WILEY | Hindawi

Research Article

Application of Mung Bean Protein Separation and Purification Combined with Artificial Intelligence MLR Classifier Technology in the Study of Protein Physical and Chemical Properties

Xianqing Wang ⓘ, Jimin Bai, and Yueting Luan

College of Food Science, Heilongjiang Bayi Agricultural University, Daqing 163319, China

Correspondence should be addressed to Xianqing Wang; wangxianqing@byau.edu.cn

Within the scope of this project, a spectroscopy-dependent machine learning (ML) method will be utilized to estimate the optimal harvest time for mung bean, which will be used to examine the changes in physical and chemical attributes of the bean as it develops. It was decided to harvest mung bean from the R5 (initial seed), R6 (full seed), and R7 (beginning maturity) stages. The spectral reflectance of the pods was measured, and their physical and chemical characteristics were characterized. The experiment was carried out using a spectrophotometer with a wavelength range of 360–740 nm. On the basis of the qualities that have been identified so far in the study, early, ready, and late specimens have all been included. The results showed that the pod/bean weight and pod thickness reached their maximum at R6. After that, everything remained the same as before. Around R6, there was an increase in sugar, carbs, amino acids, and glycine, among other things. The ML approach (random forest classification) achieved an accuracy of 0.95 for the classification of pods dependent on their spectral reflectance. Specimens can be classed as "early" or "late" depending on whether or not they are "ready" or "not ready" when they are collected or processed. As a result, this procedure is the most effective choice available. It can figure out when the best time is to harvest mung bean.

1. Introduction

One of the most popular soy products in the United States is mung bean, which has been consumed in East Asia for millennia and is represented in Figure 1. Isoflavones and other nutrients such as vitamins C and E and monounsaturated fatty acids are found in significant concentrations in this food [1]. Although several factors contribute to the nutritional value and eating quality of mung bean [2, 3], these quality characteristics change over time as the bean matures into an adult bean. Mung bean's high marketability and customer acceptance can be attributed in part to the fact that it is harvested at the height of morphological and eating quality at the optimal period, according to some authors [3]. It is

also easier to process mung bean if it is dependably good quality. Picking during the R6 and R7 growth phases, when the pods are beginning to turn yellow and when moisture and bean weight are nearing their maximum values, will ensure the highest yield and quality.

Collecting mung bean outside of the right harvest window has the potential to diminish its economic viability due to the dynamic nature of bean growth during the R6 and R7 phases. Picking beans too soon may result in a lower quality bean, while harvesting them too late results in fibrous and yellow beans (as opposed to green ones) [5]. Farmers are limited in the time they have to harvest mung bean [5] due to the fact that they have only a one-week window after they reach their optimal harvest time. It is possible to distinguish

Figure 1: Mung bean.

a plant in the R6 growth stage from one in R7 by the presence of pods that have beans that have completely filled the pod cavity [6]. There are a number of biological variations during the transition from R6 to R7, which indicate that reproductive growth has been completed and that senescence has begun (R8). Bean growth takes up 85–90% of the limited pod space, leaves, pods, and beans turn yellow, and sugars and other chemical elements accumulate in the early beans, to name a few of the changes.

In this work, we propose a framework in order to predict the harvest time using the spectral machine learning technique which is a standard computation procedure that encompasses the random forest classification technique.

2. Related Works

Research on mung bean and soybean has been conducted at various stages of development in order to better understand their physical properties and chemical contents. The physical, chemical, and antinutritional characteristics of mung beans were studied by [7], who investigated the effects of mung bean development on the beans themselves. [8] Protein, oil, starch, and soluble saccharides in soybean seeds, as well as differences in seed length, were analyzed for modification content ranging from R1 to R8. The oil, protein, carbohydrates, starches, organic acids, and amino acids of growing soybean seeds were examined in a study by [9]. The metabolism and accumulation of oligosaccharides in the plant were studied during the development of soybeans. The physical and chemical features of mung bean or soybean seeds that alter over the course of their reproductive life were well depicted by all of these studies. However, just a few researches have looked at using these differences to predict the best harvest period for mung bean.

Mung bean growers' capacity to notice changes in the plant's look, texture, or flavor is the basis for most present harvesting systems for mung beans. Inexperienced or inexperienced mung bean growers may find it difficult to use these approaches due to their subjective character, and they may suffer financially as a result of the lower grade mung bean harvested outside of the optimal window. As a result, it would be excellent if technologies for quickly, consistently, and uniformly establishing the appropriate harvest time could be developed. Strawberries [11], cherry tomatoes [12], and apples [11] have all recently had their optimal harvest times established using spectroscopic methods [13]. This is an important drawback of our study because no previous research has used spectroscopic methods to estimate the optimal harvest time for mung bean. Portable spectroscopic equipment also shows promise because it gives a short (typically a few seconds) way for picking the best time to harvest mung bean in the field, as opposed to the longer chemical investigation. Even more minute differences in pod color during mung bean growth can be detected via spectroscopic analysis, which is otherwise impossible to see with the naked eye. Secondarily, in order to establish the best harvest time for the crops of mung beans, spectroscopy-dependent analysis must be calibrated against an appropriate reference technique. Multivariate regression testing is frequently used for calibration, but because spectra are so complicated, it does not always provide satisfactory results [5]. Despite this, recent improvements in machine learning techniques have provided a chance to analyze the complex spectroscopic information in order to deliver accurate and trustworthy calibration [14]. "Random forest" (RF) is an ensemble learning technique that has grown in popularity in recent years due to its high classification accuracy and speed when applied to huge datasets, according to [15]. Radiofrequency (RF) technology has been used to classify various types of food using multispectral and hyperspectral data. As an example, infrared spectroscopy was used to distinguish between genuine and fake nutmeg, and the results showed that RF was superior to other classification approaches, such as partial least squares-discriminant analysis (PLS-DA) and soft independent modeling of class analogies (SIMCA). For the second time, a machine learning study was done to categorize bananas into separate categories dependent on their measurable features (i.e., artificial neural network, random forest, and support vector machines). Following the experiments, it was shown that the RF strategy had the best classification accuracy among the three machine learning methods.

Mung bean's physical and chemical properties will be studied to identify when to harvest, and spectroscopy-dependent machine learning will be used to make this determination. The physical and chemical features of mung bean harvested at various stages, ranging from R5 to R7, were examined in this study. Pod weight, 20-bean weight, pod dimensions (width, length, and thickness), and color were among the physical characteristics. Sucrose, fructose, glucose, alanine, glycine, oligosaccharides (raffinose and stachyose), moisture of fresh beans, protein and starch, neutral detergent fiber (NDF), and ash were all included in the chemical compositions. Mung bean harvesting can be improved by using the physical and chemical features of harvested beans to determine the optimal harvesting stage and to identify the beans that were harvested "too early" and "too late." In addition, handheld, portable spectrophotometers were used to measure the reflectance of the harvested mung bean pods between 360 and 740 nm. A machine learning approach was utilized to evaluate whether the mung bean harvesting was ready dependent on the gathered spectral reflectance, using the observed spectra. For the first time, researchers have used spectroscopy to develop a method for
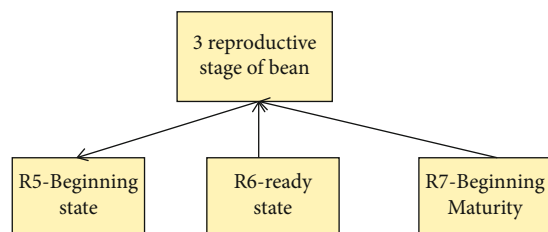
Figure 2: Reproductive phases of mung bean.

Table 1: Various harvest phases with planting date.

Dates	Harvest phase	Days from the planting
21.02.2021	First harvest phase (R5-1)	0
19.06.2021	Second harvest phase (R5-2)	119
25.06.2021	Third harvest phase (R6-1)	125
30.06.2021	Fourth harvest phase (R6-2)	130
05.10.2021	Fifth harvest phase (R7-1|)	135
11.10.2021	Sixth harvest phase (R7-2)	141

quickly and correctly determining the best time to pick mung bean, a step necessary to ensuring a steady supply of marketable, high-quality soybeans.

