

# **Milled Rice Grain Grading Using Raspberry Pi With Image Processing and Support Vector Machines With Adaptive Boosting**

by

**Ariel D. Catli**  
**Karl Jethro R. Cui**  
**Joy Abigail F. Tan**  
**Earlvic Owen S. Zuñega**

A Thesis Report Submitted to the School of Electrical, Electronics, and  
Computer Engineering in Partial Fulfilment of the Requirements for the Degree

Bachelor of Science in Computer Engineering

Mapúa University  
September 2017

## **Chapter I**

### **INTRODUCTION**

Rice is the staple food of the Philippines. It is a major agricultural commodity that is mass produced in the country and also in other developing countries (Juliano, 2016). In the first half of 2015, around 7.6 million metric tons of rice is produced by the Philippines (Philippine Statistics Authority, 2016). The Philippine Grain Standardization Program is a government program spearheaded by the National Food Authority to provide commercial assessment standards for the determination of the grade and quality of milled rice products. The implementation of the program started on September 21, 2002. From its establishment, the National Grains Standard has been formed (National Food Authority, 2001). The National Grains Standards defined the characteristics classification of the rice grain samples. Factors for determining grade include dimensional length, degree of milling, percentage by weight of broken kernels, brewers, red kernels, immature kernels, chalky kernels, damaged kernels, yellow kernels, age-related changes, and other characteristics. The grades are based on the percentage by weight of the classified grains to the overall weight of the product. The Grade 5 is the lowest and the Premium grade is the highest grade a milled rice product can be classified to. Moreover, the implementation of these standards in the market is expected to boost the quality of the rice products in the Philippines. The NGS not only defined provisions about grading, but several packaging regulations, labelling, and quality testing procedures are also outlined. The program institutionalizes the standards to promote inclusive growth and better-quality products. Using the standards retailers, farmers, and distributors can grade their products accordingly. However, the process of grading is still manual and is highly subjective.

Assessment experts rely on their own perceptual inference and the manual measurement using simple tools only. The differences in the assessment could render the standards pointless.

Existing studies are aimed to develop simple, affordable, and accessible grading methodologies based on different standard specifications (Concepcion, et al., 2015). The US Department of Agriculture provided the US Standards for rough, brown, and milled rice (US Standards for Rice, 2009). Moreover, several other foreign standards are based on the International Rice Standards set by the Codex Standard 198-1995 of the Codex Alimentarius (Food Code) including the NGS (Standard for Rice, 1995). In the Philippines, the manual grading of milled rice depends on the assessors' use of measuring tools like rulers and calipers. Furthermore, grain type classification also depends on the experience of the assessor. Image processing techniques are the most common computer-aided methods used to classify and grade rice grains based from different standards all over the world. These techniques are often applied to color analysis. Thresholding techniques were used to distinguish the chalky region of a grain and ultimately quantify its region percentage (Chandra, et al., 2014). The amount of the chalky region signifies the breaking capacity of the grain and this degrades the quality of the product. A lot of studies correlate the degree of milling of the rice products to its quality. A study made in 2001 monitors the degree of milling of rice grain samples using the whiteness of the rice grains (Yadav, et al., 2001). An image of the rice grain samples is obtained using CCD Camera mounted to a platform equipped with image enhancing components. The image is analyzed by a computer running an analysis software. Several studies even use machine learning algorithms to determine the grade of the milled rice. The machine learning program learned how to distinguish between grades when fed with the training data obtained from manual methods (Neelamegam, et al., 2013).

Even though there are a lot of studies directed towards fast and affordable rice grading, the standards from where they were based are diverse and generic. They also tend to determine the grade of the rice products based on few factors. This lack of study using the NGS exposes the vulnerability of the current assessment methods of rice grain products in the Philippines. The grading procedure remains manual, highly subjective, and rely heavily to the expertise of the assessors. This has the potential to create sampling and grading variations which may affect the general Philippine rice market.

The general objective of this study is to create a milled rice grain grading system using Raspberry Pi microcomputer with image processing and adaptive boosted support vector machines. Using image processing methodologies, the study aims to develop a milled rice grading system that is portable and accessible to people and organizations who are working directly on rice like millers, distributors, and farmers. The specific objectives of this study are (1) to design a portable standalone device for milled rice grain grade classification and image acquisition; (2) to develop a milled rice grain grade classification process using grain validation, color and area analysis, and support vector machines with adaptive boosting; and (3) to evaluate the precision and accuracy of the device in classifying the grade of the milled rice grain samples.

The grade of a milled rice product affects its price. Therefore, it is essential to find a less subjective and more precise method of grading milled rice products. Fortunately, the NGS provided the data to accurately grade a product. A device that will produce precise and accurate grading based on the NGS can improve the profit of a rice business. Moreover, a device that is accessible and portable can reach even the remotest parts of the Philippines. This effectively extends the reach of the Philippine Standardization Program. In general, the potential impact

of this study is related to increasing the productivity of the business by decreasing grading time, providing a fair and accurate grading method, and increasing the accessibility of the method to bigger demographics. Moreover, the National Food Authority's goal to proliferate the compliance of the NGS could be boosted with the automation of the grading procedure. The International Rice Research Institute, which has its headquarters here in the Philippines, can expand its knowledge bank using this study.

The study is limited to the definitions of the National Grain Standards and the reference values from rice-grading assessors. Moreover, the factors that will be considered by the portable grading system are: dimensional length, degree of milling, percentage by weight of broken kernels, brewers, red kernels, immature kernels, chalky kernels, damaged kernels, presence of foreign materials, and yellow kernels. The moisture content is excluded since this measurement is dependent on the age of the paddy rice gain (palay) and the study is limited to milled rice. The study aims to develop a portable device that grade a rice sample based from an acquired image of non-overlapping rice grains in the sample. For portability, the device will be a standalone system powered by a battery. An image acquisition platform with constant lighting setup will be developed to reduce the perception variation of the images. The device is expected to display the grade of the rice sample along with the measured values of the following factors: rice gain count, dimensional length, degree of milling, percentage by weight of broken kernels, brewers, red kernels, immature kernels, chalky kernels, damaged kernels, foreign material count, degree of milling, and yellow kernels. Furthermore, the open-source Raspberry Pi will be used as the main computer of the device and the Open Source Computer Vision Library (OpenCV) with the Python interface and the other Python modules for the image processing application.

## **Chapter II**

### **REVIEW OF RELATED LITERATURE**

#### **Rice Economy**

Rice is one of the most cultivated source of carbohydrate and calorie requirements. In 2014, the global production of rice reached 490 million metric ton with 402 million of it being used as food and the remaining as feed and other purposes (FAO, 2014). In the Philippines, it is estimated that the national production reached 7.6 million metric ton in first half of 2015 with the 23.59% of the total production coming from the Central Luzon region. The average yield for the same region was 5.64 metric tons per hectare. It is also forecasted that there will be a total of 12.27% increase in national production in the first half of 2017. (Philippine Statistics Authority, 2016). Rice farming is a big contributor to the development and progress of the Philippine economy.

#### **Philippine Grains Standardization Program**

The rice market in the Philippines is one of its biggest. The reason for this is that Filipinos are one of the nationalities with rice as its staple food. Though rice is relatively bland in taste, its consumers could differentiate quality among varieties of rice. Quality can be assessed through physical, chemical, and market preferences. The chemical methods of assessing the quality of rice employ the analysis of the percentage composition and moisture content. Though this method is highly selective and accurate, its cost is not feasible for frequent evaluation. Sensory evaluation through tasting is highly subjective to the tester's ability to differentiate and 'tasting' skills. The physical method of evaluation the quality of rice is deemed to balance the trade-off between economic feasibility and precision.

The Philippine Grains Standardization Program of 2002 is a government program spearheaded by the National Food Authority to integrate recommended industrial and commercial assessment that will provide inclusive growth, uniformity, compliance, and food quality and safety standards for the labelling and quality assessment of corn and rice grains produced in the Philippines. The National Grains Standard provides the standard specifications on the quality assessment, labelling, and recommended packaging for corn and rice products. The significance of providing quality assessment specifications is mainly to classify rice products so that the appropriate prices are set fairly and justifiably based on the superiority of the products. The NGS provides grading criteria to classify the rice product into Premium or any from Grade 1 to Grade 5. The specifications for milled rice grading are provided in Table 2.1.

**Table 2.1** The National Grains Standard

PARAMETER	GRADE					
	PREMIUM	GRADE	GRADE	GRADE	GRADE	GRADE
		1	2	3	4	5
Grain Size	Very Long, Long, Medium, Short					
Degree of Milling	Over milled, Well milled	Well milled	Regular milled			
GRADE FACTORS (% by weight)	GRADE					
	PREMIUM	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5
Broken, max.  (total including brewers)	5.00	10.00	15.00	25.00	35.00	45.00
Brewers, max.	0.10	0.20	0.40	0.60	1.00	2.00
Defectives:						
Damaged kernel, max.	0.50	0.70	1.00	1.50	2.00	3.00
Discolored kernel, max.	0.50	0.70	1.00	3.00	5.00	8.00
Chalky kernel, max.	4.00	5.00	7.00	7.00	10.00	15.00



Immature kernel, max.	0.20	0.30	0.50	2.00	2.00	2.00
Contrasting type, max.	3.00	5.00	10.00	-	-	-
Red kernel, max.	1.00	2.00	4.00	5.00	5.00	7.00
Foreign matters, max.	0.025	0.10	0.15	0.17	0.20	0.25
Paddy, max. (no. per 1000 grams)	10.00	15.09	20.00	25.00	25.00	25.00
Moisture content	14.00					
Milling degree	OMR, WMR	WMR	RMR, WMR(Super), UMR(Ordinary)			

## Definitions and classification of the characteristics

The National Grains Standard defined the factors and parameters of the specifications. The following definitions are directly referenced from the NGS.

### *Grain Size*

The grain size of a particular sample is the average of the individual sizes of the grain's measured major axis length. With the specification of the National Grains Standards, only the major axis length of the grain is measured with disregard to the minor axis length. The size classifications are defined in Table 2.2.

**Table 2.2** The National Grain Standards Grain Size Classification

Grain Size	Description
Very Long	Rice with 80% or more of whole milled rice kernels having a length of 7.5mm and above.
Long	Rice with 80% or more of whole milled rice kernels having a length of 6.4 to 7.4mm.
Medium	Rice with 80% or more of whole milled rice kernels having a length of 5.5 to 6.3mm.
Short	Rice with 80% or more of the whole milled rice having a length of less than 5.5mm.

### *Degree of Milling*

The rice seed is coated with plant material called bran. The degree of milling is defined as the extent of how much bran layers and germ have been removed in the milled rice. The classifications of the degree of milling are defined in Table 2.3.

**Table 2.3** The National Grains Standard Degree of Milling Classification

<b>Degree of Milling</b>	<b>Description</b>
Regular milled	Rice kernel from which the hull, the germ, the outer bran layers and the greater part of the inner bran layers have been removed but parts of the lengthwise streaks of the bran layers shall be within the range of 20-40% of the kernels.
Well milled	Rice kernels from which the hull, the germ, the outer bran layers and the greater part of the inner bran layers have been removed, but parts of the lengthwise streaks of the layers shall be less than 20% of the kernels.
Over milled	Rice kernel from which the hull, the germ and the bran layers have been completely removed.

*Broken Kernels*

The broken kernels are described as the pieces of kernels smaller than 75% of the average length of the unbroken kernel.

**Figure 2.1** Broken kernels

### *Brewers*

The brewers are grain samples that can pass through sieves having round perforations of 1.4 mm in diameter.

### *Damaged Kernels*

The damaged kernels are those that are sprouted or distinctly damaged by insects, water, fungi, and/or any other means. Up to 40% of rice crops can be infested and damaged by the insects targeting the grains (Banlawe, et al., 2014).



**Figure 2.2** Damaged kernels

### *Discolored Kernel*

The discolored kernels are kernels that have changed their original color as a result of heating and other means. They are also known as ‘yellow kernels’ or ‘fermented kernels’.

### *Chalky Kernel*

The chalky kernels are those, whole or broken, one-half or more of which is white like the color of white chalk and is brittle upon removal of the hull for palay.



**Figure 2.3** Rice grains with varying chalkiness (Chalkiness degree, 2014)

### *Immature Kernel*

The immature kernels are those, whole or broken, which are light green and chalky with soft texture.



