

University of The Philippines Los Baños

College of Arts and Sciences

Institute of Computer Science

CMSC 190

Special Problem

**Peanut Mold Detection Via Marker-based Watershed Segmentation and An Artificial
Neural Network With Feedforward Backpropagation**

Samuel Darwin D. Lagrosas

Proponent

Prof. Lei Kristoffer R. Lactuan

Adviser

Abstract

Peanut seeds are susceptible to mold infestation such as *Aspergillus*, which causes the release of aflatoxins, a carcinogenic substance that can harm human and animal health. The aim of this study is to develop a method for detecting peanut mold using computer vision and artificial intelligence. Marker-based watershed segmentation was used to extract the peanut seeds from captured images for color and texture analysis. Color features are extracted from RGB, HSV, and Grayscale color spaces, while texture features are obtained using Gray Level Co-occurrence Matrix properties. Selected relevant features were fed into a feedforward backpropagation neural network, which generates values ranging from 0 to 1, with values closer to 1 indicating mold contamination and values near 0 indicating non-contamination. In detecting mold infestation in peanut seeds, the neural network produced an accuracy rate of 89.33%.

1. Introduction

Peanuts (*Arachis hypogaea*) are an annual leguminous crop found in tropic regions worldwide, including the Philippines. With 47 million metric tons produced globally in 2020 [1], peanuts are considered a major field legume worldwide. Besides being inexpensive and high in protein, peanuts are a healthy alternative for livestock, such as pigs and cows [2].

Agricultural crops like peanuts, corn, cottonseed, and tree nuts can be contaminated with aflatoxins produced by a particular type of fungi. Aflatoxin-producing fungi, such as *Aspergillus flavus* and *Aspergillus parasiticus*, thrive in warm, humid climates [3]. Balendres et al. identified *Aspergillus* as one of the spreaders of mycotoxins on major crop commodities in the Philippines [4]. A significant amount of exposure to aflatoxins may increase the risk of liver cancer [3].

Mold-infested peanuts may have aflatoxin which is found to be highly carcinogenic, posing risk to humans and livestock [5]. Due to the microscopic size of fungi, basic cleaning equipment may not be able to remove all contamination. Before reaching manufacturers and the peanut market, peanuts must be inspected and segregated effectively [5].

1.1. Problem Statement

Peanut farmers deal with pre- and post-harvest problems caused by *Aspergillus* fungi. In humid countries with warm, tropical temperatures like the Philippines, local peanut produce is more at risk of being contaminated with aflatoxins from *Aspergillus*. Aflatoxins are highly carcinogenic [3] and they pose a threat to human and livestock health.

In most cases, peanut kernels are inspected through manual inspection, biochemical testing, and chromatography [6]. Peanut products are checked by passing them through UV light in a dark room. This manual segregation procedure exposes workers to long periods of light and possible mold contamination, which is detrimental to their health and reduces productivity [7]. An article reported that “due to their high moisture content, fresh peanuts cannot be stored long-term (10-14 days)” [8].

This study seeks to develop a computer vision system as an approach to peanut mold detection.

1.2. Objectives

The study had the following objectives:

1. Build an artificial neural network model for detecting an early onset (7 days) of peanut mold contamination
2. Create a computer vision system capable of detecting mold infection in peanuts with cost-efficient and readily-available tools such as open-source software, consumer-grade smartphone camera, and LED lighting.

1.3. Significance of the Study

Peanut production in the Philippines is limited with our peanut supply being 75% import-dependent in 2019 [2]. Even so, peanuts have tremendous market potential in the country. Peanuts are a healthy snack food and an ingredient for a wide variety of Filipino dishes and confectioneries that are growing in popularity [2]. As mentioned, manual inspection of peanut seeds is laborious and risks their health which reduces productivity [7]. To keep up with the demand for peanut production, processing, and export in the country, it is essential to develop an effective and low-cost approach to ensuring peanut quality, one of which is detecting peanut mold contamination using computer vision and machine learning.

1.4. Scope and Limitations

The primary objects of this study are peanut kernels that have been removed from their shells. Peanuts were supplied by local suppliers in Tayabas City, Quezon, which were of the Runner and Java variety and were imported from India. The purpose of this study is to detect mold-infected peanut kernels after contamination with mold. Peanut kernel samples will be obtained from the proponent's family business of selling roasted peanuts.