3. Methodologies

3.1. Basic Material of Mung Bean Plant. Plantings of R15-10280, V16-0547, and the UA-Kirksey variety were made at Kentland Farm, Whitethorne, VA, in May 2019. A randomized complete block design (RCBD) with 7-m-long rows and 76-cm spacing between each row was used for the plots. The planting rate is around 18 seeds every meter; thus, there are 126 seeds per row in total. Rows include roughly 95 plants with an emergence rate of 75%. There were a total of three replications in the study. The first nine feet of each entry were tagged with three dates once blossoming had begun. Depending on the genotype's flower availability, anything from 25 to 30 nodes was marked in each replication. As a result of all genotypes being in maturity group V and having been planted in the same field on the same date, the duration phases of all genotypes are comparable. To make things even more interesting, the three genotypes' flower tagging dates were only two days apart. Genotypes R15-10280, V16-0547, and UA-Kirksey had blooming tag dates of October 5 and 11 and July 30 for the following years. When necessary, the nodes were cleaned to remove all of the younger flowers, making harvesting easier and allowing us to accurately track the flowering time of each node. At six separate times, pods were hand-harvested. There were six harvests in 2020 corresponding to R5-1, R5-2, R6-1, R6-2, R7-1, and R7-2 of the growth stages (Figure 2).

It is important to note that the morphological and chemical features of pods and beans changed swiftly due to the dry weather and a drought field, as there was only a one-day difference between R7-1 and R7-2 growth phases. Table 1 presented offer specifics on when to sow and harvest soybeans. For each of the six harvest periods, a total of 10 pods were selected from each genotype in all three replications, resulting in a total of 90 pods per harvest period. Six harvests yielded a total of 540 pods. In order to eliminate any dirt or debris from the pods, they were brushed off during harvest, although they were not thoroughly cleaned. So that the pubescence could be preserved and the pod's color would not be altered by possible harm, this was done.

Once harvested, the samples were stored in Ziplock bags and brought to the lab where their physical parameters were measured using an ice pack in a chiller. Plantings of R15-10280, V16-0547, and the UA-Kirksey variety were made

at Kentland Farm, Whitethorne, VA, in May 2019. A randomized complete block design (RCBD) with 7-m-long rows and 76-cm spacing between each row was used for the plots. The planting rate is around 18 seeds every meter; thus, there are 126 seeds per row in total. Rows include roughly 95 plants with an emergence rate of 75%. There were a total of three replications in the study. The first nine feet of each entry were tagged with three dates once blossoming had begun. Depending on the genotype's flower availability, anything from 25 to 30 nodes was marked in each replication. As a result of all genotypes being in maturity group V and having been planted in the same field on the same date, the duration phases of all genotypes are comparable.

3.2. Physical Properties. A computerized fractional caliper was used to measure the pods' length, width, and thickness in the lab (Husky Tools). From the thickest to the tiniest section of each pod, it was measured. Additionally, pod weight was determined using analytical balances. In order to get a 20-bean weight, random pods from each genotype and replication were opened. Sample lightness (+)/darkness (), redness (+)/greenness (), and yellowness (+)/blueness were measured with a portable Konica Minolta CM-700 Spectrophotometer. The pods were returned to the ultra-low freezer at 80 C and labeled Ziploc bags until further chemical composition testing could be undertaken.

3.3. Chemical Composition

3.3.1. Alanine, Glycine, and Free Sugars. Hand-shelling and freeze-drying mung bean from the pods yielded freeze-dried mung bean. A blender was utilized to crush the dry beans into a powder, which was then analyzed for chemical composition using a 500-m sieve. All of the sugars and amino acids were extracted using the technique reported by [19] with minor adjustments, as well as the raffinose and stachyose oligosaccharides (alanine, glycine, and fructose). In a 2-mL centrifuge tube, dried samples weighing 0.15 g were combined with 1.5 mL of deionized water (DI water). After 2 hours of shaking at room temperature, the combination was centrifuged for 10 minutes at a speed of 13,500 g. After that, 750 L of supernatant and 750 L of acetonitrile were collected and combined. At room temperature, the mixture was mixed for ten minutes before being centrifuged for a further 10 minutes. This was followed by the use of high-performance liquid chromatography (HPLC)

Table 2: Relative sweet of certain amino acids.

Amino acids	Relative sweet
Alanine	0.95-1.65
Glycine	0.45-1.15
Glucose	0.35-0.75
Fructose	0.95-1.70
Sucrose	0.99

with an index detector (RID) to analyze 750 L of supernatant (Agilent Technologies, Santa Clara, CA, USA).

3.3.2. Total Sweet Taste. The sweet of various free sugars and amino acids varies. The sweet of various sugars and compounds was compared to sucrose in Table 2 to determine their relative sweet (RS), which was then used to compute the overall sweet. The following equation was used to determine the overall sweet taste as shown in equation (1).

$$\text{Sum Total(sweet)} = RS_{Alanine} \times C_{Alanine} + RS_{Glycine} \\ \times C_{Glycine} + RS_{Fructose} \times C_{Fructose} \\ + RS_{Sucrose} \times C_{Sucrose} + RS_{Glucose} \\ \times C_{Glucose}. \quad (1)$$

3.3.3. Moisture, Protein, Fat, Neutral Detergent Fiber (NDF), Starch, and Ash Assessment. Oven drying at 105°C for an hour and a half was used to measure fresh bean moisture levels [18]. According to [18], total nitrogen content was determined by multiplying 6.25 times (equation (2)) the protein conversion factor to get the total protein content (AOAC, 2001.11). The fat content was measured using AOAC 2003.05.

$$\text{Total Nitrogen} = 6.25 * \text{total protein content[AOAC]}. \quad (2)$$

The fat content was assessed using AOAC 2003.05. The fat content was extracted using petroleum ether and determined using AOAC 2003.05. The ANKOM fiber analyzer was used to determine the amount of NDF present (ANKOM Technology, Macedon, NY, USA). Using the digested dry weight, the NDF content was estimated after the nonfiber portion was removed using a neutral detergent solution [21]. According to AOAC 942.05, the weight difference before and after incineration was used to calculate the ash content in a muffle furnace operating at 550°C for 12 hours as shown in equation (3). Measurement of hydrolyzed glucose was done using the HPLC with RID and Bio-Rad Aminex HPX-87H, as reported by [22].

$$\text{Total Ash Content at 550°C} \\ = \text{weight of AOAC (before)} - \text{weight of AOAC(after)}. \quad (3)$$

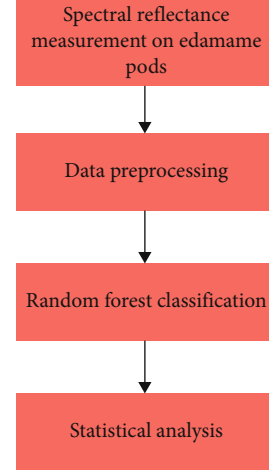*3.4. Prediction of Harvest Time Using the Spectral Machine Learning Technique*

Figure 3: Steps in prediction of harvest time depends on spectral ML technique.

3.4.1. Spectral Machine Learning. Before presenting our primary findings, we briefly review the various spectral approaches used in machine learning and explain how our findings apply to both traditional and modern algorithms. Let $x1, \cdots, xn$ be a set of Rm data points.

Spectral approaches are classified by their reliance on:

Outer point cloud properties: these include PCA and Fisher discriminant analysis. That is, they need spectral analysis of a positive-definite kernel of dimension m.

MDS and newer variations rely on it (more or less) to execute an embedding of the data points. They require spectroscopic features of a point cloud of dimension n. Large datasets (intrinsic or extrinsic) make spectrum analysis difficult. Our essay focuses on obtaining the optimum rank-k approximation to a symmetric, positive semidefinite (SPSD) matrix for methods like PCA and MDS.