**Figure 2.4** Immature kernels

### *Contrasting Type*

Palay/rice kernels of different varieties other than the variety designated, wherein the size, shape, and color differ distinctly from the characteristics of kernels of the variety designated.

### *Red Kernel*

The red kernels are those that have red bran covering, wholly or partly.

### *Foreign Matter*

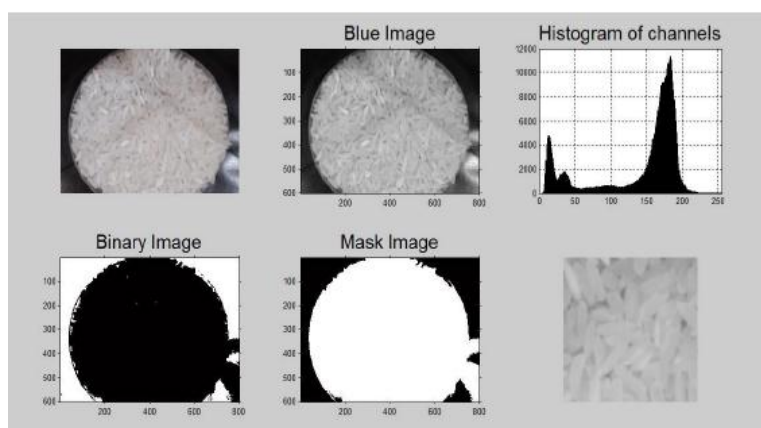
Organic and inorganic components other than whole or broken rice kernels (e.g. foreign seeds, husks, bran, sand, dust, and other crop seeds).

### *Paddy*

Paddy is the cut part of the rice plant other than the seeds.

### *Moisture Content*

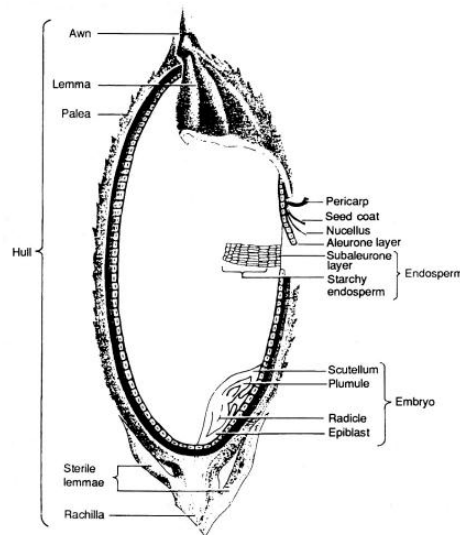
The moisture content is the water content of palay, milled rice and corn, expressed in percent (%) as received. The image data of a rice sample could indicate the moisture content. One study developed a system wherein an image of the sample is analyzed for color properties like homogeneity, intensity, and energy as seen in figure 2.4. These properties were then fed as input feature vector for a perceptron neural network. The network determines the moisture content and the results were compared to the measurements read from a standardized industry moisture content tester. A percentage difference of 2.1169% was achieved which presents the system as a viable equivalent to industry grade testers (Cruz , et al., 2017).



**Figure 2.5** Histogram features and masking (Cruz, et al., 2017)

## Morphological Indicators

The NGS is based on the physical characteristics and morphology average quantification of the rice grains. The characteristic set of each sample could indicate and be classified into grades.



**Figure 2.6** Longitudinal section of a rice grain (Belsnio, 1988)

The rice grain is the seed of a rice plant that is sexually mature. The color of a rice grain begins from being light green to progressively yellow to golden. This change in morphology could indicate the ripeness of the grain for harvest. Milled rice is rice grain where its bran is removed to an extent. The NGS defined the extent classification of milling. In Figure 2.5, the hull is the outer layer of the grain which defines the color of the grain. As the hull is removed, the bran will be left behind. The bran is a layer in the grain structure which lines the starchy endosperm of the grain.

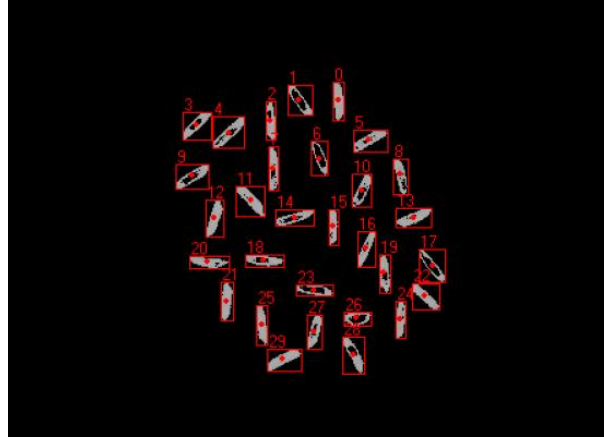
### **Grain size, broken and brewer grains**

The white part of a hulled rice grain is the starchy endosperm cells. These cells are composed of starch structures that provide sustenance for the embryo of the rice grain. The grain size is determined by the major axis length of the endosperm. Although the NGS does not provide the length to width ratio specification, it classifies rice grain size based on the major axis length. This means that the width or length of the minor axis could be disregarded for classification parameter. The grain size is significant to assess the volumetric property of the rice product.

Broken rice grains degrade the quality of the overall product. The breaking of the rice grains is caused by milling procedures or the general property and chemical quality of the rice grains. Broken rice grains are originally whole rice grains but have lengths that are less than 75% of the average length. Market preference indicates that the lesser the amount of broken grains per whole grains means the quality of the grain products is higher (Dalen, 2004). Brewers are similar to the broken grains. However, brewers are rounded. The brewer grains also decrease the perceived quality of the product. The major axis length of the rice grain is used to indicate if the sample grain is a broken, brewer, or a regular desired rice grain.

Image processing dimensional measurements are often used in the grading. In a study made in 2004, the dimensional length and width of the rice grains are used to determine if it is broken (Dalen, 2004). Broken kernels have length less than 75% of the average length. The author used image processing tools to segment, binarize, and measure the dimensions.





**Figure 2.7** Individual connected component analysis of grain threshold image with count labels (Bambole, et al., 2015)

The classification of grain size depends on the average length of the sample. The individual size must be compared to the mean length to provide a statistically significant classification of the rice. Using connected-component algorithm, threshold rice grains can be analyzed individually by labelling them. In this way, the total count of the rice grain sample can also be obtained (Bambole, et al., 2015).

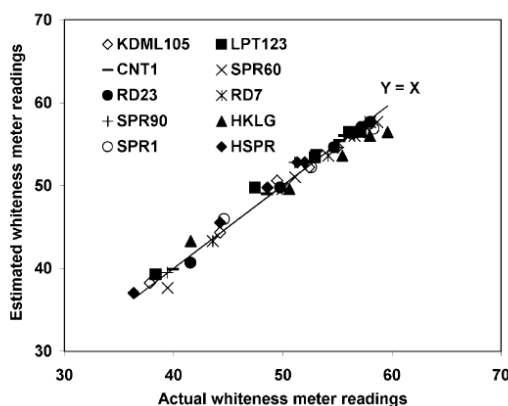
Since the connected-component labelling requires that the grains be separated or maintain a safe distance from each other, there exists a problem when the grains are overlapping each other and forming a single ‘blob’ after binarization. This cluster of pixels often makes the counting mechanism erroneous. To solve this problem, a study developed an algorithm that use edge reconstruction method to separate and distinguish individual grains with 96.53% accuracy (Alcala, et al., 2014).

### **Degree of milling**

Milling is the process of removing the husk, bran, and the embryo of the rice grain. The main purpose of milling to a degree is to manage the starch to protein content of the resulting grain. As mentioned, the starch resides in the endosperm. However, proteins and lipids are

found in the bran which is removed in the milling process. High degree of milling means the starch to protein content ratio is higher (Paiva, et al., 2014).

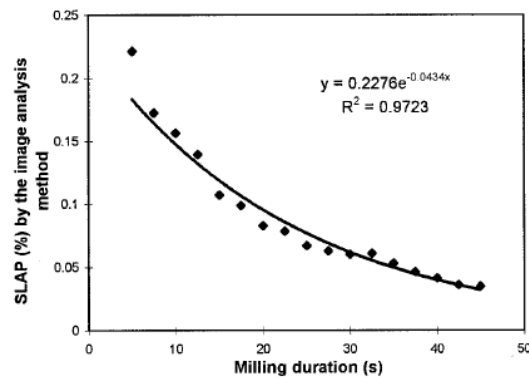
A study indicated that the degree of milling and the whiteness of the rice grains are related with each other (Yadav, et al., 2001). In this study, using image processing techniques, the head rice yield and the degree of milling are estimated. The head rice yield is the ratio of the weight of the milled rice grains to the total amount of unhusked rice grains. This ratio aims to provide a general quantification of the extent of the removal of the husks and other internal coverings. The whiteness of the rice is said to be proportional to the lipid concentration in the rice grain. High lipid concentration means the rice grain is whiter. As the degree of milling increases, the amount of whiteness in the grain intensifies. The whiteness is defined by the overall white level of the sample and not by region only. Therefore, a gray level distribution mean was used to indicate the overall whiteness. Increasing degree of milling corresponds to increase in mean gray level.



**Figure 2.8** Comparison between whiteness estimation using mean gray levels and the actual whiteness (Yadav, et al., 2001)

In another study, the surface lipid concentration on the grain was used to estimate the degree of milling and found that higher degree of milling results in lower surface lipid

concentrations (Liu, et al., 1998). A vision system is used to acquire images of the both sides of the rice grain. Applying thresholding, a threshold value of 10 is used to create a binary image. The total area of the grain is obtained based on the pixel binary values. The same method was used to get the total area of the surface lipid. A threshold value of  $10 < T < 255$  was used to normalize the pixels within the range. The total area of the surface lipid is obtained based on the pixel binary values. The T value is adjusted based on the amount of feature extraction for the surface lipid concentration. The surface lipid area percentage is the ratio between the total area surface lipid and the total area of the grain.



**Figure 2.9** The relationship between the surface lipid area percentage versus the milling duration with  $T = 130$  (Liu, et al., 1998)

By measuring the relative whiteness of the rice grain and the region percentage of the color imperfections caused by varying surface lipid, the degree of milling can be estimated.



**Figure 2.10** Detecting bran and other internal coverings using thresholding (Liu, et al., 1998)

Furthermore, low to medium degree of milling retains some bran coverings on the rice grain. Through image processing, the bran coverings' area can be compared to the whole grain area. Comparing the ratio of bran to whole area to some calibration values, the degree of milling can be estimated. Ultimately, this is similar to the method mentioned previously by measuring the amounts of imperfection on the surface of the grain. A similar method was developed in determining the degree of milling. By measuring the average whiteness of the grain sample using photodiode arrays, the degree of milling is mapped over the output voltage of the photodiode array (De Jesus, et al., 2014).

### **Damaged kernel**

External factors like the temperature, humidity, presence of pests, fungi, and rots causes the damage in the kernels of the rice. In excess humidity, unhusked rice grains could begin to sprout. This is visible even when the rice grains are milled. Also, fungal contamination like molds contribute to the decreasing quality of the rice grains. Damaged kernels can be distinguished by the visible and unnatural looking spots in the rice grain or the presence of sprout like structures.



**Figure 2.11** Collection of damaged kernels

Damaged kernels show unnatural discolorations. In a study made for classifying defects in rice grains, gray scale images were analyzed for features of discoloration (Chandra, et al., 2014). Using statistical parameters, the texture properties of the rice grain were analyzed. The gray level co-occurrence was analyzed to provide an insight to the texture of the rice grains. The damage ratio is described as the ratio of damage texture to the healthy texture. As with the detection of milling degree, the surface imperfections can be used to estimate the damage in the kernels by getting the ratio of the classified damaged region and the total area of the region.

#### **Discolored kernel**

Discolored kernels are those that have gone through the process of ‘fermentation’. Rice products that have undergone this process indicates that the products are stored for a long time. Unnecessary high storage temperature also causes discolored kernels. Fermentation in the rice can be indicated by the collective yellowness of the grains.

By estimating the mean yellow of the rice grains and then compared to calibrated values, the degree of fermentation can be obtained. Also, the yellowness tends to form a gradient on the rice grain meaning there is a low tendency for a region to be sharply yellow compared to the other regions. The yellowness tends to diffuse from the starting point of the fermentation.