2. Review of Related Literature

Several researchers have explored multispectral and hyperspectral imaging techniques, as well as machine learning techniques, to identify and separate kernels and seeds contaminated with aflatoxin-producing molds. Following are a few related studies that inspired the methods used in this study:

Suyantohadi and Masitoh (2016) developed a machine vision system with UV lighting to classify contaminated and uncontaminated peanut kernels using the K-means algorithm. The clustering

process used the RGB values as centroid values to represent uncontaminated and contaminated peanut seeds. A validation process was then carried out on the performance of the computer vision, and it resulted in the detection of mold-contaminated peanuts with an accuracy of 100% [9].

Ziyadee, et al. (2021) reported an accuracy of 99.2% using ANN with UV lighting and an accuracy of 99.7% using ANN with LED lighting in detecting mold infection in peanuts. This was after being exposed to contamination for 72 hours. Moreover, the researchers realized that mold contamination spread to all kernel samples after long periods of time rendered the samples irrelevant to their study [10].

In a study by Manhando et al. (2021), they developed a method for early detection of mold contamination in peanuts. The approach uses optical coherence tomography (OCT) imaging technology and a Support Vector Machine (SVM) based on ECOC that is trained based on features extracted via a Deep Convolutional Neural Network (DCNN). Results showed that after incubation periods of 48 and 96 hours, the developed method can detect mold-contaminated peanuts with an accuracy of approximately 85% and 96% [11].

For peanut variety classification, G and Kini (2018) employed multiple machine learning classifiers to obtain the most accurate predictive model. The researchers used the Gray level co-occurrence matrix (GLCM) to extract image texture features. The co-occurrence matrix was used to compute 12 texture features. The Correlation-based Feature Selection (CFS) technique was applied to reduce the dimensionality of the feature set and achieve high accuracy in selecting each classification model [12].

Zhao and Han (2009) developed a method for identifying peanut cultivars and quality by examining their appearance characteristics. From their sample images, 49 features were extracted in total. Principal Component Analysis was used to reduce, standardize, and de-correlate the feature data. Feature components with greater than a 90% contribution rate were selected for the recognition model. The recognition system was able to identify moldy peanuts with an accuracy of 96% and an average of 93% for all classifications using a forward neural network [13].

A study conducted by Jiang, et al. (2016) successfully identified mold-contaminated peanuts using Near-Infrared (NIR) hyperspectral images. The Principal Component Analysis (PCA) technique is employed in the spectral dimension to select sensitive bands and to project the spectral vector in the desired direction in order to identify moldy information. In parallel, the kernel-scale objects were segmented into spatial dimensions by using a marker-controlled watershed algorithm. Results demonstrated that the proposed method was 87.14% accurate in identifying contaminated kernels in the learning image and it was 98.73% accurate in the validation image [14].

3. Materials and Methods

The setup and tools that will be used and the process that will be followed for the study are described in this section. All of the implementation and libraries involved in the computer vision system are using **Python**. The figure below depicts the order of the activities that will be conducted during the study.