The forecasting of the harvest time depends on the spectral machine learning method which consists of four steps as shown in Figure 3:

3.4.2. Data Preprocessing. Each sample was categorized into three categories: "early class," "ready class," and "late class" depending on physical and chemical data gathered from the sample. Every one of the ten pods in every sample was given the same class label in order to match the spectral reflectance dataset ($n = 10$ for each sample). As a result, the data set comprises of mung bean pod spectral reflectance readings in the 540-nm range. Some 220 of these datasets were delayed class, 180 early class, and 140 ready-class observations. To train and evaluate the RF classifier, we employed a feature matrix that included both the core spectral data and the first-order derivatives (FOD). It is standard procedure to use FOD transformations of the spectral curve in order to enhance spectral characteristics and reduce random noise in order to improve classification quality.

3.4.3. Random Forest Classification. A data mining technique known as random forest (RF) can be used to solve classification and regression issues. Voting to determine the class type

and then growing a group of trees has greatly increased classification accuracy. These ensembles are grown using random vectors. Using a random set of vectors, a tree is created for each one. Classification and regression trees are used in RF. Trees are used to address classification difficulties. The RF prediction is determined by the majority of class votes. Since overfitting does not occur in big RFs, the generalization error merges to a limiting number when adding more trees to the RF [26]. To improve accuracy, it is essential to have low bias and correlation. With no pruning and randomization of variables applied at each node, low correlation can be achieved in the trees. The following summarizes the RF's general growth and voting process:

Each RF tree is created from a bootstrap sample taken from the training data. The remaining one-third of samples is used to calculate out-of-bag (OOB) error, with the remaining two-thirds of samples being utilized to develop each tree.

  (i) A random sample of n variables is chosen from a pool of N variables during the tree-growing process

  (ii) As a starting point, $\log 2(\mathrm{N}) + 1$ or N can be chosen, and various values can be attempted until the least OBB error is obtained. Each node uses only one of the n specified variables to make the best split possible. Trees can be tested using OOB datasets once

they have been grown. The OOB data collection is used by RF to compute an unbiased error estimate as more trees are planted. It is also used to calculate the relevance of variables in RF classification using OOB data set

An adaptation of the RF analysis approach was developed using the R programming language and VSURF/caret packages [23]. The general equation is

$$\mathrm{RFC} = \frac{1}{n}\sum_{i=1}^{n}(r_i - a_i). \qquad (4)$$

From equation (4), n indicates the total data points, r_i indicates the returned value by the model, and the actual value of data point is represented by a_i.

We used 10-fold repeated cross-validation on the training data and divided the dataset $80:20$ between training and test data. This procedure was repeated 100 times, and the mean accuracy was obtained. In this study, RF classifiers trained to assign each spectrum to one of the three classes or two classes (e.g., early vs. late) were used to compare the model performance as evaluated by classifier accuracy depending on cross-validation findings. In order to figure out accuracy, we used

$$\mathrm{Accuracy} = \frac{\mathrm{True\ negative} + \mathrm{True\ positive}}{\mathrm{True\ positive} + \mathrm{False\ Positive} + \mathrm{True\ negative} + \mathrm{False\ Negative}}. \qquad (5)$$

The 39 predictor factors were used to classify the three groups (waveband at 10-nm resolution). According to training data, the best classifier was found to have an mtry of 30 and n-tree of 2000. The VSURF program was used to pick the characteristics, and the accuracy of the forecast was assessed using specified spectral bands.

4. Experimental Analysis

4.1. Physical Properties. Weight of pods, 20-bean weight, size, and color of pods, as well as the breadth and length of pods, were all documented as the mung bean grew from R5 to R7 (Table 3). Childish pods with green beans that are completely formed are ideal for high-quality mung bean harvests, and the beans should be dry at the harvest period [24]. The pod and 20-bean weights of all three genotypes grew significantly from R5 to R6 and peaked at stage R6-1 or R6-2. The genotypes V16-0547 and UA-Kirksey showed a minor but not statistically significant decline through stage R7, while the genotype R15-10280 showed a slight rise from R7-1 to R7-2. According to the results of pod width and length measurements, this increase was caused by the selection at random of larger mung bean pods. Seed filling starts during stage R5 dependent on soybean growth and develop-

ment, and dry mass builds at the same time (Purcell et al., 2014). Growth and dry mass buildup slow to a halt at the conclusion of R5, which marks the beginning of R6. A modest but not substantial drop in pod and bean weights occurred when the beans entered stage R7 [26]. Because the pods had already reached their maximum size and had a set pod width and length by the end of R4, pod width and length changes were not as significant as weight changes. The larger seeds, on the other hand, caused a considerable rise in pod thickness from R5 to R6. Over R7, there were no notable changes in pod thickness following seed filling. Consumers like brighter, greener pod [2].

Each of the three genotypes saw a substantial growth in $G*$ values from the beginning to the end of the study. The weight of the pods increased as they grew, as indicated by the increase in the $G*$ value. There were no significant changes in $x*$ values between R5-1 and R6-2, which indicated that the pods were green. All three R6-2 to R7-2 genotypes showed an increase in $x*$, but only the R15-10280 and UA-Kirksey genotypes were statistically significant. This shows that the green color was preserved until R7-1, at which point it began to fade away. Over the course of pod growth, $y*$ values rose as shown in equation (6). A rise in genotype V16-0547 was significant, but not for R15-10280

Table 3: Calculated physical properties of mung bean in various harvest phase.

Harvest phase	Pod weight/g	20 bean weight/g	Pod thickness/mm	Pod length/mm	Pod width/mm
R15-10280 (genotype phase)					
R5-1	0.85 ± 0.11^z	3.85 ± 0.31^z	10.85 ± 0.53^z	40.85 ± 0.76^z	5.85 ± 0.43^z
R5-2	1.35 ± 0.23^{yz}	5.35 ± 0.53^{yz}	11.35 ± 0.63^{yz}	41.35 ± 0.54^{yz}	7.35 ± 0.93^{yz}
R6-1	1.97 ± 0.41^x	8.97 ± 0.61^x	12.97 ± 0.43^x	45.97 ± 0.64^x	6.97 ± 0.74^x
R6-2	1.79 ± 0.13^{xy}	7.79 ± 0.53^{xy}	11.79 ± 0.54^{xy}	43.79 ± 0.78^{xy}	8.79 ± 0.48^{xy}
R7-1	1.49 ± 0.03^{xy}	11.49 ± 0.303^{xy}	12.49 ± 0.76^{xy}	42.49 ± 0.64^{xy}	5.49 ± 0.35^{xy}
R7-2	1.90 ± 0.19^{xy}	10.90 ± 0.14^{xy}	10.90 ± 0.76^{xy}	46.90 ± 0.63^{xy}	8.90 ± 0.90^{xy}
V16-0547 (genotype phase)					
R5-1	0.97 ± 0.29^z	8.97 ± 0.23^z	10.97 ± 1.76^z	40.97 ± 0.54^z	7.97 ± 0.49^z
R5-2	1.31 ± 0.13^{yz}	3.31 ± 0.33^{yz}	12.31 ± 0.83^{yz}	47.31 ± 0.43^{yz}	6.31 ± 0.63^{yz}
R6-1	1.93 ± 0.29^{xy}	6.93 ± 0.39^{xy}	13.93 ± 0.87^{xy}	43.93 ± 0.76^{xy}	7.93 ± 0.89^{xy}
R6-2	2.20 ± 0.31^x	4.20 ± 0.61^x	12.20 ± 0.37^x	42.20 ± 0.65^x	8.20 ± 0.91^x
R7-1	1.79 ± 0.23^{xy}	9.79 ± 0.34^{xy}	11.79 ± 0.54^{xy}	41.79 ± 0.56^{xy}	9.79 ± 0.13^{xy}
R7-2	1.71 ± 0.23^{xy}	5.71 ± 0.342^{xy}	13.71 ± 0.72^{xy}	44.71 ± 0.86^{xy}	5.71 ± 0.33^{xy}
UA- Kirksey (genotype phase)					
R5-1	0.85 ± 0.11^z	7.85 ± 0.432^z	10.85 ± 0.11^z	0.85 ± 0.57^z	6.85 ± 0.61^z
R5-2	1.11 ± 0.17^{yz}	11.11 ± 0.343^{yz}	13.11 ± 0.17^{yz}	1.11 ± 0.23^{yz}	8.11 ± 0.87^{yz}
R6-1	1.69 ± 0.25^x	10.69 ± 0.43^x	12.69 ± 0.98^x	1.69 ± 0.98^x	5.69 ± 0.25^x
R6-2	1.76 ± 0.98^{xy}	8.76 ± 0.32^{xy}	11.76 ± 0.25^{xy}	1.76 ± 0.56^{xy}	9.76 ± 0.45^{xy}
R7-1	1.65 ± 0.24^{yz}	9.65 ± 0.54^{yz}	13.65 ± 0.34^{yz}	1.65 ± 0.72^{yz}	5.65 ± 0.94^{yz}
R7-2	1.96 ± 0.02^x	4.96 ± 0.54^x	12.96 ± 0.87^x	1.96 ± 0.23^x	7.96 ± 0.22^x