#### **Chalky kernel**

Rice grains with colors that can be compared to the color of a white chalk tends to break easily (Bambole, et al., 2015). Chalky grains, when stored or milled, turns into broken grains and powdery substrates. The degree of whiteness on a region compared to the overall whiteness of the sample is used to indicate the chalkiness of the rice grain.

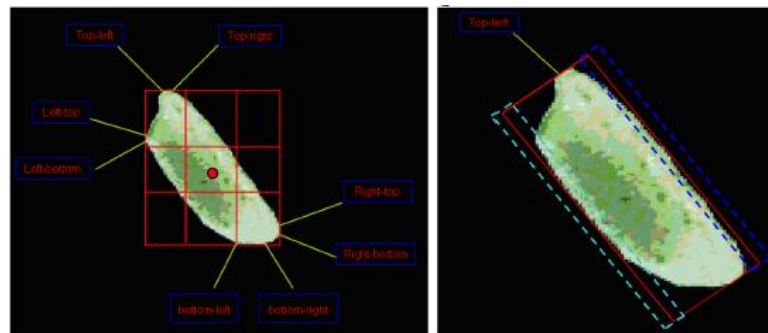
The degree of chalkiness could be processed as the measured opaqueness of the rice grains by passing light through them. However, more recent methodologies use thresholding algorithms that compare the degree of whiteness of the rice grains with the threshold value and then computing the summed ratio of the degree of whiteness (Bambole, et al., 2015). Based from the study, the chalkiness of the sample is obtained by the ratio of the number of the grains classified as chalky and the total number of the sample grains. To determine if a grain is chalky or not, a threshold was applied to the binarized image. Chalky regions are represented by black pixels. The ratio between the chalky region to the whole region represents the chalkiness of the grain. If the chalkiness is over 20%, then the grain is considered chalky.

Using thresholding, the colors can be differentiated and reduced to standard preset values. This makes the color distribution easier to analyze. The chalkiness of the rice can be measured by the ratio of the chalky white regions to the overall area of the region (Chandra, et al., 2014).



**Figure 2.12** (a) original image to apply threshold, (b) resulting threshold binary image to extract chalky regions (Chandra, et al., 2014)

In another study, a support vector machine coupled image processing method was developed to identify the chalky regions and the dimensional features of the grain. The application of linear classifiers proved to be successful with a maximum success rate of 98.5% classification. By segmenting the chalky threshold regions, the area of those regions was quantified and compared to the overall grain area region. This method is similar to the previously mentioned classification techniques (Sun, et al., 2014)



**Figure 2.13** Segmented chalky regions through thresholding (Sun, et al.,2014)

The average color of the rice grain can also be used to determine the chalkiness level. One study employed the use of photodiodes and LEDs to obtain the average ‘whiteness’ or chalkiness of a sample (De Jesus, et al., 2016). The corresponding voltage output of the

photodiode array are mapped into reference levels thus obtaining a proportionality indicator of chalkiness.

### **Immature kernel**

Immature kernels are those that are harvested before the desired maturation of the seeds. Immature kernels exhibit their young age through the color of the endosperm. As mentioned, the desired maturity of the grains is indicated by the yellow to golden color of the husk. Immature kernels have greenish endosperms. The degree of green in the endosperm indicates the immaturity of the rice kernel.



**Figure 2.14** Immature rice grains

Similar to fermented kernels, the diffusion of the color is gradient. In this case, the green color tends to diffuse from tips radiating towards the center of the grain. No visibly sharp regions are present.

### **Red kernel**

Red kernel properties are almost similar to the degree of milling. The bran is the covering of the endosperm. Red kernels still have this covering depending on the milling. Well-milled rice grains are supposed to contain less red kernels. However, for regular milled rice



grains, the red kernels are relatively abundant. Red kernels could determine the degree of milling of the rice grains.



**Figure 2.15** Degree of milling comparison (Paiva, et al., 2014)

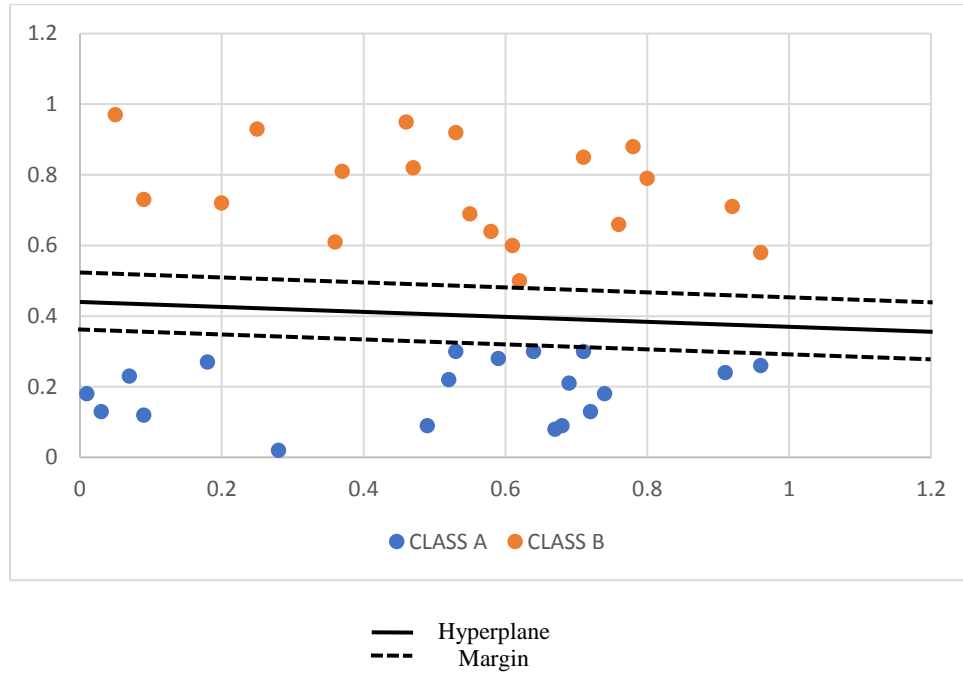
### Machine Learning Classifiers

Machine learning is the use of computing process in order to mimic or model a behavior, classification, or identity of a certain application which can include image recognition, object detection and more. One type of machine learning encompasses supervised learning methods. In a supervised learning environment, the algorithm generates a model fed with a set of training data which are usually made up of feature vector and class pairs. The feature vector is a vector space containing the quantification of the characteristics that distinguish a class. An example of a supervised machine learning algorithm is the support vector machine algorithm (SVM).

A machine learning approach using artificial neural networks (ANN) was developed in determining the variety of a rice sample. In the study, 52 varieties of Philippine rice were classified using a computer and a camera module with 85.81%-97.39% accuracy for the different categories of rice (Guzman, et al., 2008). For rice quality grading, a probabilistic neural network (PNN) was used to classify the color histograms of each rice grains in a sample with 94% accuracy (Agustin, et al., 2008).

### Support vector machine (SVM)

A support vector machine is often used as a classifier based on supervised learning. The algorithm separates or classify data by an optimal hyperplane inside a hyperspace. The position of the data when plotted against the hyperplane determines its class. The input vectors' dimensional space are increased non-linearly (Cortes, et al., 1995). The SVM's hyperplane is derived to maximize the margins on both sides of the hyperplane. For low dimensional linearly separable data with  $n$ -dimensions, the hyperplane could be described in a  $(n-1)$ -dimensional space. As illustrated in figure 2.12, the two-dimensional data set is separated with a one-dimensional hyperplane which margin sizes are optimized by the support vector machine. The margin size is the distance of the margin plane to the hyperplane. The training data used is a set described by  $(y_1, x_1) \dots (y_i, x_i)$ . Where  $y \in \{-1, 1\}$ , and  $x \in \mathbb{R}^p$  for all  $i=1, 2, \dots, u$ . The optimal hyperplane,  $w_0 \cdot z + b = 0$ , is the basis of the decision where to classify a point.



**Figure 2.16** An SVM Classifier Plot with Optimized Hyperplane

Finding the hyperplane is an optimization problem. The support vectors are the nearest vectors to the hyperplane and in which the equation 2.1 holds true. The margin size can be

described by the inequality 2.2. The optimal hyperplane is a linear combination of support vectors. The vector for the hyperplane,  $w_0$ , can be found by equation 2.3.

$$y_i(w \cdot x_i + b) = 1 \quad (2.1)$$

$$\rho < \sqrt{\frac{2}{w_0}} \quad (2.2)$$

$$w_0 = \sum_{i=1}^l y_i \alpha_i^0 x_i \quad (2.3)$$

### Adaptive boosting

Weak classifiers are often evaluated as those whose accuracy is below random chance (Freund, et al., 1997). Several algorithms are developed in order to ‘boost’ the accuracy of these classifiers. In general, this process is called boosting. The most common method of boosting is the adaptive boosting or AdaBoost . In general, the algorithm takes any classification algorithms and statistically boost the probability of the erroneous data labelled data of being able to train the model better. Each labeled data is randomly selected and given weight based on their influence on the error. A new model will be generated and the weights will be updated also based on the individual errors of each labeled data. The error of the model data is:

$$\varepsilon_t = \sum_{i=1}^N p_i^t |h_t(x_i) - y_i|$$

Where  $\varepsilon_t$  = error

$p_i^t$  = probability distribution of labelled  $i$ th data at iteration  $t$

$h_t(x_i)$  = output of the classifier given  $x_i$

$y_i$  = supposed output of the classifier given  $x_i$

Normalizing the error,

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$$

Thus, the weights of the vectors for the next iteration can be updated by:

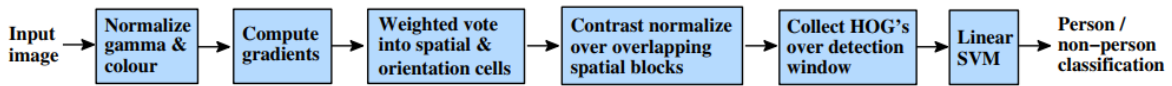
$$w_i^{t+1} = w_i^t \beta_t^{1-|h_t(x_i)-y_i|}$$

Since the AdaBoost algorithm generates many models of different accuracy, at the end of the algorithm, a convening step is needed to produce a single output of the model and is described by equation 2.4. The  $H$  is the overall hypothesis of the models generated. The ‘1’ and ‘0’ represents the classes of the weak classifier.

$$H = \begin{cases} 1 & \text{if } \sum_{t=1}^T (\log \frac{1}{\beta_t}) h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \log \frac{1}{\beta_t} \\ 0 & \text{Otherwise} \end{cases} \quad (2.4)$$

### Histogram of oriented gradients

The histogram of oriented gradients (HOG) descriptor can be used as a feature set for classification. The HOG descriptor is a histogram (distribution) of the orientation of gradients (Dalal, et al., 2005). The process begins with the calculation of gradients over a small section of an image. The image will be divided into subsections and the histogram of the gradients from the subsections will be created. The block normalization process is applied to the image by sliding. A normalized vector is the result of process which is called the HOG feature vector. These feature vectors are concatenated into a single feature vector with a high dimension and can be used as input vector to a classifier.



**Figure 2.17** The process of using HOG features as input to a Linear SVM (Dalal, et al., 2005)

## Confusion Matrix

Confusion matrix is the summarization of the performance of a classification model on a group of tested data in which the true or actual values are known. This method is frequently used in statistical classification problems, most especially in machine learning. The matrix has two-dimensions, the row is indexed by the actual class of an object, while the column is indexed by the class that the classifier predicts (Deng, et al., 2016).

		Predicted				
		$A_1$	$\dots$	$A_j$	$\dots$	$A_n$
Actual	$A_1$	$N_{11}$		$N_{1j}$		$N_{1n}$
	$\vdots$			$\vdots$		
	$A_i$	$N_{i1}$	$\dots$	$N_{ij}$	$\dots$	$N_{in}$
	$\vdots$			$\vdots$		
	$A_n$	$N_{n1}$		$N_{nj}$		$N_{nn}$

**Figure 2.18** Confusion Matrix (Deng, et al., 2016)

The table shows the basic form of confusion matrix for a multi-class classification. In the confusion matrix,  $N_{ij}$  represents the number of samples belonging to class  $A_i$  but classified as class  $A_j$ . To fully simplify and analyze the predicted values, the matrix will become two by two that classifies as true positives, false negatives, false positives and true negatives.