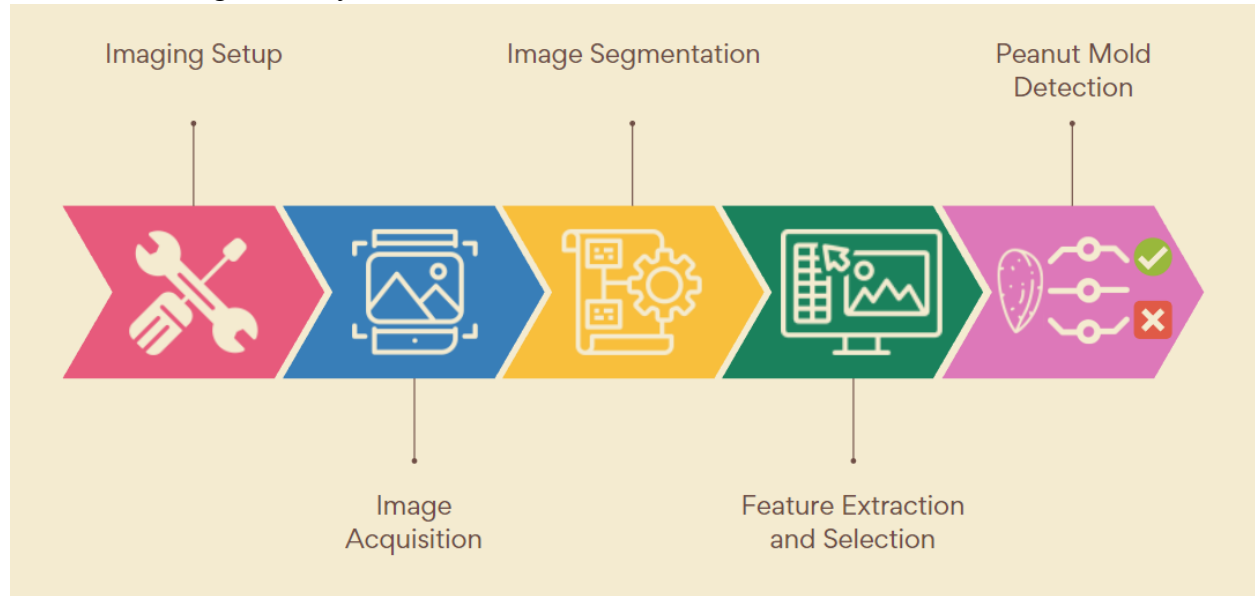


Figure 1. Activity Diagram / Flow.

3.1 Imaging Setup

As part of the computer vision system, an imaging box is to be built with the materials listed below. A **container box** can be made from any sturdy material (e.g. cardboard, Karton box) and must be able to shield the inside from external light sources. **Lights** must be able to illuminate the bottom of the container while producing a minimum amount of shadows. In this case, an **LED strip light** or LED ring light would be recommended. An **object container** with non-reflective blue background is recommended because it contrasts strongly with the color of peanut seeds. Most importantly, a **smartphone** (Mi10T 64MP camera) will be placed on top of the imaging box with a small hole fitted for the camera.

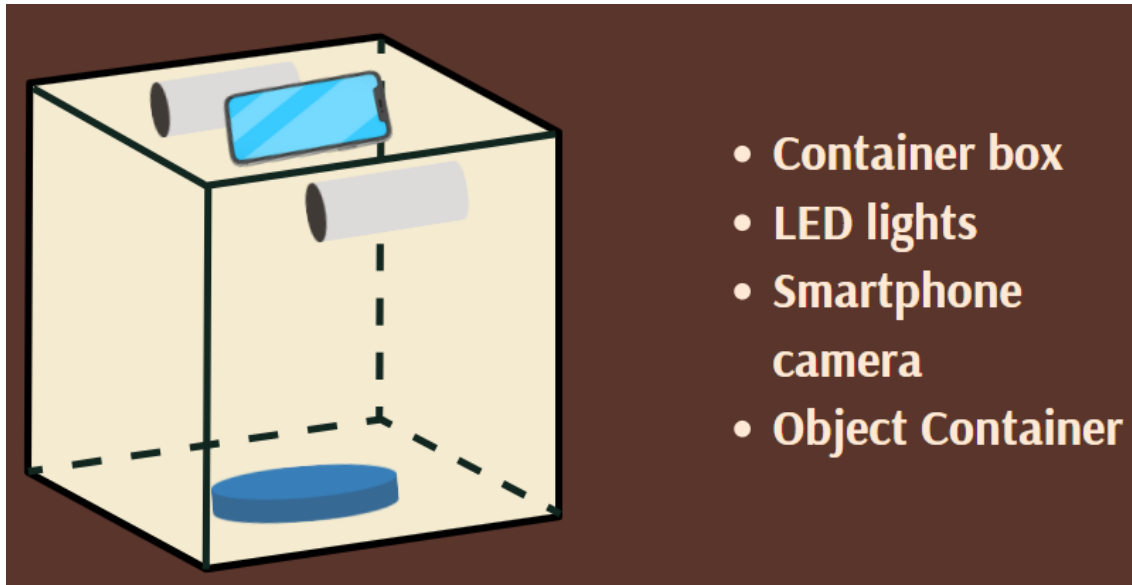


Figure 2. Imaging setup plan and the list of materials used.

