193, 2018.
- [12] A. D. D. Medeiros, D. T. Pinheiro, W. A. Xavier, L. J. D. Silva, and D. C. F. D. Dias, "Quality classification of *Jatropha curcas* seeds using radiographic images and machine learning", *Ind. Crops Prod.*, Vol. 146, p. 112162, 2020.
- [13] A. D. D. Medeiros, M. J. Z. León, L. J. D. Silva, A. M. S. Oliveira, and D. C. F. D. Dias, "Relationship between internal morphology and physiological quality of pepper seeds during fruit maturation and storage", *Agron. J.*, 2020.
- [14] É. F. L. Araújo, F. G. G. Junior, A. R. D. Silva, N. Peixoto, and E. R. B. D. Souza, "Evaluation of the desiccation of campomanesia adamantium seed using radiographic analysis and the relation with physiological potential", *Agron. J.*, Vol. 111, pp. 592–600, 2019.
- [15] D. Kusumaningrum, H. Lee, S. Lohumi, C. Mo, M. S. Kim, and B. K. Cho, "Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy", *J. Sci. Food Agric.*, Vol. 98, pp. 1734–1742, 2018.
- [16] Y. W. Seo, C. K. Ahn, H. Lee, E. Park, C. Mo, and B. K. Cho, "Non-Destructive Sorting Techniques for Viable Pepper (*Capsicum annum* L.) Seeds Using Fourier Transform Near-Infrared and Raman Spectroscopy", *J. Biosyst. Eng.*, Vol. 41, pp. 51–59, 2016.
- [17] G. C. Andrade, C. M. M. Coelho, and V. G. Uarrota, "Modelling the vigour of maize seeds submitted to artificial accelerated ageing based on ATR-FTIR data and chemometric tools (PCA, HCA and PLS-DA)", *Heliyon*, Vol. 6, p. e03477, 2020.
- [18] A. Ambrose, S. Lohumi, W. H. H. Lee, and B. K. Cho, "Comparative non-destructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy", *Sens. Actuators B Chem.*, Vol. 224, pp. 500–506, 2016.
- [19] R. M. Silverstein, F. X. Webster, and D. Kiemle, *Spectrometric Identification of Organic Compounds*, 7th ed., John Wiley & Sons, Inc., 2005.
- [20] C. Li, T. Zhao, C. Li, L. Mei, E. Yu, Y. Dong, J. Chen, and S. Zhu, "Determination of gossypol content in cottonseeds by near infrared spectroscopy based on Monte Carlo uninformative variable elimination and non-linear calibration methods", *Food Chem.*, Vol. 221, pp. 990–996, 2017.
- [21] P. Mukasa, C. Wakholi, C. Mo, M. Oh, H. J. J. Joo, H. K. Suh, and B. K. K. Cho, "Determination of viability of *Retinispora* (Hinoki cypress) seeds using FT-NIR spectroscopy", *Infrared Phys. Technol.*, Vol. 98, pp. 62–68, 2019.
- [22] G. L. Jiang, "Comparison and Application of Non-Destructive NIR Evaluations of Seed Protein and Oil Content in Soybean Breeding", *Agronomy*, Vol. 10, p. 77, 2020.
- [23] C. Pasquini, "Near infrared spectroscopy: A mature analytical technique with new perspectives—A review", *Anal. Chim. Acta*, Vol. 1026, pp. 8–36, 2018.
- [24] B. Nugraha, P. Verboven, S. Janssen, Z. Wang, and B. M. Nicolai, "Non-destructive porosity mapping of fruit and vegetables using X-ray CT", *Postharvest Biol. Technol.*, Vol. 150, pp. 80–88, 2019.
- [25] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, and O. Busto, "Data fusion methodologies for food and beverage authentication and quality assessment—A review", *Anal. Chim. Acta*, Vol. 891, pp. 1–14, 2015.
- [26] L. Benedet, W. M. Faria, S. H. G. Silva, M. Mancini, L. R. G. Guilherme, J. A. M. Demattê, and N. Curi, "Soil subgroup prediction via portable X-ray fluorescence and visible near-infrared spectroscopy", *Geoderma*, Vol. 365, p. 114212, 2020.
- [27] A. Chaugule, "Application of image processing in seed technology: a survey", *Int. J. Emerg. Technol. Adv. Eng.*, Vol. 2, No. 4, pp. 153–159, 2012.



- [28] B. Rar, A. Ali, S. Naeem, F. Jamal, Elgarhy. M, Chesneau C. Discrimination of sunflower seeds using multispectral and texture dataset in combination with region selection and supervised classification methods. *Chaos*. 2020 Nov;30 (11):113142. doi: 10.1063/5.0024017. PMID: 33261340.