		Assigned Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

**Figure 2.19** Classification of positives and negatives (Sammut, et al., 2011)

### Otsu's Algorithm

One of the most used techniques in image processing is the Otsu's method. The basic idea of Otsu's method is allowing the image thresholding or the gray level image to become a binary image or separate into two groups, which can be either the foreground or background of the image. This algorithm is also flexible when it comes to combining with other filtering methods to perform segmentation effectively (Taleb, et al., 2016).

$$\sigma_W^2 = W_b \sigma_b^2 + W_f \sigma_f^2 \quad (2.5)$$

Where  $\sigma_W^2$  = within class variance

$W_b$  = weight of the background

$\sigma_b^2$  = variance of the background

$W_f$  = weight of the foreground

$\sigma_f^2$  = variance of the foreground

$$\sigma_B^2 = W_b W_f (\mu_b - \mu_f)^2 \quad (2.6)$$

Where  $\sigma_B^2$  = between class variance

$W_b$  = weight of the background

$W_f$  = weight of the foreground

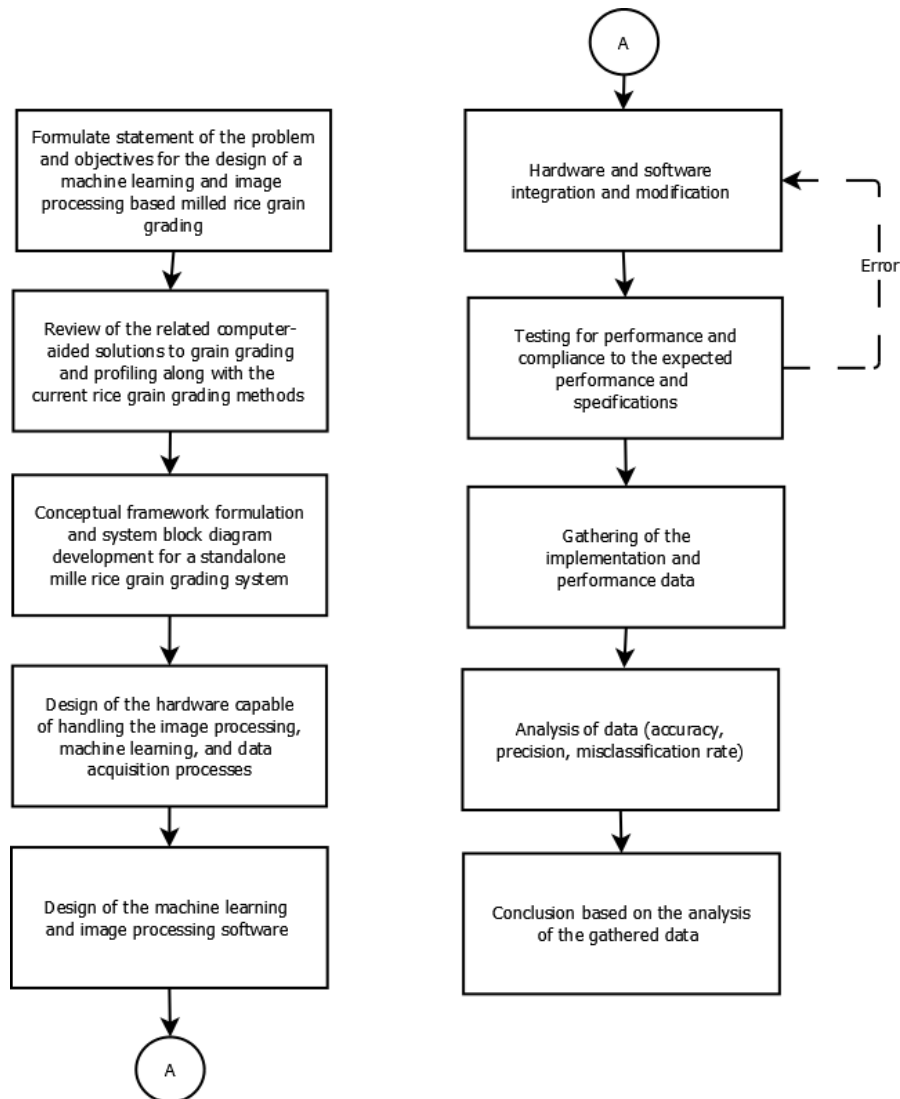
$\mu_b$  = mean of the background

$\mu_f$  = mean of the foreground

## Chapter III

### METHODOLOGY

#### Introduction



**Figure 3.1** The Quantitative Research Methodology Flow

The study will be based on a quantitative methodology as seen in figure 3.1. The goal of a quantitative research is to provide empirical evidence in the form of statistical or numeric data to prove, assess, or dispute a certain phenomenon, condition, or specification (Dewey, et

al., 2012). The study aims to evaluate the performance of the standalone device to be developed in grading milled rice grains. The study will begin by formulating the statement of the problem based on the current milled rice grain grading methods in the Philippines and other international industries. The review of literature related to current computer-aided grading methods and profiling will supplement the theoretical formulation of the structure of the conceptual framework. The objectives, scopes, and delimitations of the study will be based on the statement of the problem.

The conceptual framework defines the procedures on the fulfillment of the objectives. It is composed of three parts: (1) input, (2) process, and (3) output. For the input, the image acquisition process will provide the necessary raw data that will be processed for profiling. The process contains the image processing techniques: connected-component labelling, color analysis, image data acquisition, dimensioning, and counting; machine learning processes: support vector machines with adaptive boosting. Finally, the output is mainly the display of the grade report of the milled rice grain.

After designing the whole flow of the system to be used for the study, the hardware component development will start. In this part, the image platform and acquisition module will be developed based on the software requirements. Other development considerations include the minimization of the excessive light interference and variation for image data acquisition.

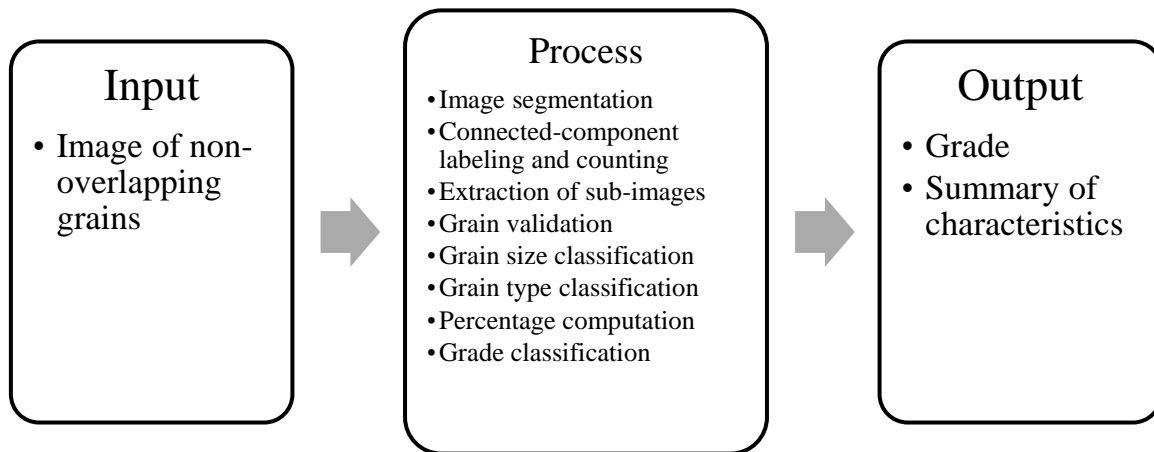
The software development consists of all the machine learning and image processing techniques to be used. The machine learning, specifically the support vector machines with adaptive boosting, components need to be trained with reference data. Reference data from the NFA-accredited establishments will be used to train the system.



The software and hardware integration will begin after training the SVMs. Pre-production testing will be done to make sure that the system is ready for production testing. If the expected performance (accuracy, precision, and misclassification rate) is not achieved, then a modification to the system will be done, accordingly. Until the individual SVMs and classifiers are performing as expected, the whole system will be tested for overall performance.

The testing consists of statistical analysis using contingency tables and confusion matrices. Specifically, the main performance indicators to be assessed are the precision, accuracy, and the misclassification rate.

A conclusion will be derived based on the statistical data analyzed. Recommendations will be given for further improvement of the process flow or technique utilization.



**Figure 3.2** The Conceptual Framework

The figure 3.2 illustrates the conceptual framework of the study. An input image of non-overlapping grains will be segmented with a threshold found using Otsu's method. The background will be subtracted based from this value. A connected-component labelling will be done along with the counting of the total individual labels in the image. The image must have more than 100 grains. Otherwise, the system outputs an error stating that the sample size is below the recommended count and requires the user to increase the sample count. These individual objects may or may not be grains (e.g. foreign materials). Bounding boxes will be overlaid on each of the labelled objects and they will be extracted into new images.

The grain validation process will be performed for every extracted image. In this process, the histogram of gradients (HOG) features will be created. The object from the image will be classified into 'grain' or 'foreign material' by the support vector machine (SVM) called SVM-VAL which is trained to determine if an input image is a 'grain' or 'foreign material' with the HOG features as the feature vector. If the object is a grain, then it will be processed

further. Otherwise, it will be classified as a ‘foreign material’, the respective counters will be incremented, and the next grain image will be processed.

After validating that the object is a grain, the grain size classification will start. In this process, the grain image’s HOG features obtained from the previous HOG determination will be used as an input feature vector for the SVM-BKN that determines if the input image is of a ‘broken’ or ‘unbroken’ grain. Respective counters will be incremented based on the classification. If the grain is ‘unbroken’, the major axis length will be recorded for later averaging process.

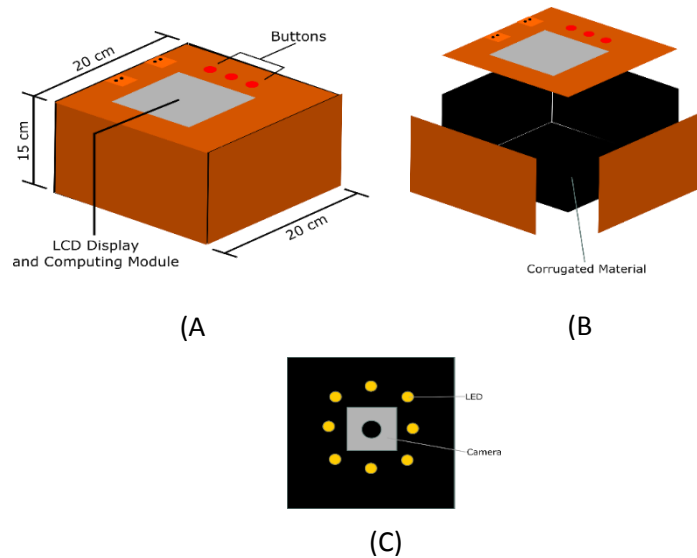
The next process is the grain type classification. In this process, different SVMs will be used: SVM-GRN, SVM-YLW, SVM-RED, SVM-DOM, SVM-PAD, and SVM-DAM. The SVM-GRN is used to classify if the grain is immature or not. Using the SVM-YLW, the grain will be classified as fermented or not. The SVM-RED will classify the grain as red kernel or not. Moreover, SVM-DOM is a multiclass classifier that determines the degree of milling of the grain. The SVM-PAD determines if the grain is a paddy or a milled rice grain. Lastly, the SVM-DAM determines if the grain is damaged. The input feature vectors for these SVMs are based on the histogram characteristics of the image. Respective counters will be incremented based on the classification of the image. Based from the average length of the unbroken grains obtained earlier, the grain will be classified into regular, broken, and brewer and the respective counters will be incremented.

The final classification of the grain is the determination if it is chalky. A thresholding process will be performed to separate the area of the chalky region from the non-chalky. The percentage area of the chalky region will be computed and based from this, the grain will be determined if it is chalky or not. Again, the respective counter will be incremented.

After all the processing, the percentage by count of the classifications will be computed. The count of the red, green, yellow, damaged, chalky, broken, brewer, paddy kernels and foreign materials will be compared to the total number of objects in the rice grain sample for percentage. The degree of milling is determined by the majority degree of the grains processed. Based from these percentages, the grade will be determined. The rice grains will be classified into PREMIUM, or GRADE 1 to GRADE 5 accordingly based from the specification of the National Grain Standards (National Grain Standards, 2004).