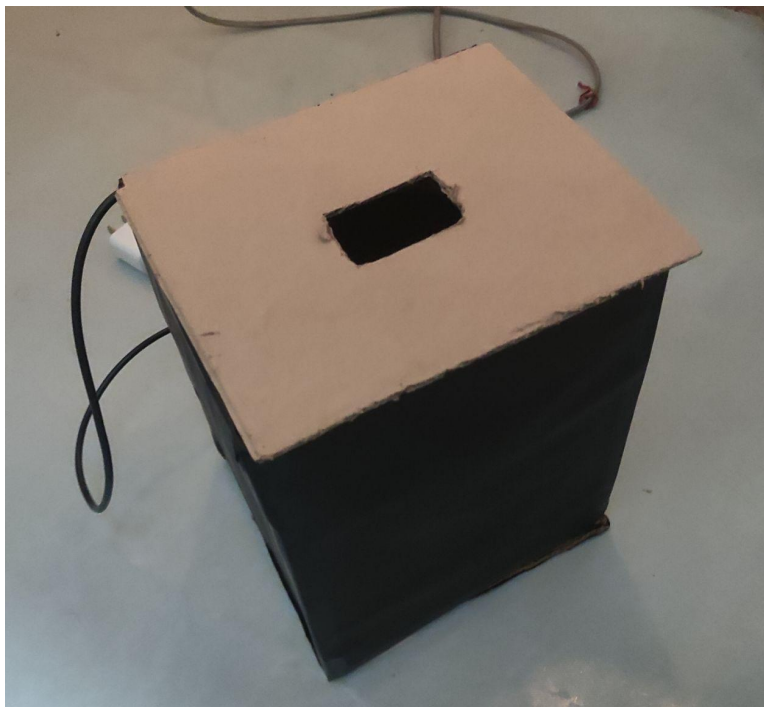


Figure 3. External View of Actual Imaging Setup.



Figure 4. *External View with Smartphone View of Actual Imaging Setup.*

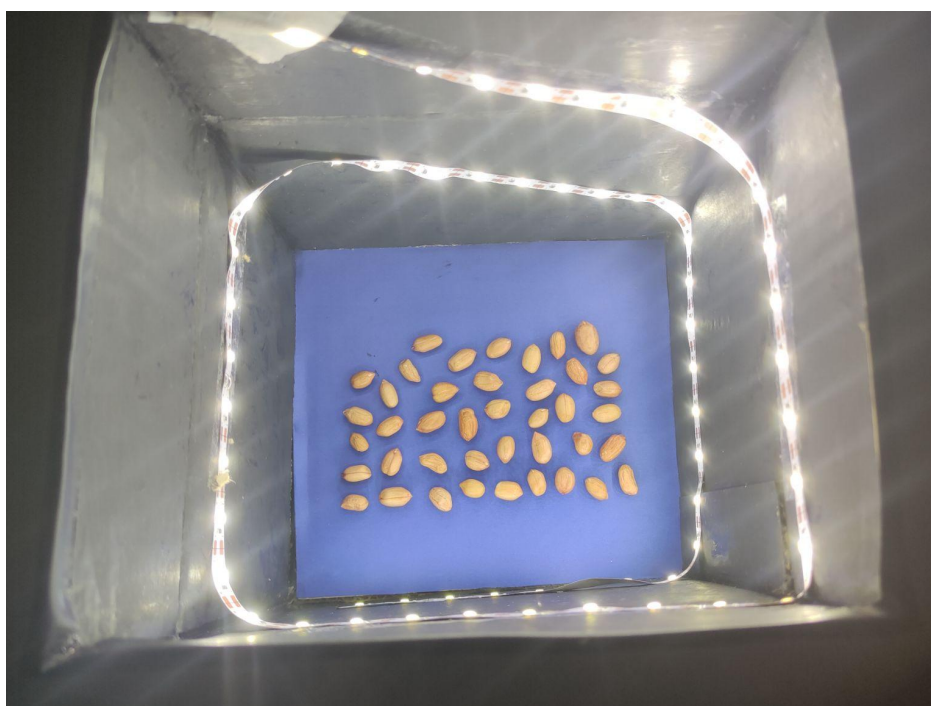


Figure 5. *Internal View of Actual Imaging Setup.*

3.2 Image Acquisition

Samples of peanut kernels were obtained from the proponent's family business of roasted peanuts. Peanuts were supplied by local suppliers in Tayabas City, Quezon, which were of the Runner and Java variety and were imported from India. The peanut kernels' images were taken in batches, meaning multiple peanut kernels were present in one image. In the case of mold-infested peanut seeds, a batch was stored in a breathable container in a dark room. Peanut kernels were sprayed with water sparingly if the air is too dry for molds to grow. After 5-7 days or as soon as gray-green or yellow-green spots appear on the peanut seeds, batches of samples were taken images. The following are a few of the non-contaminated and mold-contaminated samples.



Figure 6. Sample Image of Non-Contaminated Peanut Seeds.



Figure 7. Sample Image of Mold-Contaminated Peanut Seeds.

3.3 Image Segmentation

The image samples went under **color segmentation** to isolate the peanut kernels from the background. This process was performed using **OpenCV** which involves the following steps:

1. Transport the raw RGB images into HSV
2. Find the most appropriate hue range that contrasts with the blue background.
3. Isolate the peanut seed samples by removing the background.
4. Clean the edges of the peanut kernels using morphological operations such as opening and closing.