Depending on one-way ANOVA with Tukey's HSD analysis (p less than 0.05), various letters (xyz) from each column show a statistically significant variation.

or UA-Kirksey (both of which remained stable at 32.62 and 35.98) (32.88 to 39.43). Mung bean growth resulted in yellower pods as seen by rising $y*$ values (a measure of yellowness).

$$\text{HSD test} = [G \times a \times b] \text{ of edamame in various harvest phase.} \tag{6}$$

Overall, the pods' colors became lighter, less green, and more yellow as they matured. Odds are the green chlorophyll catabolites in mung bean that have decreased from R6 to R8 [27] as shown in Table 4.

4.2. Alanine, Glycine, and Free Sugars

4.2.1. Various Free Sugars and Amino Acids That Contribute to Sweetness. After R6-1, the amount of sucrose in the blood decreased. Sucrose concentrations in R15-10280 and V16-0547 genotypes were 82.0 mg/g at R5-1, but jumped to 125.3 mg/g and 111.54 mg/g, respectively, at R6-1. Free sugars (sucrose, fructose, and glucose) and sweet are the content of mung bean. There were more sucrose molecules than fructose and glucose in mung bean [18]. Even though sucrose concentration began to decline after R6-1, samples obtained at R6-2 showed no significant difference from those harvested at R6-1. Immediately following R6, the sugar concentration decreased precipitously. UA-Kirksey showed the same sucrose changes as R15-10280 and V16-0547, starting at 68.7 mg/g at R5-1, climbing to 76.3 mg/g at R6-1, and then decreasing at R7-2. It was less noticeable in this group compared to the other two genotypes (R15-10280 and V16-0547). Sucrose content peaked in the early stages of R6 and then began to fall as time went on. In the growing bean embryos, most of the sugar is sucrose, and it builds up in the beans during the filling process [28]. The seed's metabolism and storage are influenced by a variety of enzyme changes throughout its growth [29]. Sucrose synthase activity was shown to be correlated with mung bean sucrose concentrations, according to [3]. As the beans progressed from stage R4 to R6, sucrose synthase activity increased five-fold, according to a study published in [10]. The increased activity of sucrose synthase led to an increase in sucrose content from R5 to R6. During bean growth, the sucrose concentration decreases as sucrose is converted to alternative storage sugars such raffinose and stachyose. When oligosaccharide levels are examined further, a change in galactinol synthase activity may also account for the drop in sucrose content. The fructose and glucose concentrations of all three genotypes were substantially lower than sucrose. When compared to the other three genotypes, genotypes R15-10280, V16-0547, and UA-Kirksey had the highest fructose (17.0 mg/g) and glucose (6.4 mg/g) levels at R5. UA-Kirksey R5-2, R6-1, and R6-2 genotypes all showed significant declines in fructose and glucose over time. There was no change in these two sugars until R7-2, when they dropped along the same stage. Free alanine and glycine have

TABLE 4: Calculated color of $[G * a * b *]$ value of mung bean in various harvest phases.

Harvest phase	$G *$	$a *$	$b *$
R5-1	70.45 ± 4.26^{yx}	8.97 ± 0.61^{x}	35.87 ± 2.542^{zx}
R5-2	68.45 ± 1.69^{x}	7.79 ± 0.53^{xy}	23.32 ± 7.23^{z}
R6-1	58.45 ± 0.23^{x}	6.93 ± 0.39^{xy}	33.12 ± 4.54^{xz}
R6-2	64.54 ± 0.34^{y}	4.20 ± 0.61^{x}	32.87 ± 2.542^{zx}
R7-1	43.76 ± 1.53^{zy}	5.49 ± 0.35^{xy}	23.32 ± 7.23^{z}
R7-2	68.45 ± 1.69^{x}	3.85 ± 0.31^{z}	32.87 ± 2.542^{zx}
R8-1	69.76 ± 1.43^{zy}	7.90 ± 0.61^{x}	35.12 ± 4.54^{xz}
R8-2	57.45 ± 1.39^{x}	4.79 ± 0.53^{xy}	32.87 ± 2.542^{zx}
R5-2	58.45 ± 0.23^{x}	8.97 ± 0.61^{x}	34.87 ± 2.542^{zx}
R6-1	64.54 ± 0.34^{y}	6.97 ± 0.74^{x}	39.45 ± 1.43^{yz}
R6-2	43.76 ± 1.53^{zy}	8.79 ± 0.48^{xy}	32.87 ± 2.542^{zx}
R7-1	68.45 ± 1.69^{x}	5.49 ± 0.35^{xy}	33.87 ± 2.542^{zx}
R7-2	58.45 ± 0.23^{x}	7.93 ± 0.89^{xy}	38.87 ± 2.342^{zx}
R8-1	59.76 ± 1.53^{zy}	8.97 ± 0.61^{x}	32.12 ± 4.54^{xz}
R8-2	47.45 ± 1.69^{x}	4.79 ± 0.54^{xy}	36.87 ± 2.92^{zx}
R5-1	64.54 ± 0.34^{y}	8.20 ± 0.81^{x}	23.72 ± 9.23^{z}
R5-2	43.76 ± 1.53^{zy}	9.79 ± 0.13^{xy}	33.87 ± 1.2^{zx}
R6-1	68.45 ± 1.69^{x}	4.20 ± 0.61^{x}	36.87 ± 0.42^{zx}
R6-2	58.45 ± 0.23^{x}	3.31 ± 0.33^{yz}	34.87 ± 0.982^{zx}
R7-1	64.54 ± 0.34^{y}	3.85 ± 0.31^{z}	39.45 ± 1.43^{yz}
R7-2	43.76 ± 1.53^{zy}	12.49 ± 0.76^{xy}	23.32 ± 7.23^{z}
R8-1	71.76 ± 1.63^{zy}	6.97 ± 0.71^{x}	32.12 ± 5.53^{xz}
R8-2	59.45 ± 1.69^{x}	11.79 ± 0.33^{xy}	31.87 ± 2.459^{zx}

from the V16-0547 genotype that had the highest overall sweetness were picked between R5-2 and R6-2 of the ripening cycle. The total sweetness of samples taken at different growth phases did not change significantly for UA-Kirksey.

$$\text{total sweet of } R6-1 \text{ and } R6-2 > \text{total sweet of } R5 \text{ and } R7. \tag{7}$$

Although all sweet components in mung bean were evaluated, the general trend of total sweet was similar to increases in sucrose concentration, especially for genotypes R15-10280 and V16-0547.

As sugar content in mung bean is higher than in most other foods, it has a greater impact on the overall sweet of the food. Author [18] found a strong connection among sucrose concentration and overall sweet in a prior study.

4.2.3. Raffinose and Stachyose. Raffinose and stachyose are galactosyl derivatives of sucrose and belong to the raffinose family of oligosaccharides (RFOs) [31]. Seeds of legume plants store energy in the form of oligosaccharides that build during maturation [32]. There are two galactose molecules coupled to one sucrose in the raffinose and stachyose structures, respectively. There are two types of sugars that are not good for you: raffinose and stachyose, which are not digestible by humans and may cause flatulence or more serious gastrointestinal problems [18]. Raffinose and stachyose are change from harvest stage 5 to 7 A major increase of raffinose began in genotype R6-2 for all three genetic lines, which had previously shown no or very low levels of raffinose. Beans harvested at R6-1 have a trace amount of stachyose; following R6-2, however, the amount increases significantly. The findings were consistent with previous investigations by [8], which reported the similar pattern of buildup of raffinose and stachyose. Because raffinose synthase is ultimately responsible for the RFO synthesis, sucrose and galactinol are both required for this process. Galantine synthase activity remained high in ripening soybean from R6 to R7, which triggered the buildup of galactinol. The galactinol content in mature soybean increased from practically 0 to 3.0 mg/seed. To produce raffinose, an enzyme known as raffinose synthase uses the beans' sucrose and galactinol as the raw materials. Stachyose was found to be more abundant than raffinose, as seen in Figures 4 and 5. In order to make stachyose, raffinose is required as a starting material, which results in the addition of an extra galactinol [33]. As a result, stachyose buildup began at R6-1 and escalated after R6-2. Sucrose levels had risen and fallen in lockstep. Sulfate content reduced after R6-1; a greater decrease was reported in the next R6-2. Accumulation of raffinose and stachyose may explain the drop in sucrose following R6-1.