The output of the system is the grade of the sample along with the summary of the characteristics.

### Hardware Development



**Figure 3.3** The Proposed Hardware Device Form. **(A)** dimensions and outer components, **(B)** the inside of the box, **(C)** the underside of the top hinged door.

A proposed form to the standalone device is illustrated in figure 3.3. The standalone device is divided into two parts: (a) the image acquisition platform and (b) the image acquisition module.

### **Image acquisition platform**

The image acquisition platform is a box with insides painted black as illustrated in figure 3.3.B. The black cover is supposed to reduce the interference of excessive light reflection from the constant lighting source. In this way, there will be less image pre-processing. The color black is chosen since it is the most contrasting color to the normal color of rice grains. Furthermore, the flooring of the box is made from corrugated material (e.g. corrugated cardboard). The corrugation of the flooring aims to orient the grains into a uniform direction. This is done to maximize the potential of the image processing algorithms to approximate the length of the grain. The grains will be placed inside the platform in a non-overlapping manner.

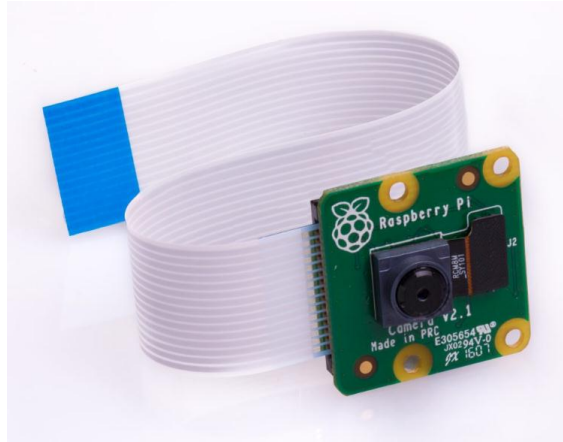
### **Image acquisition module**



**Figure 3.4** The Raspberry Pi 3 Model B

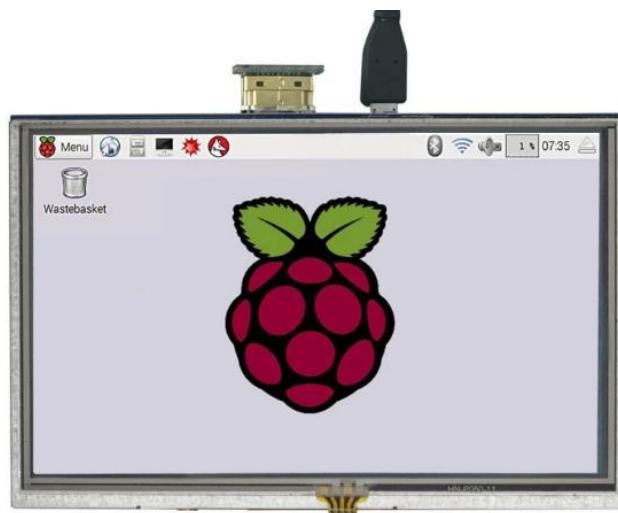
The Raspberry Pi 3 Model B will be the main computing platform of the device. It will be programmed with the classifiers and image processing algorithms. Furthermore, it will be

used for the training of the classifiers. This computer drives the LED lights along with the display of data using the LCD display. The model has a 1.2GHz 64-bit quad-core ARMv8 CPU with 1GB of RAM with camera (CSI) and display (DSI) interface. The model also supports Raspbian images.



**Figure 3.5** The Raspberry Pi Camera Module V2

A modified Raspberry Pi Camera Module V2 will be used to take the images of the grains in the sample. A 15cm ribbon cable is provided for the connection to the CSI port of the Raspberry Pi computing module. It uses the Sony IMX219 sensor with a sensor resolution of 3280x2464 pixels and a still resolution of 8 Megapixels.



**Figure 3.6** The Raspberry Pi 3.5 in LCD

The system outputs a grade report of the analyzed sample. The LCD Camera will be used to display the grade of the sample and the report about its characteristics. The LCD has a 320x480 resolution and it will be connected to the Raspberry Pi computing module.



**Figure 3.7** 5 mm White LED

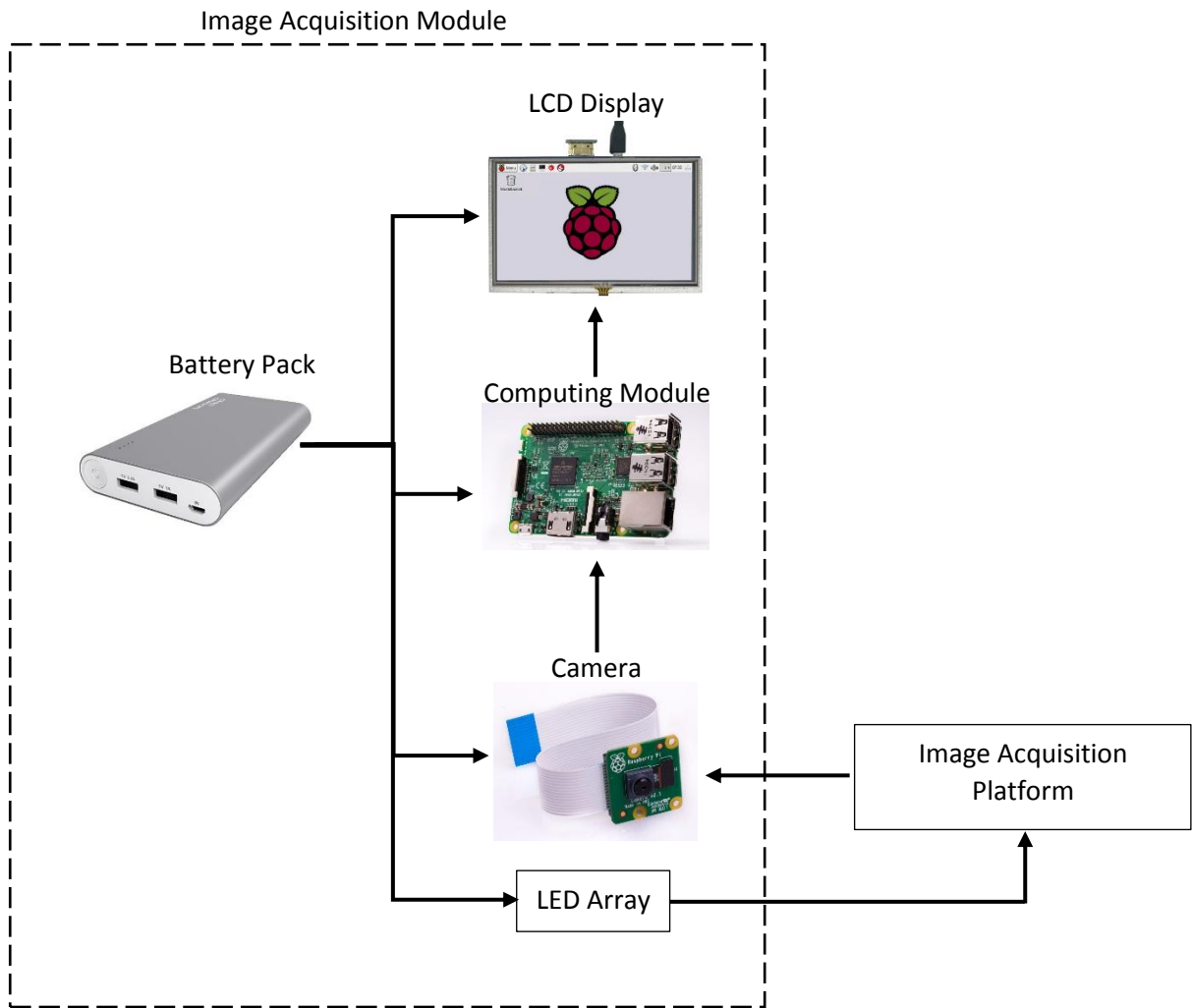
A total of eight light-emitting diodes will be placed in series to the underside of the hinged door. These LEDs will provide the constant lighting to the platform to minimize the variations of color among samples. The LEDs are also controlled by the Raspberry Pi computing module.



**Figure 3.8** The USB Power Bank

A USB power bank will be used to power the standalone device. The power bank has a capacity of 50000mAh enough to power the platform for a non-stop 24 hours, approximately. Although the device will not be on for that long, the power bank ensures that the device will work its accessibility.





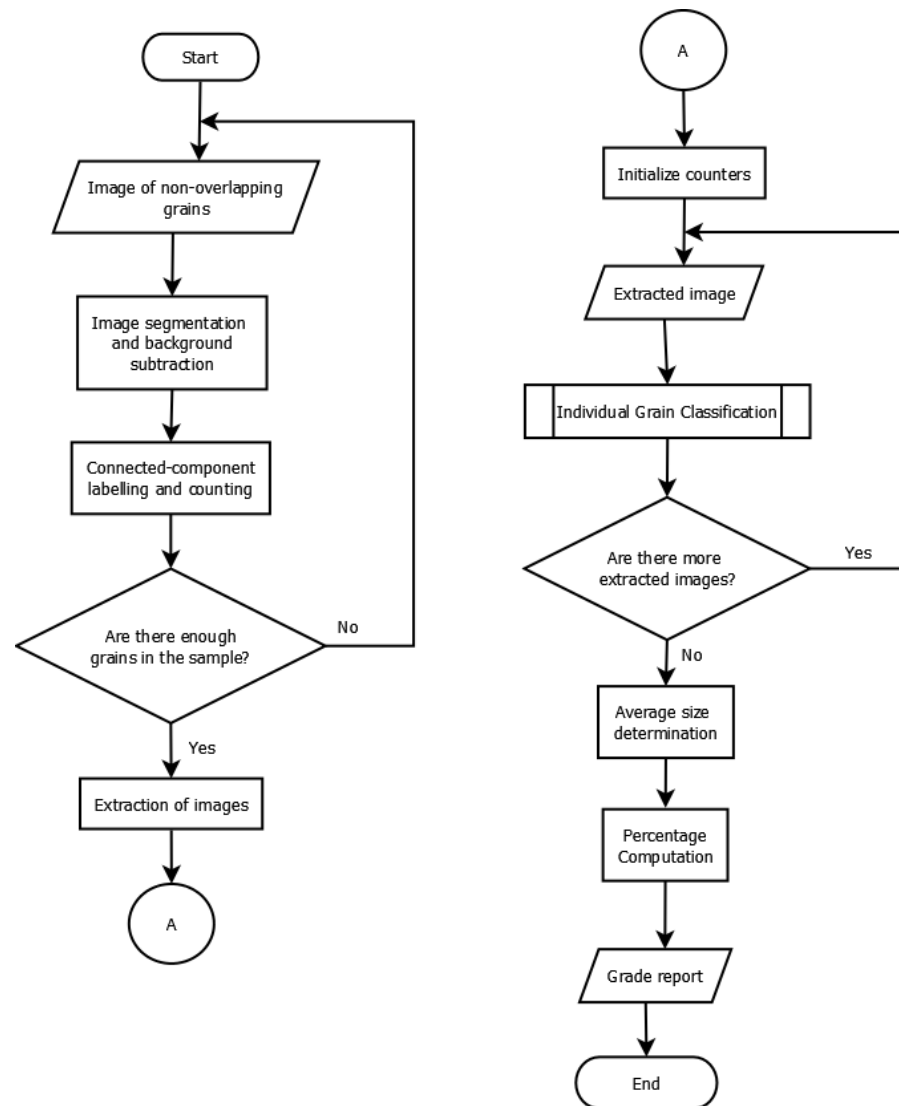
**Figure 3.9** The Block Diagram of the Hardware Setup

The figure 3.9 illustrates the workings of the standalone device. As mentioned, it will have two components: image acquisition platform and module. The user puts the grain sample inside the platform and lets the corrugation inside orient the grains accordingly. The only consideration is that the grains must not touch or overlap each other. The battery pack powers every component in the module. The camera is illuminated by the LED array and it will send images of the grains in the platform to the Raspberry Pi computing module. The computing

module classifies the grade of the sample and displays it on the LCD display along with the grade report.