In order to improve the image segmentation process, the **marker-based watershed algorithm** was implemented. In this process, the color segmentation procedure was extended with the following steps:

1. Find guaranteed foreground from the background using distance transformation.
2. Create object markers using connected component analysis.
3. Apply watershed segmentation based on the marked objects.

Examples of color segmentation and marker-based watershed segmentation are shown in Figures 8-11.



Figure 8. *Sample Image of Non-Contaminated Peanut Seeds with Color Segmentation.*



Figure 9. *Sample Image of Non-Contaminated Peanut Seeds with Marker-based Watershed Segmentation.*



Figure 10. *Sample Image of Mold-Contaminated Peanut Seeds with Color Segmentation.*



Figure 11. *Sample Image of Mold-Contaminated Peanut Seeds with Marker-based Watershed Segmentation.*

3.4 Feature Extraction and Selection

The color characteristics of each peanut kernel were extracted. **OpenCV** was used to convert RGB image samples to HSV and grayscale color spaces. The averages of each color characteristic were calculated. Table 1 lists the color features gathered in the study.

Table 1. Color Characteristics.

| | |
|---------------|--|
| 1) Red | |
| 2) Green | |
| 3) Blue | |
| 4) Hue | $\begin{cases} 60(G - B)/(V - \min(R, G, B)) & \text{if } V = R \\ 120 + 60(B - R)/(V - \min(R, G, B)) & \text{if } V = G \\ 240 + 60(B - R)/(V - \min(R, G, B)) & \text{if } V = B \\ 0 & \text{if } R = G = B \end{cases}$ |
| 5) Saturation | $\begin{cases} \frac{V - \min(R, G, B)}{V} & \text{if } V \neq 0 \\ 0 & \text{otherwise} \end{cases}$ |
| 6) Value | $\max(R, G, B)$ |

To obtain texture features, the **Gray Level Co-occurrence Matrix (GLCM)** properties were computed using **Scikit-Image** as shown in Table 2.

Table 2. Texture Features.

| | |
|--------------------------|---|
| 1) Contrast | $\sum_{i,j=0}^{levels-1} P_{ij}(i - j)^2$ |
| 2) Dissimilarity | $\sum_{i,j=0}^{levels-1} P_{ij} i - j $ |
| 3) Homogeneity | $\sum_{i,j=0}^{levels-1} \frac{P_{ij}}{i + (i - j)^2}$ |
| 4) Angular Second Moment | $\sum_{i,j=0}^{levels-1} P_{ij}^2$ |
| 5) Energy | $\sqrt{\sum_{i,j=0}^{levels-1} P_{ij}^2}$ |
| 6) Correlation | $\sum_{i,j=0}^{levels-1} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$ |

To avoid overfitting, the **Python Data Analysis Library (pandas)** will be used to create the correlation matrix, from which the features will be filtered to be inducted into the artificial neural network.

3.5 Peanut Mold Detection

The artificial neural network model for peanut mold detection will be built, trained, and tested using **Keras from TensorFlow**. The dataset will be distributed into **70% for training** and **30% for testing**. The following flowchart illustrates the process involved in designing the prediction model for detecting mold-contaminated and non-contaminated peanut kernels. Figure 12 illustrates the process that transpired in designing the models to detect peanut mold.