4.2.4. Macronutrient Organization. Mung bean's macronutrient content is just as important to consumers as its look and sweet. It was for this reason that measurements were made of the contents of the foodstuff (Table 5). More than 60% of the weight of mung bean was water, making it the most prevalent component. Among the most essential

sweet tastes, which may increase the sweetness of mung bean. Table 5 shows the shift in free alanine and glycine levels from R5 to R7. The amount of free alanine and glycine in proportion to the number of free sugars was extremely low. At stage R6-1, R15-alanine 10280's and glycine concentrations were 7.9 mg/g and 1.60 mg/g, respectively. All of the genotypes showed no significant change in alanine content between R5 and R7, indicating that the bean's development stage had no effect on alanine production. A large rise in glycine content occurred from R5 to R6 and a subsequent fall by R7 in genotype R15-10280; however, in R6-1 to R6-2 and then no change in genotype UA-Kirksey while growing beans, the glycine concentration in genotype V16-0547 did not change. An apparent trend in the alanine and glycine levels of mung bean could not be identified between developmental stages 5 and 7. Because alanine and glycine quantities in mung bean are so low, they have no impact on the flavor.

4.2.2. Sweet Total. According to reports, mung bean with a superior level of customer acceptance was found [30]. The total sweetness of mung bean harvested at R6-1 and R6-2 was higher than that of mung bean harvested at R5 and R7 for genotype R15-10280 as shown in equation (7). Samples

TABLE 5: Chemical composition of mung bean in various harvest phases.

| Harvest phase	Fat (%)	Starch (%)	Protein (%)	Moisture content of fresh beans (%)	Ash (%)	NDF (%)
R5-1	18.80 ± 7.23^{yx}	11.45 ± 1.43^{xy}	39.45 ± 1.43^{yz}	69.45 ± 7.23^{zx}	5.08 ± 0.72^{zx}	7.08 ± 0.72^{zx}
R5-2	16.16 ± 7.23^{zx}	10.45 ± 0.32^{y}	40.54 ± 0.23^{y}	70.45 ± 4.26^{yx}	5.34 ± 0.72^{yx}	5.34 ± 0.72^{yx}
R6-1	11.18 ± 7.23^{xy}	11.45 ± 6.23^{zx}	42.45 ± 1.43^{z}	68.45 ± 8.52^{y}	3.66 ± 0.43^{y}	3.66 ± 0.43^{y}
R6-2	14.14 ± 7.23^{x}	12.45 ± 1.23^{x}	43.45 ± 0.232^{xy}	63.45 ± 1.45^{z}	4.75 ± 0.32^{z}	5.08 ± 0.72^{zx}
R7-1	13.08 ± 7.23^{yz}	9.45 ± 2.342^{yz}	44.45 ± 0.87^{xz}	62.45 ± 6.34^{y}	5.32 ± 0.545^{y}	5.34 ± 0.72^{yx}
R7-2	12.19 ± 7.23^{yx}	6.45 ± 5.87^{y}	45.45 ± 0.54^{z}	66.45 ± 8.54^{zx}	5.54 ± 0.78^{zx}	3.66 ± 0.43^{y}
R5-1	17.15 ± 7.23^{y}	10.45 ± 9.45^{yz}	46.45 ± 1.87^{xy}	69.45 ± 4.78^{x}	5.63 ± 0.43^{zx}	5.08 ± 0.72^{zx}
R5-2	15.19 ± 7.23^{yz}	10.45 ± 6.38^{xz}	47.45 ± 1.54^{x}	67.45 ± 9.32^{xy}	3.76 ± 0.23^{yx}	5.34 ± 0.72^{yx}
R6-1	13.12 ± 7.23^{xy}	11.45 ± 8.32^{yz}	49.45 ± 0.23^{xz}	63.45 ± 3.54^{x}	4.65 ± 0.98^{y}	3.66 ± 0.43^{y}
R6-2	13.17 ± 7.23^{xz}	12.45 ± 0.43^{zy}	43.45 ± 0.65^{zy}	61.45 ± 6.3^{xy}	3.54 ± 0.43^{z}	6.08 ± 0.72^{zx}
R7-1	12.15 ± 7.23^{xy}	9.45 ± 0.543^{x}	40.45 ± 0.87^{x}	68.45 ± 1.69^{x}	3.76 ± 0.34^{y}	7.34 ± 0.72^{yx}
R7-2	11.13 ± 7.23^{zx}	8.45 ± 0.342^{y}	44.45 ± 1.54^{yz}	58.45 ± 0.23^{x}	4.76 ± 0.56^{zx}	5.66 ± 0.43^{y}
R5-1	16.14 ± 7.23^{yx}	10.45 ± 0.54^{yx}	39.45 ± 0.54^{xy}	64.54 ± 0.34^{y}	5.34 ± 0.23^{zx}	7.08 ± 0.72^{zx}
R5-2	14.13 ± 7.23^{zy}	9.45 ± 1.786^{xy}	46.45 ± 0.34^{x}	43.76 ± 1.53^{zy}	3.7 ± 0.26^{yx}	7.34 ± 0.72^{yx}
R6-1	12.12 ± 7.23^{yx}	9.45 ± 0.32^{xz}	42.45 ± 0.47^{xz}	32.12 ± 4.54^{xz}	5.89 ± 0.45^{y}	5.66 ± 0.43^{y}
R6-2	11.11 ± 7.23^{z}	9.45 ± 0.543^{xy}	44.45 ± 1.23^{xy}	32.87 ± 2.542^{zx}	5.32 ± 0.41^{z}	7.08 ± 0.72^{zx}
R7-1	16.14 ± 7.23^{y}	8.45 ± 1.32^{xz}	43.45 ± 0.57^{yx}	23.32 ± 7.23^{z}	3.54 ± 0.6^{y}	8.34 ± 0.72^{yx}
R7-2	15.11 ± 7.23^{xy}	7.45 ± 0.23^{xy}	42.45 ± 1.78^{xz}	34.66 ± 7.23^{yx}	4.32 ± 0.5^{zx}	5.66 ± 0.43^{y}
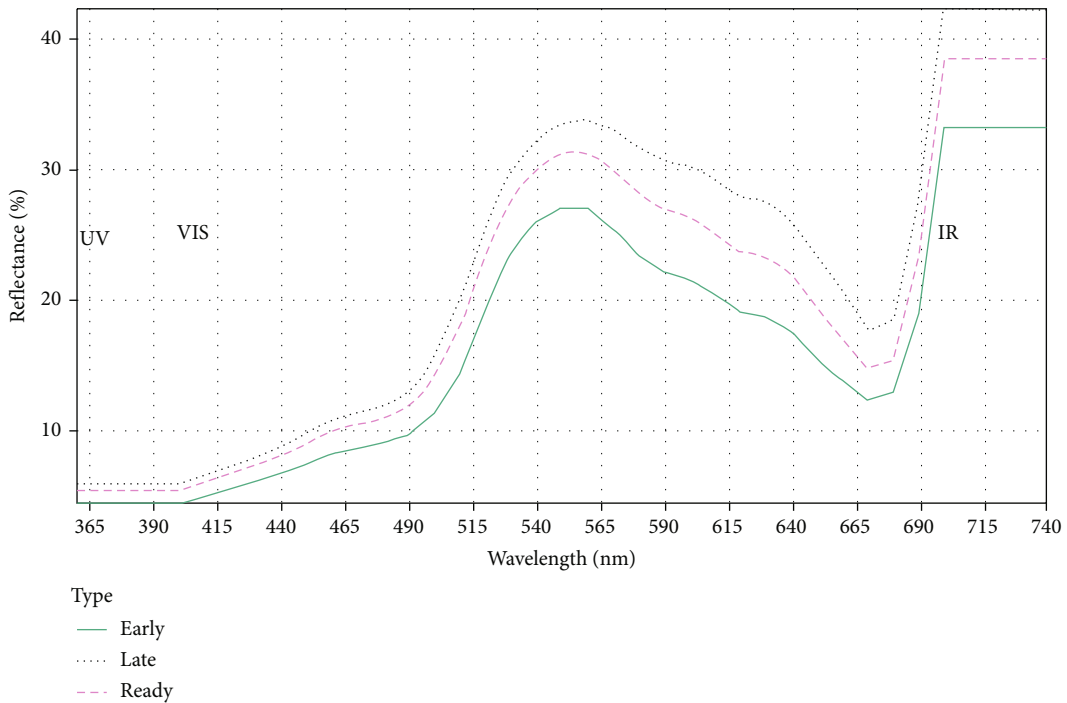
FIGURE 4: Analysis of spectral reflectance using machine learning. Average spectral reflectance of mung bean pods that are in the three categories: early, late, and ready to harvest.