### **Software Development**

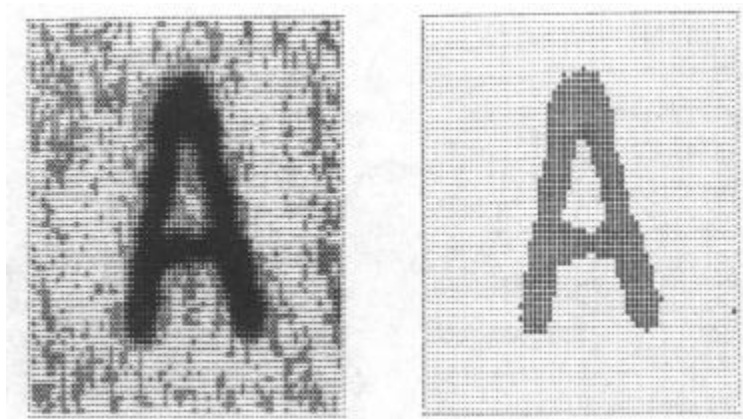
The system software will employ the image processing algorithms along with machine learning methods. Specifically, the system will use thresholding with the optimum threshold value obtained using the Otsu's method for the image processing. With the machine learning algorithms, the support vector machine (SVM) classifier boosted with the adaptive boosting algorithm (AdaBoost) will be used to classify the grains into different types.



**Figure 3.10** The System Flowchart

The figure 3.10 illustrates the system flowchart of the classification process. The input to the system is an image of non-overlapping grains. A segmentation process will be applied first to the image. The Otsu's method is used to find the optimum threshold value to maximize the separation of the foreground and the background (Otsu, 1979) as illustrated in figure 3.11. The image will be duplicated. The duplicate will be converted to gray-level and its histogram will be generated. The optimum threshold will be determined using the Otsu's method. After that, thresholding will binarize the duplicate image. The binarized image will contain the

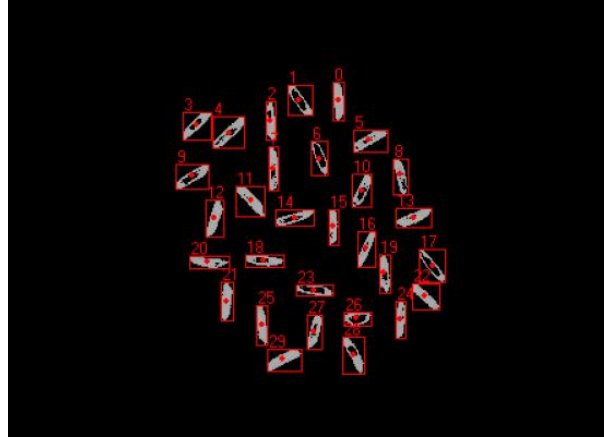
subtracted background. However, the color property was also binarized. The binarized image will be used as mask for the original image. Doing this process subtracts the background from the foreground (i.e. grains).



**Figure 3.11** Segmentation using the optimum threshold found using Otsu's method (Otsu, 1979)

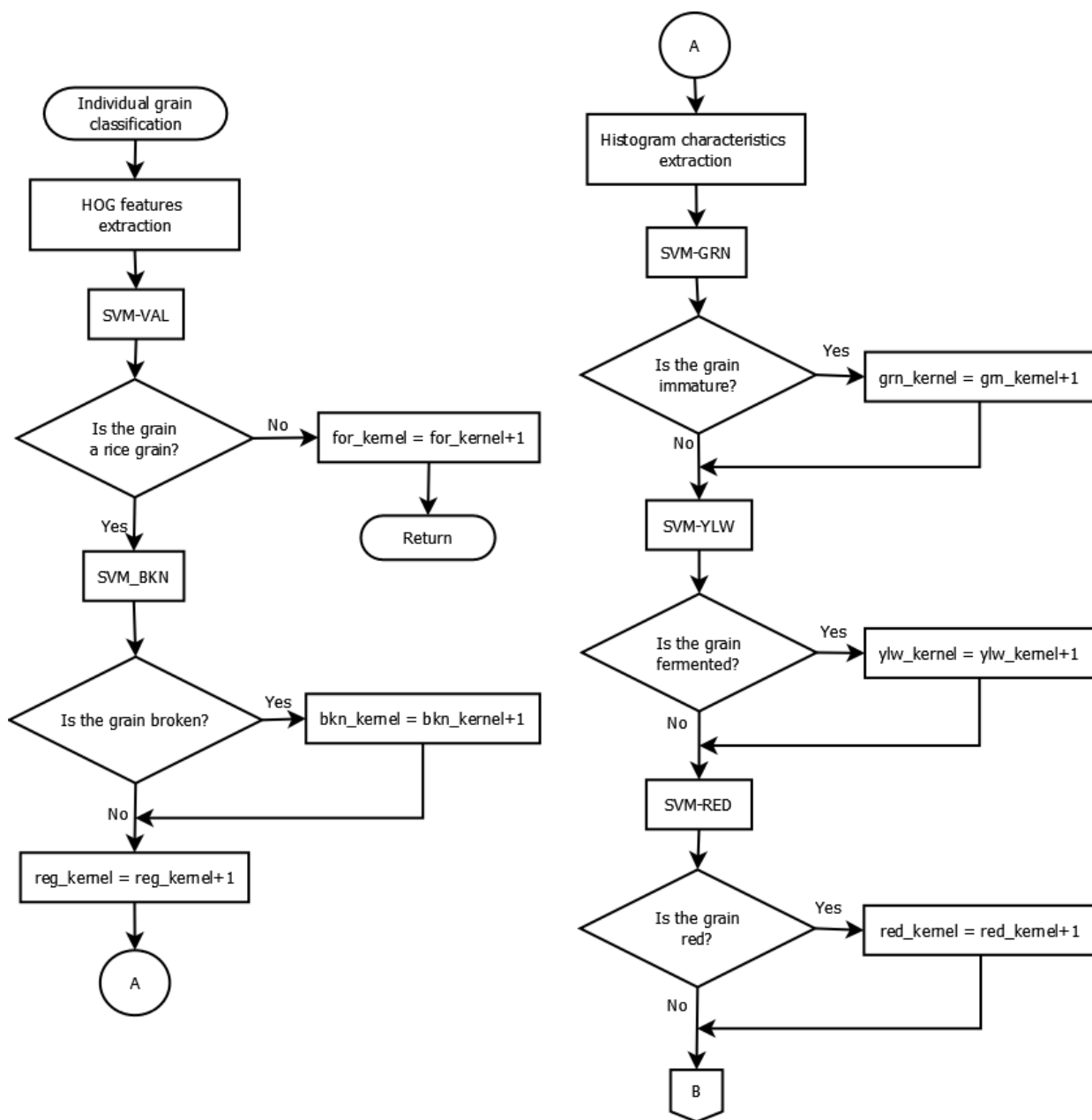
After image segmentation, a connected-component algorithm will distinguish every grain in the image. The connected-components will be counted. If the number of grains in the sample is below the requirement, then the process will not proceed. Furthermore, a minimum bounding box will be drawn around each grain. This bounding-box is used to extract images of the individual grains. The number of extracted images will be recorded along with the collection of these images.

For every extracted image of grains extracted, an individual grain classification will be performed. These classifications are recorded in counters. After all extracted images are classified, the percentage computation and average size determination will begin.

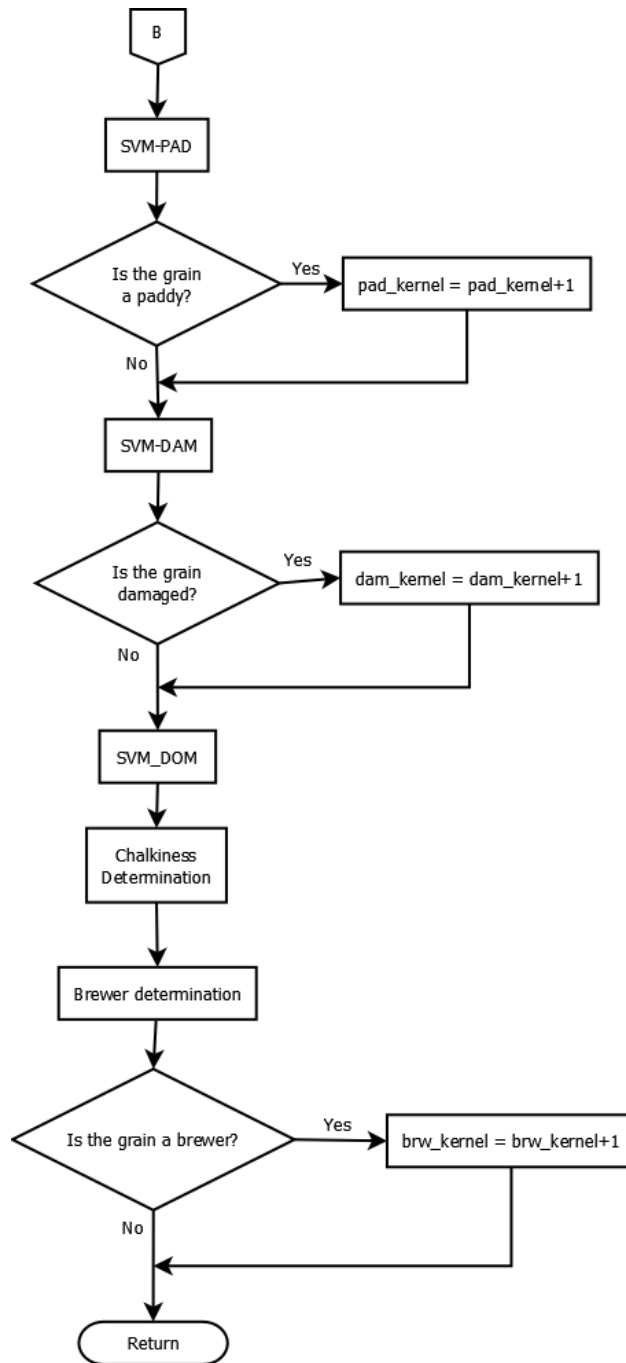


**Figure 3.12** Individual connected component analysis of grain threshold image with count labels (Bambole, et al., 2015)

The average size of the rice grains will be computed as this will classify if the sample is very long, long, medium, or short. The percentage computation compares the number of each classifications to the total number of the grains classified as milled rice. The percentage by number is equated to the percentage by weight. Using the National Grain Standards for Milled Rice, the grade (Grade 1-5 or Premium) will be determined and the grade report will be displayed.



**Figure 3.13** The Individual Grain Classification Flowchart



**Figure 3.14** The Individual Grain Classification Flowchart (continuation)

The process of individual grain classification begins by analyzing the extracted image's histogram of gradients (HOG) features. The classification process uses these features for the support vector machines (SVM) as illustrated in figures 3.13 and 3.14. The HOG features can be used as descriptor for the SVMs. By concatenation of the HOG features of the subsections

of an image, a single feature vector of higher dimensionality can be obtained (Dalal, et al., 2005). Moreover, the HOG features can be used to detect shapes and edges. This detection ability makes it possible for the SVMs to recognize the shape of a rice and foreign material grain.



**Figure 3.15** Histogram of Gradients (Dalal, et al., 2005)

The HOG features will be an input feature vector for the SVM-VAL. The SVM-VAL is a support vector machine classifier that will be trained to recognize if the grain in the image is a rice grain or a foreign material. A support vector machine is a type of supervised learning algorithm which separates two classes using a hyperplane. In a high-dimension space, the optimal hyperplane is the one that maximizes two margins on both of its sides (Cortes, et al., 1995). The nearest vectors to the optimal hyperplane are called support vectors. In this study, several SVMs will be trained to classify the grains in extracted images.

Although SVMs can be used as multiclass classifiers, in this study, one SVM will be trained for each types of grains: immature, yellow, damaged, broken, fermented, and red. Other trained SVMs will include those for determining the degree of milling and the validation if the grain is not a foreign material. These classification occurrences can overlap with each other.



For example, an immature kernel may be damaged as well as the broken. Therefore, each classification warrants their own SVMs.

If the SVM-VAL detects a foreign material, an update to the foreign material counter will be done and will immediately proceed to the processing of the next extracted image. Otherwise, the HOG features of the current image will be the input for the next SVMs. The table 3.1 summarizes the SVMs included in the study along with their description and classes.