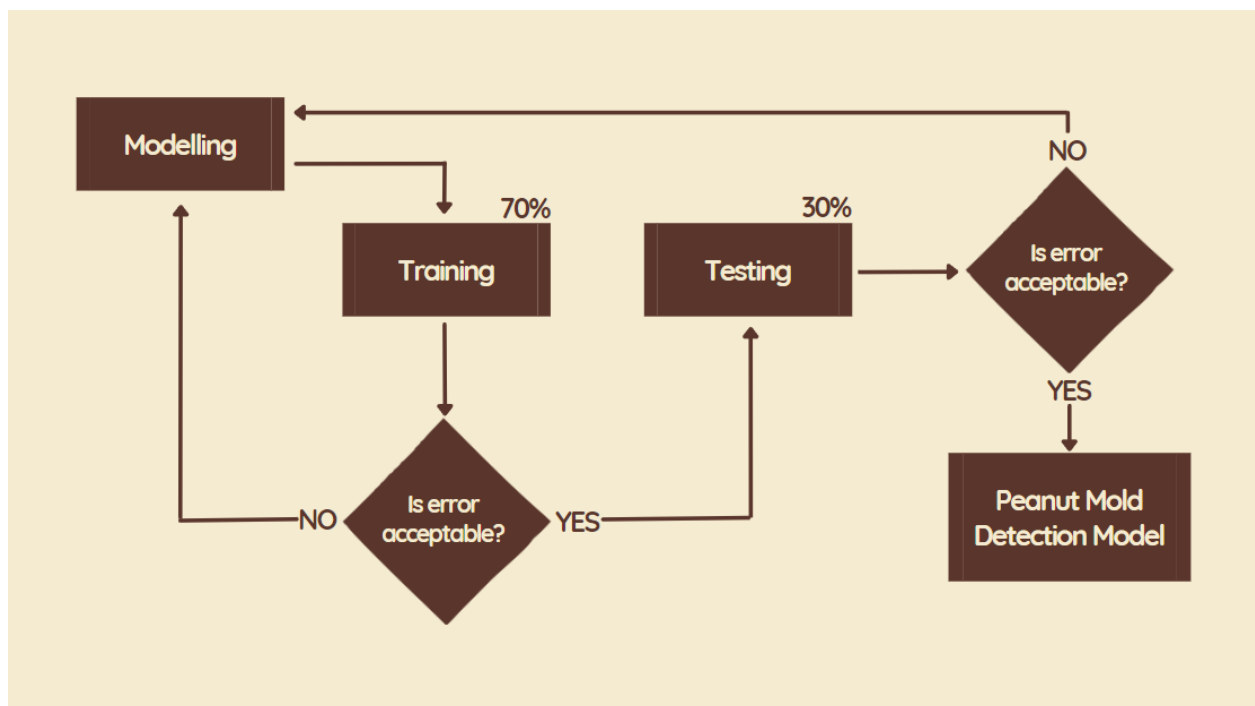


Figure 12. *Peanut Mold Detection Model Flowchart.*

3.6 Algorithms To Be Used

Marker-based Watershed Segmentation

When the images are transferred to the binary space, adjacent target objects may overlap and merge due to the limitations of the imaging setup. To improve the image segmentation, the images were further processed via the Marker-based Watershed Segmentation. This multi-stage algorithm interprets a grayscale image as a topographic surface (Figure 13) where high-intensity values denote peaks and hills while low-intensity values denote valleys. Each local minima

(catchment basins) are gradually adjusted until they have reached a certain threshold. To prevent the labels from overlapping, markers were placed to strengthen the peaks (watershed lines) [15]. As a result of the marker-based watershed process, the watershed lines should be able to define more accurately segmented objects.

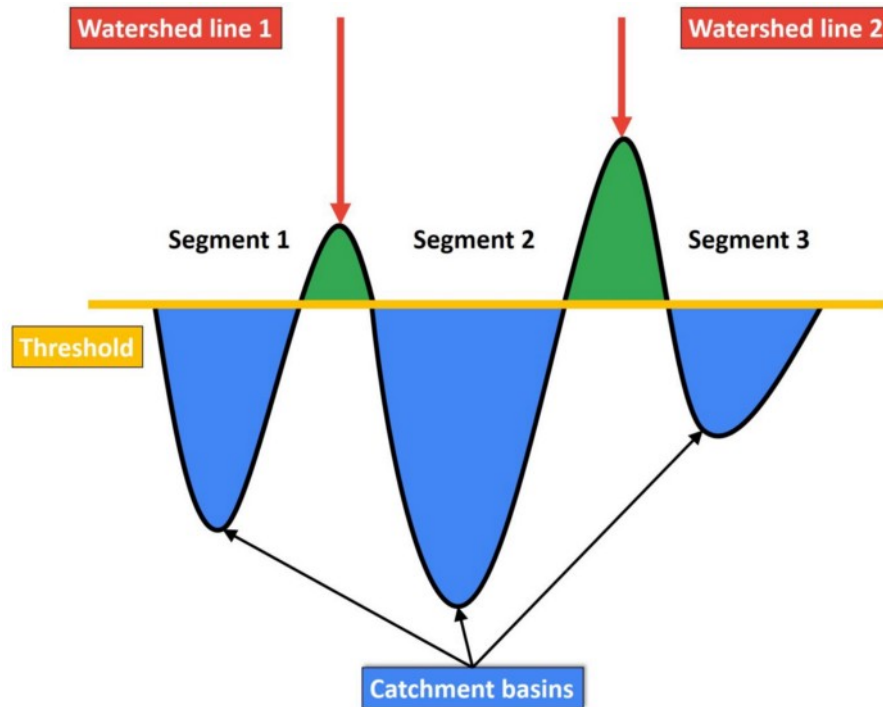


Figure 13. Visual Representation of Marker-based Watershed Algorithm [16].