qualities of mung bean is its high moisture content. A high moisture level of fresh beans was found in all genotypes at R5-1, and the moisture content subsequently declined between R5-1 and R7-2. For genotypes V16-0547 and UA-Kirksey, the reduction from R5-1 to R6-1 was significant, but not for R15-10280. Mung bean's chewiness and hardness are largely due to the starch that makes up the majority of the seed's energy storage [34]. In all three genotypes, the
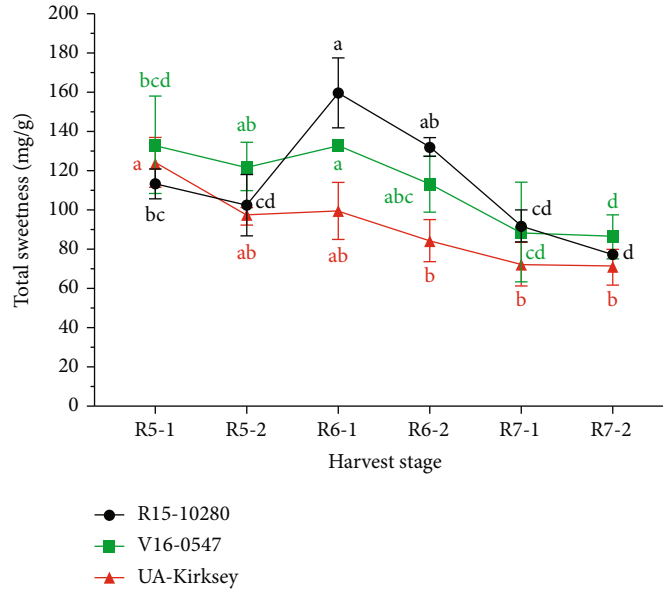
FIGURE 5: Total sweetness of mung bean in different harvest stages.

initial starch level was around 11%; however, the starch content of genotypes R15-10280 and V16-0547 increased somewhat from R5-1 to R6-1, whereas for UA-Kirksey, the starch content did not change. From R6-1 to R7-2, the starch content declined dramatically in all three genotypes. Reports of this modification were also made. Starch accumulated at the beginning of the bean's development and subsequently decreased dramatically when the bean had nearly filled its pods in their study. The physical qualities of mung bean are influenced by the amount of starch they contain after harvesting. The enzymes in mung bean can be deactivated, and the shelf life extended by blanching. It softens mung bean by causing starch and pectin gelatinization and solubilization. The higher the starch content, the softer the mung bean will be after blanching. Mung bean is a rich source of vegetable protein because it contains about 39.5% protein by dry weight. All harvested mung bean samples had protein levels ranging from 38.9% to 40.90%, with a minor rise of 3–4% from R5-1 to R7-2.

Researchers [35] found a small rise in this variable (2011). The protein content of mung bean increased between R5 and R7 by 2–5 percent. From R5 to R7, the protein content remained largely consistent across the board. During the R5-1 stage, the fat content of R15-10280, V16-0547, and UA-Kirksey was 13.8 percent in R15 and 12.7 percent in V16, respectively. When it collected up to R6-2, no substantial difference was discovered. Initially, R5-1 had an NDF concentration of 7% to 8% and an ash content of 6%, respectively. Genotypes V16-0547 and UA-NDF Kirksey's content declined dramatically from R5-1 to R6-1 and then marginally increased subsequently. There was no substantial variation in NDF for R15-10280 compared to the other two genotypes. In all three genotypes, the ash concentration decreased from R5 to R6 and increased from R6 to R7.

The best time to harvest mung bean is in the early stages of R6 (R6-1) depending on physical and chemical features.

This phase of mung bean development results in a large bean size, a pleasant light green color, a higher level of sweetness, a higher protein and starch content, and lower levels of raffinose and stachyose. There was no significant difference in physical and chemical attributes between the mung bean that was harvested at R6-2 and that was collected at R6-1.

However, more improvements in accuracy are required (Figures 6 and 7).Similar spectroscopy-dependent approaches were previously utilized to predict the time of apple harvest. To determine the ripeness of the red apple, researchers employed UV–Vis and near-infrared spectroscopes paired with partial least square regression to monitor chlorophyll content (the green hue) on the skin. The biggest distinction between our study [13] that in our case, the response variable is categorical rather than uninterrupted. As a result, mung bean and other vegetables and fruits may be harvested at the appropriate moment utilizing machine learning technologies like RF.

4.2.5. Selected Wavelength Using Three Class Classifiers. There may be difficulties in gathering data in the field and repetitive data if a large number of spectral bands are used. Analyzing massive amounts of data provided by spectroscopic techniques is a difficult task. Data analysis can be improved by picking wavelengths that contain the most useful information with the least amount of unnecessary data. When assessing the performance of models, 12 wavelengths from the Prim spectral data and 9 wavelengths from the FOD spectral data were picked for examination. In the RF technique, a machine learning process known as feature selection automatically selected the wavelengths (11 and 8). Prim and FOD spectral data from these specified wavelengths were used to train three-class RF classification models, and the results are shown in Figure 6 as Prim/Prim, FOD/FOD, FOD/Prim, and Prim/FOD. Intriguingly, the accuracy of the models constructed using only a few key
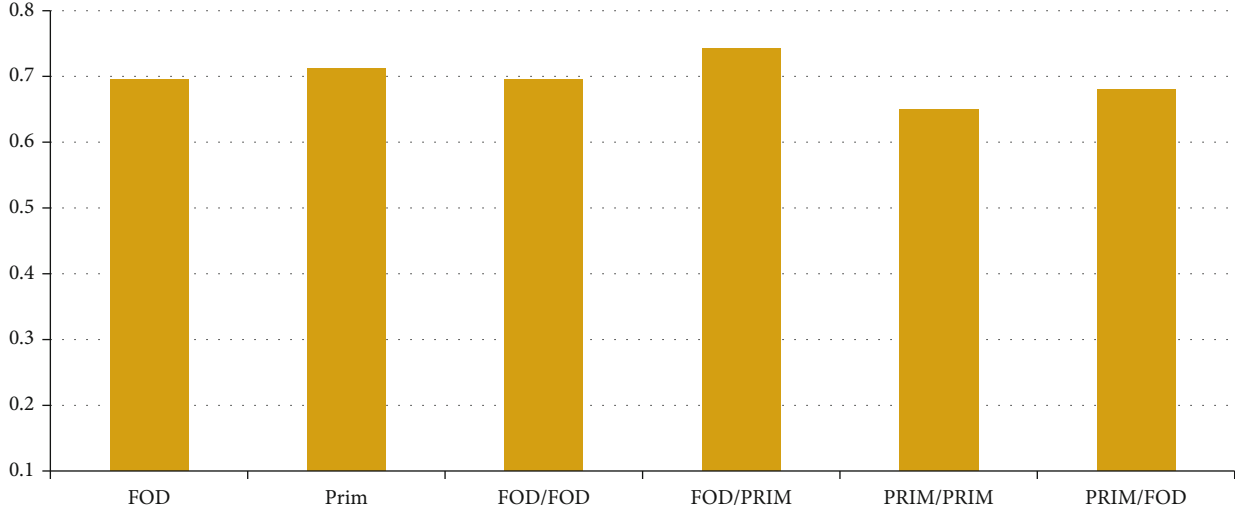
FIGURE 6: Comparison of model accuracy between various classification techniques.
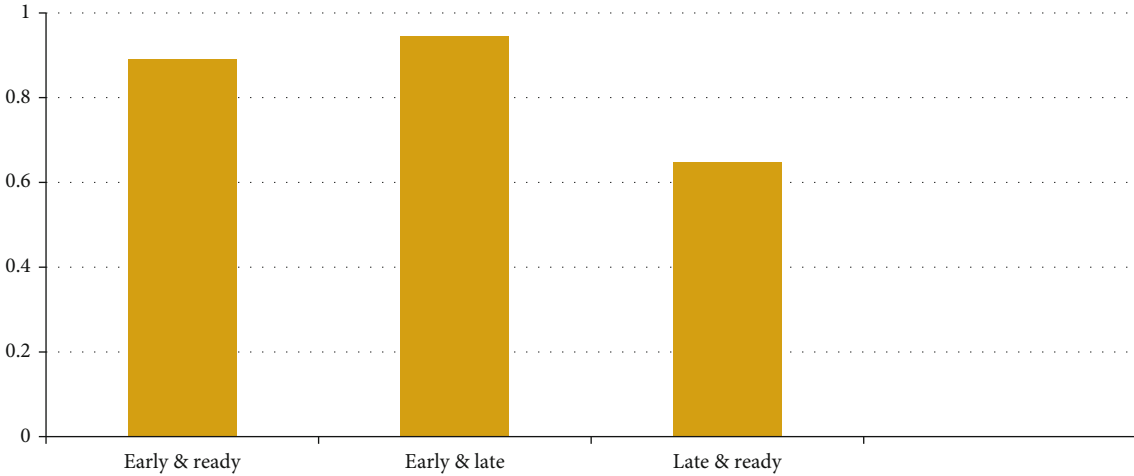