- [29] D. A. Carroll, J. C. Brown, and P. Robson, "Using k-NN to analyze images of diverse germination phenotypes and detect single seed germination in *Miscanthus sinensis*", *Plant Methods*, Vol. 14, p. 5, 2018.
- [30] R. Masteling, L. Voorhoeve, I. Jsselmuiden, J. "DiSCount: computer vision for automated quantification of *Striga* seed germination", *Plant Methods*, Vol. 16, No. 60, (2020).
- [31] N. Genze, R. Bharti, M. Grieb, S. J. Schultheiss, D. G. Grimm, "Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops", *Plant Methods*, 2020 Dec Vol. 22, No. 16(1), p. 157. doi: 10.1186/s13007-020-00699-x. PMID: 33353559; PMCID: PMC7754596.
- [32] D. L. Claros, X. Meng, E. Custovic, G. Deng, O. Berkowitz, J. Whelan, and M. G. Lewsey, "Developmental normalization of phenomics data generated by high throughput plant phenotyping systems", *Plant Methods*, 2020 Aug Vol. 12, No. 16, p. 111. doi: 10.1186/s13007-020-00653-x. PMID: 32817754; PMCID: PMC7424680.
- [33] M. Rodrigues, F. G. G. Junior, and J. M. Filho, "Vigor-S: System for Automated Analysis of Soybean Seed Vigor", *Journal of Seed Science*, Vol. 42, 2020.
- [34] U. Škrubej, Č. Rozman, and D. Stajnkó, "Assessment of germination rate of the tomato seeds using image processing and machine learning", *Eur J Hortic Sci*, Vol. 80, pp. 68–75, 2015.
- [35] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata, "Satellite Image Time Series Classification With Pixel-Set Encoders and Temporal Self-Attention", *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 12322-12331, doi:0.1109/CVPR42600.2020.01234.
- [36] T. T. Nguyen, V. N. Hoang, T. L. Le, T. H. Tran, and H. Vu, "A Vision Based Method for Automatic Evaluation of Germination Rate of Rice Seeds", In: *Proc. of 2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, Ho Chi Minh City, Vietnam, 2018, pp. 1-6, doi: 10.1109/MAPR.2018.8337511.
- [37] R. K. Madapudi, A. A. Rao, and G. Merugu, "Change requests artifacts to assess impact on structural design of SDLC phases", *Int'l J. Computer Applications*, Vol. 54, No. 18, pp. 21-26, 2012.
- [38] M. R. Kumar and V. K. Gunjan, "Review of machine learning models for credit scoring analysis", *Ingeniería Solidaria*, Vol. 16, No. 1, 2020.
- [39] R. K. Madapudi and P. C. S. Mahesh, "HBS-CRA: scaling impact of change request towards fault proneness: defining a heuristic and biases scale (HBS) of change request artifacts (CRA)", *Cluster Computing*, Vol. 22, No. 5, pp. 11591-11599, 2019.
- [40] A. D. Medeiros, L. J. D. Silva, J. P. O. Ribeiro, K. C. Ferreira, J. T. F. Rosas, A. A. Santos, C. B. D. Silva, "Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-ray Imaging", *Sensors (Basel)*, 2020 Aug, Vol. 3, No. 20(15), p. 4319. doi: 10.3390/s20154319. PMID: 32756355; PMCID: PMC7435829.
- [41] Available online: <https://www.kaggle.com/datasets/c6caf46ef419a4cb653b4f8872a8136bb17151b106abd84821742b19d6db2447>
- [42] M. M. V. Chalapathi, "Self-Learning Based Emotion Recognition using Data Analytics", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 9, No. 1, pp. 2278-3075, Nov. 2019.
- [43] K. Ramana, A. Rajanikanth, M. R. Kumar, G. Nagaraja, A. V. Krishna, and P. Nagendra, "Leaf disease classification in smart agriculture using deep neural network architecture and IoT", *Journal of Circuits, Systems and Computers*, Vol. 31, No. 15, p. 2240004, 2022.
- [44] R. Jhaveri, A. Revathi, K. Ramana, R. Raut, and D. R. Kumar, "A review on machine learning strategies for real-world engineering applications", *Mobile Information Systems*, Vol. 2022.
- [45] N. Penchalaiah, J. N. Emmanuel, S. S. Kamal, and K. Ramana, "IoT based automatic irrigation system using wireless sensor networks", In: *Proc. of the 3rd International Conference on Communications and Cyber Physical Engineering*, pp. 1255-1272, 2021.