Color Segmentation

In a study, Khattab et al. described color segmentation as the process of segmenting images based on the color features of pixels, and on the assumption that homogenous colors in an image represent separate clusters of objects [17]. For example, relevant objects in an image may have a similar hue range that can be distinguished from others, in this case, the peanut seeds against the blue background. As a result, the background of the image can be easily removed, leaving only the peanut kernels for further analysis.

Feedforward Backpropagation Neural Network

An artificial neural network with a feedforward backpropagation algorithm was used to determine mold contamination in peanut kernels. The feedforward backpropagation algorithm, as its name implies, consists of two steps. In the first step, the features are fed forward to the network as input. Hidden neurons transform input values into linear functions that approximate the output value. Gradients are computed based on the predicted output value and are propagated back to the neurons. As a result of this iterative process, accuracy can be increased [18]. The output layer contains the mold contamination score of values 0 and 1, with 1 being contaminated.

The learning algorithm will use the binary cross-entropy as the loss function and the sigmoid tangent as the activation function.

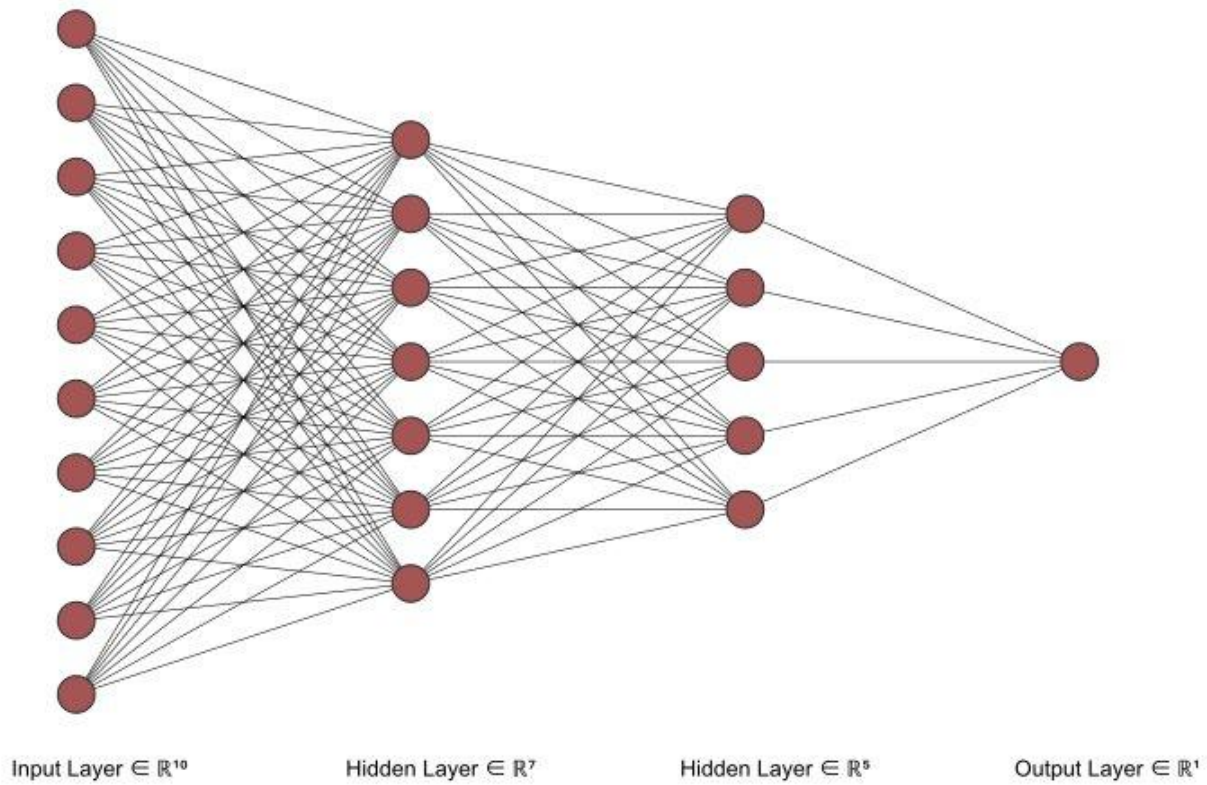


Figure 14. Visualization of Feedforward Backpropagation Neural Network.