FIGURE 7: Classification accuracy of two categories.

wavelengths was comparable (0.65 to 0.73) to that of the models constructed using all of the available wavelengths. The precision and recall for the classifier are as shown in equations (8) and (9):

$$precision = \frac{True\ positive}{True\ positive + False\ positive}, \qquad (8)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative}. \qquad (9)$$

Hyperspectral image systems were used to assess strawberry ripeness using support vector machine (SVM) models constructed on full spectra and chosen ideal spectra [36]. They discovered that SVM models dependent on ideal spectra were superior to models dependent on full spectra in terms of performance. Unique spectral ranges were used in their investigation, which is why their results varied from those of other researchers. Ideal wavelengths between 441.1 and 1013.97 nm produced satisfactory results, while wave-

lengths between 941.46 and 1578.13 nm produced inferior results. Thus, this research shows that the decrease in wavelength number may be achieved without impacting the accuracy of the model for categorizing mung bean at various maturation stages, as demonstrated by previous studies as well as overall, the model trained using FOD spectral data outperforms the model trained with Prim spectral data. Due to FOD's superior ability to resolve overlapped wavebands and reduce random noise, this discovery was not surprising. According to [37], in a thorough evaluation of spectral classification algorithms, the least improvement (over primary spectra) was found in FOD-dependent approaches for very complex datasets, whereas the most improvement was found in less complicated datasets. The use of an RF classifier in conjunction with FOD spectra improved classification accuracy in this investigation. Primary spectra and random forest classifiers still have high classification accuracy. Other research suggests that the FOD spectra's ability to increase classification accuracy is influenced by the technique and number of classes utilized
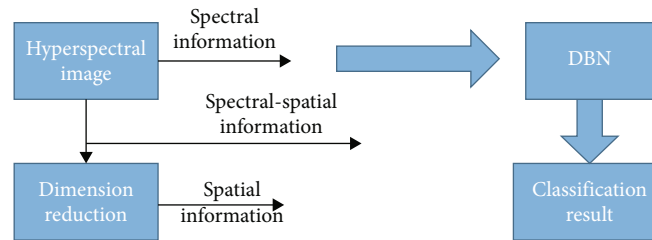
FIGURE 8: Two-class classifier for hyperspectral image.

in the classification. In spite of this, the FOD spectral data does not indicate a distinct separation of the curves.

4.2.6. Two Class Classifiers. Spectral data of the "ready" category were difficult to distinguish, while early and late data were easy to tell apart using our approach. Using solely primary spectral data, two-class classifications were made between each of the two groups studied. We did not employ FOD since its visual inspection does not allow us to distinguish clearly between the categories of "early" and "late" (data not shown). When classifying as early or late, the model accuracy climbs to 0.91, whereas when classifying as early or ready, the model accuracy rises to 0.88 (Figure 7). At best, the model's ability to discern "late" from "ready" is only 0.69 percent accurate. Twelve wavelengths were selected as the most important for feature selection. When utilized to represent the primary spectral data of adzuki beans, selective wavelength modeling produces similar precision. It was found that the model had difficulty distinguishing between the "ready" and "late" categories of spectral data. Using solely primary spectral data, two-class classifications were made between each of the two groups studied. We did not employ FOD because its visual inspection does not distinguish between early versus late stages of development categorizations of things (data not shown). It climbs to 0.91 for labeling as either early or late, whereas it rises to 0.88 for categorizing as either early or ready. The model can only tell the difference between "late" and "ready" with an accuracy of 0.69 percent. After a thorough process of feature selection, the most significant 12 wavelengths were determined. When utilized to represent the primary spectral data of adzuki beans, selective wavelength modeling produces similar accuracy.

Cucumber chilling injury was detected using a hyperspectral imaging technique (Figure 8), according to [38], and the increased model accuracy for two-class classifications matched their findings. There were no errors in any of the two-class categories, but there were errors in the three-class classifications (normal, lightly chilling, and severely chilling), with an overall accuracy of only 91.6 percent. An RF classification system was found to reasonably identify early, late, and ready stages of harvesting for the soybean crop based on data from handheld, portable spectrometers.

5. Conclusions

Using a spectroscopy-dependent machine learning method, this research examined at how mung bean's physical and chemical properties vary over the course of bean growth with the highest accuracy of 97.8%. At stage R6, pod weight, bean weight, and pod thickness reach their maximum values. With age, the greens of mung bean turn a darker shade of brown. The chemical composition of all genotypes changes from R5 to R7 in a similar way. Sweetness in mung beans is associated with higher levels of sucrose, glycine, and starch in R6. In contrast, the fat, NDF, and ash concentration are minimal at this point in the process. Mung bean should be harvested at the early R6 (R6-1) stage, which takes into account all the changes in physical attributes and chemical composition that occur during bean growth. R6-2, on the other hand, is suitable and even preferable to R5 and R7 if a longer harvest window is required. Classifying early and late samples with a spectral reflectance of 0.95 and 0.87, respectively, was achieved using the machine learning method, which was dependent on pods' spectral reflectance. Classifying "late" and "ready" samples, however, yielded a low accuracy of 0.68. Mung bean harvesting times may be predicted using machine learning methods dependent on the spectrum reflectance of pods.

Data Availability

The analyzed datasets generated during the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. Johnson, S. Wang, and A. Suzuki, "Edamame: a vegetable soybean for Colorado," *Energy (Kcal)*, vol. 582, p. 573, 2000.

[2] S. Zeipiņa, I. Alsiņa, and L. Lepse, "Insight in edamame yield and quality parameters: a review," *Research for rural development*, vol. 2, pp. 40–45, 2017.

[3] J. Konovsky, T. A. Lumpkin, and D. McClary, "Edamame: the vegetable soybean," *Understanding the Japanese Food and Agrimarket*, vol. 1, no. 1, pp. 173–181, 2020.