**Table 3.1** The Support Vector Machines

Support Vector Machine	Description	Classes
SVM-VAL	Determines if the grain in the extracted image is a rice grain or a foreign material.	<ul style="list-style-type: none"> <li>• grain (rice grain),</li> <li>• for_mat (foreign material)</li> </ul>
SVM-BKN	Determines if the grain in the extracted image is a broken or a regular grain.	<ul style="list-style-type: none"> <li>• bkn (broken)</li> <li>• unbkn (unbroken)</li> </ul>
SVM-GRN	Determines if the grain in the extracted image is immature.	<ul style="list-style-type: none"> <li>• grn (immature)</li> <li>• n_grn (mature)</li> </ul>
SVM-YLW	Determines if the grain in the extracted image is fermented.	<ul style="list-style-type: none"> <li>• ylw (fermented)</li> <li>• n_ylw (regular)</li> </ul>
SVM-RED	Determines if the grain in the extracted image is a red kernel.	<ul style="list-style-type: none"> <li>• red (red)</li> <li>• n_red (regular)</li> </ul>
SVM-PAD	Determines if the grain in the extracted image is a paddy.	<ul style="list-style-type: none"> <li>• pad (paddy)</li> <li>• n_pad (regular)</li> </ul>

SVM-DOM	Determines the degree of milling of the grain in the extracted image.	<ul style="list-style-type: none"> <li>• reg (regular milled)</li> <li>• well (well milled)</li> <li>• ovm (over milled)</li> </ul>
---------	---	---

The HOG features will be used as input vectors for the SVM-VAL and the SVM-BKN. For other SVMs, another set of input vectors will be generated and used instead of HOG features. The color histogram of the extracted image will be constructed: R, G, B, and Gray histograms. A feature vector will be created by using the characteristics of these histograms as summarized by table 3.2.

**Table 3.2** Color Histogram Characteristics

Histogram	Characteristic	Description
R, G, B, and Gray	Standard Deviation	<p>The standard deviation is the measure of the spread of the level pixels. For an <math>L</math>-level histogram, the SD is:</p> $\sigma = \sqrt{\sum_{i=0}^{L-1} (I(i) - \bar{m})^2}$ <p>Where <math>\sigma = \text{SD}</math>, <math>I(i)</math> = number of pixels at level <math>i</math>, <math>\bar{m}</math> = mean number of pixels in levels.</p>
R, G, B, and Gray	Skew	<p>The skew is the measure of the asymmetry of the data around its mean (Sergyan, 2008). For an <math>L</math>-level histogram, this is given by:</p>

		$\gamma = \frac{1}{\sigma^3} \sum_{i=0}^{L-1} (I(i) - \bar{m})^3$ <p>Where <math>\gamma</math> = skew, <math>I(i)</math> = number of pixel values at level <math>i</math>, and <math>\bar{m}</math> = mean number of pixels in levels.</p>
R, G, B, and Gray	Mean	<p>The average value of the histogram. For an <math>L</math>-level histogram, the mean is:</p> $\bar{m} = \frac{\sum_{i=0}^{L-1} I(i)}{M}$ <p>Where <math>\bar{m}</math> = mean, <math>I(i)</math> = number of pixels at level <math>I</math>, and <math>M</math> = total number of pixels.</p>
R, G, and B	Dominant	<p>The dominant level is the level of the histogram with the highest number of pixels.</p>

The chalkiness determination requires other image processing technique. The NGS defined the percentage region of the chalky part of the grain. Based from this percentage, the grain can be classified as chalky or not. The Otsu's method will be used to find the optimum threshold which separates the chalky region from the regular. A thresholding process will be done based on the Otsu's threshold. The number of pixels considered as chalky will be compared with the overall pixels of the grain as summarized by equation 3.1.

$$\%Chalkiness = \frac{C}{K} \times 100\% \quad (3.1)$$

Where  $C$  = number of pixels considered as chalky

$K$  = total number of pixels in the extracted image



**Figure 3.16** (a) original image to apply threshold, (b) resulting threshold binary image to extract chalky regions (Chandra, et al., 2014)

For every extracted image of rice grains, the length will be obtained using image processing approximation. The average major axis length of the unbroken grains will be computed. Based from this value, the brewer count will be performed. The NGS provided the size measurement classifications for these types of grains.

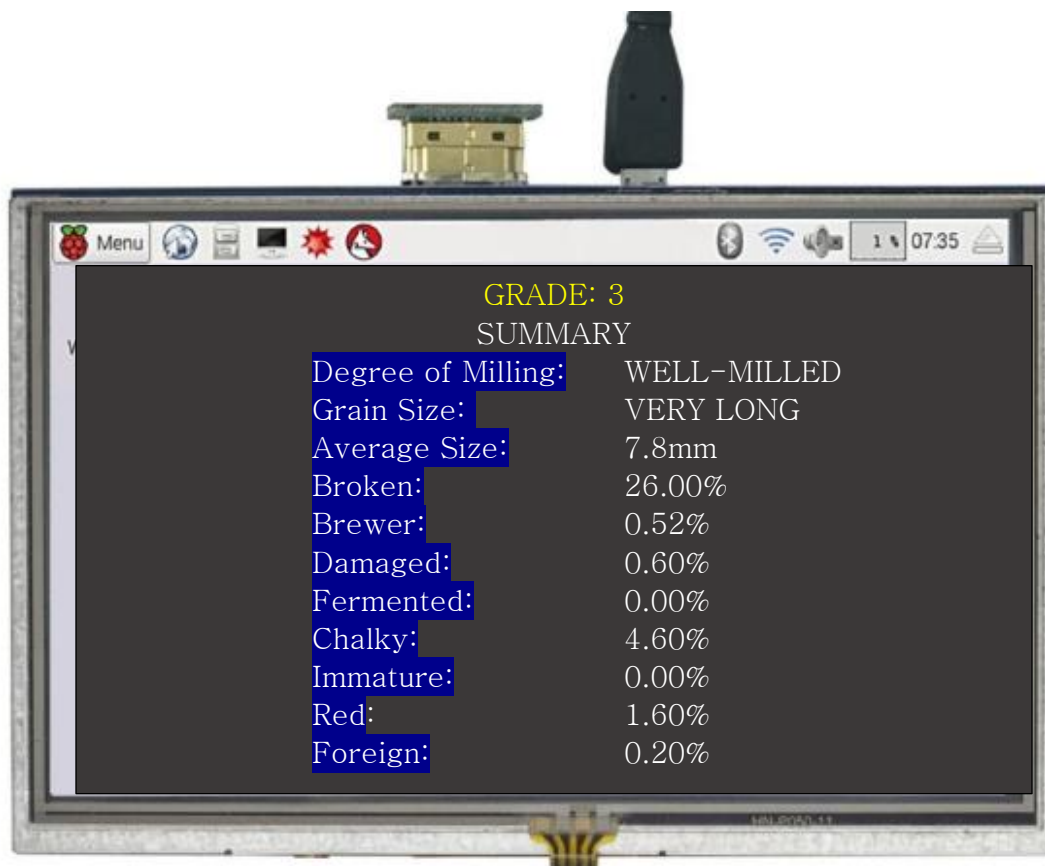
After all the extracted images are classified, the percentage computation will be performed. In general, the percentage computation is the process of dividing the number of a certain grain classification to the total number of grains. This equates to percentage by weight as an approximation and a sampling by convenience method. The NGS summarizes the grade requirements in table 3.3. Based from these, the grade will be determined.

**Table 3.3** The National Grains Standards for Milled Rice

PARAMETER	GRADE					
	PREMIUM	GRADE	GRADE	GRADE	GRADE	GRADE
		1	2	3	4	5
Grain Size	Very Long, Long, Medium, Short					
Degree of Milling	Over milled, Well milled	Well milled	Regular milled			
GRADE FACTORS (% by weight)	GRADE					
	PREMIUM	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5
Broken, max.  (total including brewers)	5.00	10.00	15.00	25.00	35.00	45.00
Brewers, max.	0.10	0.20	0.40	0.60	1.00	2.00
Defectives:						
Damaged kernel, max.	0.50	0.70	1.00	1.50	2.00	3.00
Discolored kernel, max.	0.50	0.70	1.00	3.00	5.00	8.00
Chalky kernel, max.	4.00	5.00	7.00	7.00	10.00	15.00

Immature kernel, max.	0.20	0.30	0.50	2.00	2.00	2.00
Contrasting type, max.	3.00	5.00	10.00	-	-	-
Red kernel, max.	1.00	2.00	4.00	5.00	5.00	7.00
Foreign matters, max.	0.025	0.10	0.15	0.17	0.20	0.25
Paddy, max. (no. per 1000 grams)	10.00	15.09	20.00	25.00	25.00	25.00
Moisture content	14.00					
Milling degree	OMR, WMR	WMR	RMR, WMR(Super), UMR(Ordinary)			

The grade report will be displayed on the interface of the device. The sample grade report is illustrated in figure 3.17.



**Figure 3.17** The Grade Report

### **Experimental Setup and Procedure:**

The device will be a standalone and portable platform that is powered by a battery. The following lists the procedures to be done for the grading process:

1. The user picks/scoops up sample grains from the batch being graded and puts it inside the device's platform. He/she must ensure that the grains are placed along the orientation of the corrugation and that no grains must be touching or overlapping another.
2. The user turns on the machine and let it boot the grading software and the other processes.

3. The user presses the 'START' button. The device will perform the image process and checks the count of the grains inside the sample. If the sample grain count is below the required number, then the device will prompt the user to increase the grains. Otherwise, the grading process will begin.
4. After the proper count has been obtained, the grading process will start.
5. The user will wait for the grade report. Upon completion of the classification process, the grade report will be displayed.
6. The user has a choice to turn the machine off or doing another grading process by pressing the 'RESTART' button and repeating Steps 1 to 6.

### **Training**

The SVMs used in this study are boosted using adaptive boosting (AdaBoost). The AdaBoost is a boosting algorithm that 'boosts' weak classifiers into a strong enough classifier by adjusting the weights of the training data. In general, the algorithm takes any classification algorithms and statistically boost the probability of the misclassified labelled data of being able to train the model better (Freund, et al., 1997). Each labeled data is randomly selected and given weight based on their influence on the error. A new model will be generated after training iterations and the weights will be updated also based on the individual errors of each labeled data. The error of the model data is:

$$\varepsilon_t = \sum_{i=1}^N p_i^t |h_t(x_i) - y_i|$$

Where  $\varepsilon_t$  = error

$p_i^t$  = probability distribution of labelled  $i$ th data at iteration  $t$

$h_t(x_i)$  = output of the classifier given  $x_i$



$y_i$  = supposed output of the classifier given  $x_i$

Normalizing the error,

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$$

Thus, the weights of the vectors for the next iteration can be updated by:

$$w_i^{t+1} = w_i^t \beta_t^{1-|h_t(x_i)-y_i|}$$

Since the AdaBoost algorithm generates many models of different accuracy, at the end of the algorithm, a convening step is needed to produce a single output of the model and is described by equation 3.2. The  $H$  is the overall hypothesis of the models generated. The ‘1’ and ‘0’ represents the classes of the weak classifier.

$$H = \begin{cases} 1 & \text{if } \sum_{t=1}^T (\log \frac{1}{\beta_t}) h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \log \frac{1}{\beta_t} \\ 0 & \text{Otherwise} \end{cases} \quad (3.2)$$

All of the SVMs, except the SVM-VAL and SVM-BKN, in this study will be trained with the help of the AdaBoost. Reference images are chosen based on the NGS.

There are two groups of SVMs to be trained and tested in the study: (A) SVM-VAL and SVM-BKN, and (B) All other SVMs. The group (A) needs to be trained using the HOG features and the group (B) will be trained using other histogram characteristics. At the end of each training, the SVMs will be tested for precision, accuracy, specificity, and sensitivity. The summarized results are found in tables 3.4 to 3.10. The evaluation of the confusion matrices is found in table 3.13. A sample training set for the group (B) SVMs is found in table 3.4. The grains of interest to be used as training reference include: white, brown, red, and black rice grains.