[4] D. Moseley, M. P. Da Silva, L. Mozzoni et al., "Effect of planting date and cultivar maturity in edamame quality and harvest window," *Frontiers in Plant Science*, vol. 11, 2020.

[5] L. C. Carson, J. H. Freeman, K. Zhou, G. Welbaum, and M. Reiter, "Cultivar evaluation and lipid and protein contents of Virginia-grown edamame," *HortTechnology*, vol. 21, no. 1, pp. 131–135, 2011.

[6] M. Licht, *Soybean Growth and Development*, Iowa StateUniversity Extension and Outreach, Ames, Iowa, USA, 2014.

[7] Y. Xu, A. Cartier, D. Kibet et al., "Physical and nutritional properties of edamame seeds as influenced by stage of development," *Journal of Food Measurement and Characterization*, vol. 10, no. 2, pp. 193–200, 2016.

[8] X. Saldivar, Y. J. Wang, P. Chen, and A. Hou, "Changes in chemical composition during soybean seed development," *Food Chemistry*, vol. 124, no. 4, pp. 1369–1375, 2011.

[9] B. Yazdi-Samadi, R. W. Rinne, and R. D. Seif, "Components of developing soybean seeds: oil, protein, sugars, starch, organic acids, and amino acids1," *Agronomy Journal*, vol. 69, no. 3, pp. 481–486, 1977.

[10] C. A. Lowell and T. M. Kuo, "Oligosaccharide metabolism and accumulation in developing soybean seeds," *Crop Science*, vol. 29, no. 2, pp. 459–465, 1989.

[11] F. Shen, B. Zhang, C. Cao, and X. Jiang, "On-line discrimination of storage shelf-life and prediction of post-harvest quality for strawberry fruit by visible and near infrared spectroscopy," *Journal of Food Process Engineering*, vol. 41, no. 7, 2018.

[12] H. Q. Yang, "Nondestructive prediction of optimal harvest time of cherry tomatoes using VIS-NIR spectroscopy and PLSR calibration," in *In Advanced Engineering Forum*, vol. 1, pp. 92–96, Trans Tech Publications Ltd, 2011.

[13] E. Bertone, A. Venturello, R. Leardi, and F. Geobaldo, "Prediction of the optimum harvest time of 'Scarlet' apples using DR-UV-Vis and NIR spectroscopy," *Postharvest Biology and Technology*, vol. 69, pp. 15–23, 2012.

[14] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "Machine learning for high-throughput stress phenotyping in plants," *Trends in Plant Science*, vol. 21, no. 2, pp. 110–124, 2016.

[15] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: a review of applications and future directions," *ISPRS journal of photogrammetry and remote sensing*, vol. 114, pp. 24–31, 2016.

[16] F. B. de Santana, W. B. Neto, and R. J. Poppi, "Random forest as one-class classifier and infrared spectroscopy for food adulteration detection," *Food Chemistry*, vol. 293, pp. 323–332, 2019.

[17] E. J. Peidad, J. I. Larada, G. J. Pojas, and L. V. V. Ferrer, "Postharvest classification of banana (Musa acuminata) using tier-based machine learning," *Postharvest Biology and Technology*, vol. 145, pp. 93–100, 2018.

[18] X. Yu, F. Yuan, X. Fu, and D. Zhu, "Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds," *Food Chemistry*, vol. 196, pp. 776–782, 2016.

[19] S. Machado, A. S. Costa, F. Pimentel, M. B. P. Oliveira, and R. C. Alves, "A study on the protein fraction of coffee silverskin: protein/non-protein nitrogen and free and total amino acid profiles," *Food Chemistry*, vol. 326, p. 126940, 2020.

[20] G. G. Birch and S. E. Kemp, "Apparent specific volumes and tastes of amino acids," *Senses*, vol. 14, no. 2, pp. 249–258, 1989.

[21] J. OHair, Q. Jin, D. Yu et al., "Non-sterile fermentation of food waste using thermophilic and alkaliphili Bacillus licheniformis YNP5-TSU for 2, 3-butanediol production," *Waste Management*, vol. 120, pp. 248–256, 2021.

[22] B. C. Vidal, K. D. Rausch, M. E. Tumbleson, and V. Singh, "Determining corn germ and pericarp residual starch by acid hydrolysis," *Cereal Chemistry*, vol. 86, no. 2, pp. 133–135, 2009.

[23] R. H. J. Heim, I. J. Wright, H.-C. Chang et al., "Detecting myrtle rust (Austropuccinia psidii) on lemon myrtle trees using spectral signatures and machine learning," *Plant Pathology*, vol. 67, no. 5, pp. 1114–1121, 2018.

[24] H. Born, "Edamame: vegetable soybean," 2016, https://attra.ncat.org/product/edamame-vegetable-soybean/.

[25] L. Purcell, M. Salmeron, and L. Ashlock, "Soybean Growth and Development," in *Arkansas Soybean Production Handbook*, p. 197, Arkansas Cooperative Extension Service Miscellaneous Publications, 2014.

[26] L. Wang, T. Wang, and W. R. Fehr, "Effect of seed development stage on sphingolipid and phospholipid contents in soybean seeds," *Journal of Agricultural and Food Chemistry*, vol. 54, no. 20, pp. 7812–7816, 2006.

[27] D. Borrmann, J. C. de Andrade, and U. M. Lanfer-Marquez, "Chlorophyll degradation and formation of colorless chlorophyll derivatives during soybean (Glycine max L. Merill) seed maturation," *Journal of Agricultural and Food Chemistry*, vol. 57, no. 5, pp. 2030–2034, 2009.

[28] Y. Du, Q. Zhao, L. Chen et al., "Effect of drought stress during soybean R2–R6 growth stages on sucrose metabolism in leaf and seed," vol. 21, no. 2, p. 618, 2020.

[29] K. Sitthiwong, M. Toshiyuki, O. Nobuyuki, and H. Suzuki, "Changes in carbohydrate content and the activities of acid invertase, sucrose synthase and sucrose phosphate synthase in vegetable soybean during fruit development," *Asian Journal of Plant Sciences*, vol. 4, no. 6, pp. 684–690, 2005.

[30] R. Carneiro, S. Duncan, S. O'Keefe et al., "Utilizing consumer perception of edamame to guide new variety development," *Frontiers in sustainable food systems*, vol. 4, 2021.

[31] V. Kumar, A. Rani, L. Goyal et al., "Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location," *Journal of Agricultural and Food Chemistry*, vol. 58, no. 8, pp. 5081–5085, 2010.

[32] T. M. Kuo, J. F. VanMiddlesworth, and W. J. Wolf, "Content of raffinose oligosaccharides and sucrose in various plant seeds," *Journal of Agriculturaland Food Chemistry*, vol. 36, no. 1, pp. 32–36, 1988.

[33] Y. Jing, S. Lang, D. Wang, H. Xue, and X. F. Wang, "Functional characterization of galactinol synthase and raffinose synthase in desiccation tolerance acquisition in developing Arabidopsis seeds," *Journal of Plant Physiology*, vol. 230, pp. 109–121, 2018.

[34] D. G. Stevenson, J. L. Jane, and G. E. Inglett, "Structures and physicochemical properties of starch from immature seeds of soybean varieties (Glycine max (L.) Merr.) exhibiting normal, low-linolenic or low-saturated fatty acid oil profiles at maturity," *Carbohydrate Polymers*, vol. 70, no. 2, pp. 149–159, 2007.

[35] D. L. Dornbos and M. B. McDonald, "Mass and composition of developing soybean seeds at five reproductive growth stages1," *Crop Science*, vol. 26, no. 3, pp. 624–630, 1986.

[36] Z. Gao, Y. Shao, G. Xuan, Y. Wang, Y. Liu, and X. Han, "Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning," *Artificial Intelligence in Agriculture*, vol. 4, pp. 31–38, 2020.

[37] A. Ghiyamat, H. Z. M. Shafri, G. A. Mahdiraji, A. R. M. Shariff, and S. Mansor, "Hyperspectral discrimination of tree species with different classifications using single- and multiple-end-member," *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, no. 1, pp. 177–191, 2013.

[38] H. Cen, R. Lu, Q. Zhu, and F. Mendoza, "Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification," *Postharvest Biology and Technology*, vol. 111, pp. 352–361, 2016.