**Table 3.4** The Training Set Format for Group (B) SVMs

<b>Data</b>	<b>INPUT</b>				<b>CLASS/OUTPUT</b>
	<b>Standard Deviation (R,G,B,Gr)</b>	<b>Mean (R,G,B,Gr)</b>	<b>Skew (R,G,B,Gr)</b>	<b>Dominant (R,G,B)</b>	
<b>1</b>					
<b>2</b>					
<b>3</b>					
<b>4</b>					
<b>...</b>					
<b>1500</b>					

**Table 3.5** The Confusion Matrix for SVM-VAL

		<b>Predicted classification of grain</b>		<b>Total</b>
		<b>Rice Grain</b>	<b>Foreign Material</b>	
<b>Actual classification of grain</b>	<b>Rice Grain</b>			
	<b>Foreign Material</b>			
	<b>Total</b>			

**Table 3.6** The Confusion Matrix for SVM-BKN

		<b>Predicted classification of grain</b>		<b>Total</b>
		<b>Broken</b>	<b>Regular</b>	
	<b>Broken</b>			

Actual classification of grain	Regular			
	Total			

**Table 3.7** The Confusion Matrix for SVM-GRN

		Predicted classification of grain		Total
		Immature	Regular	
Actual classification of grain	Immature			
	Regular			
	Total			

**Table 3.8** The Confusion Matrix for SVM-YLW

		Predicted classification of grain		Total
		Fermented	Regular	
Actual classification of grain	Fermented			
	Regular			
	Total			

**Table 3.9** The Confusion Matrix for SVM-RED

		Predicted classification of grain		Total
		Red	Regular	
Actual classification of grain	Red			
	Regular			
	Total			

**Table 3.10** The Confusion Matrix for SVM-PAD

	Predicted classification of grain
--	-----------------------------------

		<b>Paddy</b>	<b>Regular</b>	<b>Total</b>
<b>Actual classification of grain</b>	<b>Paddy</b>			
	<b>Regular</b>			
	<b>Total</b>			

**Table 3.11** The Contingency Table for SVM-DOM

		<b>Predicted classification of grain</b>			
		<b>Well Milled</b>	<b>Regular Milled</b>	<b>Over Milled</b>	<b>Total</b>
<b>Actual classification of grain</b>	<b>Well Milled</b>				
	<b>Regular Milled</b>				
	<b>Over Milled</b>				
	<b>Total</b>				

The chalkiness determination process will also be tested accordingly. A confusion matrix will be constructed based from the prediction of the chalky grain classifier and is illustrated at table 3.11. The average size predicted by the device will also be tested using a confusion matrix at table 3.12. The evaluation of the classification is also summarized at table 3.13.

**Table 3.12** The Confusion Matrix for Chalkiness Classification

		Predicted classification of grain		Total
		Chalky	Regular	
Actual classification of grain	Chalky			
	Regular			
	Total			

**Table 3.13** The Confusion Matrix for the Average Size Classification

		Predicted classification of grain				Total
		Short	Medium	Long	Very Long	
Actual classification of grain	Short					
	Medium					
	Long					
	Very Long					
	Total					

**Table 3.14** Evaluation of The Trained Support Vector Machines and Other Processes

Classifier	Accuracy	Precision	Misclassification Rate
SVM-VAL			
SVM-BKN			
SVM-GRN			
SVM-YLW			
SVM-RED			

<b>SVM-PAD</b>			
<b>SVM-DOM</b>			
<b>Chalky Classifier</b>			
<b>Average Size Classifier</b>			

### System Testing

The system, as a whole, will also be tested. The grading process depends on how well each classifier perform and work as a whole. Test sample of known grade will be graded again using the device. A confusion matrix will be constructed for the testing process and is illustrated in table 3.14. The grades are PREMIUM and GRADE 1-5.

**Table 3.15** The Contingency Table for the Grading Process

		Predicted classification of sample						Total
		Premium	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	
Actual classification of sample	Premium							
	Grade 1							
	Grade 2							
	Grade 3							
	Grade 4							
	Grade 5							
Total								

The confusion matrix in table 3.14 will be evaluated. The accuracy and the misclassification rates will be computed. The equation 3.3 describes the accuracy of the system. The accuracy of the system is basically the measure of how the system correctly classifies the grades of the samples (classifying to the true class and classifying as not in the certain class).

$$ACCURACY = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.3)$$

The *TP* (True Positive) is the count of the correct classification of the grade of the sample. The *TN* (True Negative), on the other hand, is the count of the correct classification that the sample does not have a certain grade. Misclassifications are also computed. The *FP* (False Positive) is the count of the misclassifications of the predicted grade of the sample when it actually belongs to another grade. Lastly, the *FN* (False Negative) is the count of the misclassifications when the grade of the sample was predicted to not have the certain grade but it actually does.

The precision (positive predictive value) of the system's performance is the measure of how correct is the system when it predicted a certain grade. This property is defined by equation 3.4.

$$PRECISION = \frac{TP}{TP+FP} \quad (3.4)$$

The misclassification rate is the measure of how wrong is the system when classifying grades. Equation 3.5 describes the misclassification rate as a function of the accuracy of the system.

$$MR = 1 - ACCURACY \quad (3.5)$$

## References

- Bambole, J. D. and Ghodinde, K. A. (2015), "Evaluation of chalkiness in basmati rice by virtual instrumentation," *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai, 2015*, pp. 604-608.
- Alcala, N. L. D., Dela, L. J. C., Lazaro, T. P. F. and Llanto, J. M (2014). MATLAB Toolbox Creation for Rice Kernel Segmentation and Characterization Including Touching Kernels Separation by Edge Reconstruction Method. *Mapua University*.
- Banlawe, I. A. P., Garcia, R. G., and Sariego, P. T., (2014). Aural and photonic spectrum based digital pest controller for Oryza Sativa (rice). *2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Palawan*, pp. 1-6
- Belsnio, B. (1988). The anatomy and physical properties of the rice grain. *Towards Integrated Commodity and Pest Management in Grain Storage. FAO, Rome*.
- Chandra, J. K., Barman A. and Ghosh A. (2014), "Classification of defects in rice kernels by using image processing techniques," *2014 First International Conference on Automation, Control, Energy and Systems (ACES), Hooghy*, pp. 1-5.
- Concepcion J. C. T., Ouk M., Zhao D., and Fitzgerald M. A. (2015). The need for new tools and investment to improve the accuracy of selecting for grain quality in rice. *Field Crops Research*. Volume 182, Pages 60-67.
- Cortes C. and Vapnik V. (1995), "Support-Vector Networks," *Kluwer Academic Publishers, Boston*, pp. 273-297.



Cruz, F.R.G., Padilla, D.A., Hortinela, C.C., Bucog, K.C., Sarto, M.C., Sia, N.S.A., Chung, W.-Y, (2017). Cloud-based application for rice moisture content measurement using image processing technique and perceptron neural network (2017) *Proceedings of SPIE - The International Society for Optical Engineering*, 10225, art. no. 102251W.

Dalal N. and Triggs B. (2005), "Histograms of oriented gradients for human detection," 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, pp. 886-893 vol. 1.

Dalen, G.V. (2004), Determination of the size distribution and percentage of broken kernels of rice using flatbed scanning and image analysis, *Food Research International*, Volume 37, Issue 1, Pages 51-58.

De Jesus, M. L. B., Mayores, A. J. C., and Recuelo, M. A., (2016). Measurement of Grain Shape, Whiteness, Chalkiness and Moisture Content of Dinorado and Sinandomeng Rice Samples using a Raspberry-Pi Based Portable Rice Grain Sampler. *Mapua University*.

Deng, X., Liu, Q., Deng, Y., and Mahadevan, S. (2016), "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Information Sciences*, Volume 340, Pages 250-261.

Dewey, A., Drahota, A., Fogg, C., Kalra, P., Kilburn, S., Markham, C., and Stores, R. (2012). Designing a cardiology research project: research questions, study designs and practical considerations. *Journal of Clinical and Preventive Cardiology*, Volume 2, Issue 1, pp. 37-47.

Food and Agriculture Organization of the United Nations (FAO), (2014). *Rice Market Monitor*, Volume XVII Issue No. 3, Pages 1-35.

Freund, Y. and Schapire, R.E. (1997), “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, *Journal of Computer and System Sciences*, Volume 55, Issue 1, Pages 119-139.

Guzman, J. D., & Peralta, E. K. (2008). Classification of Philippine rice grains using machine vision and artificial neural networks. *World conference on agricultural information and IT*, Tokyo, Japan (pp. 24-27).

[http://download.cassandralab.com/ricequality/ricequality\\_help/UNIMI.Cassandra.RiceQuality.html?Chalkinessdegree.html](http://download.cassandralab.com/ricequality/ricequality_help/UNIMI.Cassandra.RiceQuality.html?Chalkinessdegree.html) (2014), Chalkiness degree.

<http://ucanr.edu/blogs/blogcore/postdetail.cfm?postnum=11457> (2013), Image from UC Rice Blog California Rice Production.

<http://unicom-egypt.com/broken-rice/>, Image from Unicom Investment and Development.

<http://www.arkansas-crops.com/2012/12/11/factors-affecting-rice-milling-quality/> (2012), Factors Affecting Rice Milling Quality.

[http://www.buhlergroup.com/static/southamerica/es/media/02-Process\\_Technologies/Ricesorting\\_1121ricesteamed\\_reject\\_004.jpg](http://www.buhlergroup.com/static/southamerica/es/media/02-Process_Technologies/Ricesorting_1121ricesteamed_reject_004.jpg) (2017), Image from Buhler A.G.

[http://www.knowledgebank.irri.org/index.php?option=com\\_zoo&view=item&layout=item&Itemid=1025s](http://www.knowledgebank.irri.org/index.php?option=com_zoo&view=item&layout=item&Itemid=1025s), Grain moisture content and grain quality.

Juliano B. O., (2016). Rice: Overview, *Encyclopedia of Food Grains (Second Edition)*. Academic Press, Oxford, pp. 125-129.

Liu Guang-rong (2010), "Rice color inspection based on image processing technique," *2010 International Conference on Advances in Energy Engineering*, Beijing, pp. 134-137.

National Food Authority, (2001). Primer on Philippine Grains Standardization Program (Revised Edition). <http://nfa.gov.ph/files/archive/pgspnew.pdf>

Neelamegam, P., Abirami, S., Vishnu Priya K., and Valantina, S.R. (2013), "Analysis of rice granules using image processing and neural network," *2013 IEEE Conference on Information & Communication Technologies*, JeJu Island, pp. 879-884.

O. C. Agustin and B. J. Oh, (2008). Automatic Milled Rice Quality Analysis. *2008 Second International Conference on Future Generation Communication and Networking*, Hainan Island, pp. 112-115.

Otsu, N. (1979), "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66.

Paiva, F.F., Vanier, N.L., Berrios, J.D., Pan, J., Villanova, F.A., Takeoka, G., Elias, M.C. (2014), Physicochemical and nutritional properties of pigmented rice subjected to different degrees of milling, *Journal of Food Composition and Analysis*, Volume 35, Issue 1, Pages 10-17.

Philippine Statistics Authority (2016). Philippines - Palay Production Survey 2016. <http://psa.gov.ph/psada/index.php/catalog/161>

Sammur C., and Webb G. (2011), Encyclopedia of Machine Learning, *Springer Science & Business Media*.

Sergyan, S. (2008), "Color histogram features based image classification in content-based image retrieval systems," *2008 6th International Symposium on Applied Machine Intelligence and Informatics*, Herlany, pp. 221-224.

Standard for Rice (1995), Codex Standard 198-1995, pp. 1-6.

Sun, C., Liu, T., Ji, C., Jiang, M., Tian, T., Guo, D., Wang, L., Chen, Y., Liang, X. (2014), Evaluation and analysis the chalkiness of connected rice kernels based on image processing technology and support vector machine, *Journal of Cereal Science*, Volume 60, Issue 2, Pages 426-432.

Talab, A., Huang, Z., Xi, F., and HaiMing, L. (2016), Detection crack in image using Otsu method and multiple filtering in image processing techniques, *Optik - International Journal for Light and Electron Optics*, Volume 127, Issue 3, Pages 1030-1033.

United States Standards for Rice (2009), United States Department of Agriculture Grain Inspection, Packers and Stockyards Administration Federal Grain Inspection Service, Pages 1-28.

Yadav, B.K., Jindal, V.K. (2001), Monitoring milling quality of rice by image analysis, *Computers and Electronics in Agriculture*, Volume 33, Issue 1, Pages 19-33.