4. Results and Discussion

The results from the implementation of methods for the detection and classification of mold-infected peanut kernels are reported in this section.

4.1 Molds Present in the Samples

With the guidance of Prof. Julie Aiza Mandap [19], different species of molds were identified in the sample images. In the sample images below (Figure 15), four species of *Aspergillus* fungi were found: *A. parasiticus* and *A. flavus* (dark-green) [19], [20], *A. glaucus* (yellowish-green) [19], [21], and *A. niger* (black) [19], [22]. The *Aspergillus* species were described as hairy fungi with many stipes or columns. Several common molds were also observed, including *Alternaria* and *Curvularia* which shows noticeable blight and decay on the peanut seeds [19], [23]. Furthermore, *Fusarium* and *Rhizopus* species were characterized by their white, cottony appearance [19], [23].

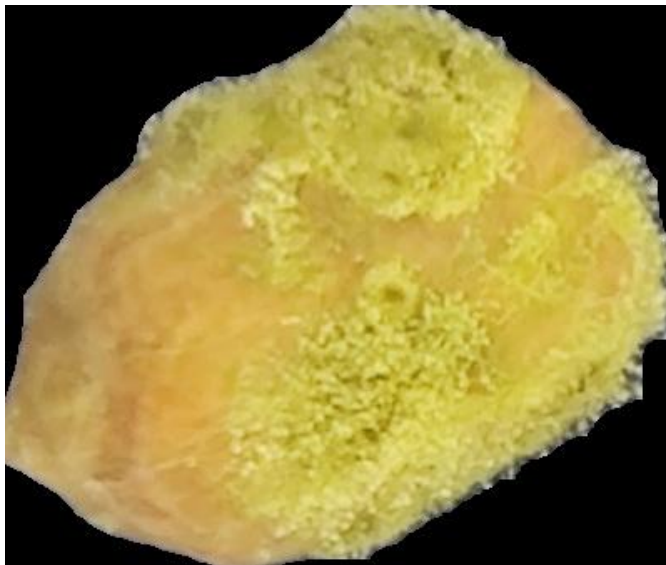


Figure 15. Peanut Seed with *Aspergillus parasiticus* and *Aspergillus flavus* molds.

Figure 15 shows the presence of *Aspergillus parasiticus* and *Aspergillus flavus* on peanut seeds. Both species of mold have dark green colors and long stalks [19], [20]. It is not possible to distinguish between the two species without a microscope [19].



Figure 16. Peanut Seed with *Aspergillus glaucus* mold.

Figure 16 shows *Aspergillus glaucus* growing on peanut seeds. Compared to *A. parasiticus* and *A. flavus*, *A. glaucus* appears more yellow and bright in color [19], [21].



Figure 17. Peanut Seed with *Aspergillus niger* mold.

On the right side of the peanut seed, *Aspergillus niger* can be seen in Figure 17. *A. niger*'s dominantly black color distinguishes them from other *Aspergillus* species [19], [22].



Figure 18. *Peanut Seeds with Alternaria mold.*

Figure 18 depicts a peanut seed infected with *Alternaria* mold all around [19], [22]. This mold is described by its dark brown color and slightly warty appearance [24].



Figure 19. *Peanut Seeds with Curvularia mold.*

Curvularia mold [19], [22] is found at the middle of a peanut seed in Figure 19. The dark brown hue and lengthy, thread-like structure distinguish this mold [25].



Figure 20. *Peanut Seed with Fusarium mold.*

Figure 20 illustrates a *Fusarium* mold-infected peanut seed [19], [22]. Gray-white, dry-appearing structures with feathery edges characterize this mold [26].



Figure 21. *Peanut Seed with Rhizopus mold.*

In Figure 21, a peanut kernel infested with *Rhizopus* mold [19], [22] can be seen near its bottom side. This mold is distinguished by its white cottony nature [27].

4.2 Selected Features

The **Python Data Analysis Library (pandas)** was used to compute the Pearson correlation p-values between one feature and another, from which the features will be filtered to be inducted into the artificial neural network. The following figures (Figures 22 and 23) show the pairwise correlation matrices showing the linear correlation between features in the datasets produced from color segmentation and watershed segmentation. The closer the correlation value between two features is to 1, the more it indicates that including two of these features is redundant and only one should remain. Pruning redundant features reduce overfitting and the neural network's computational overhead.

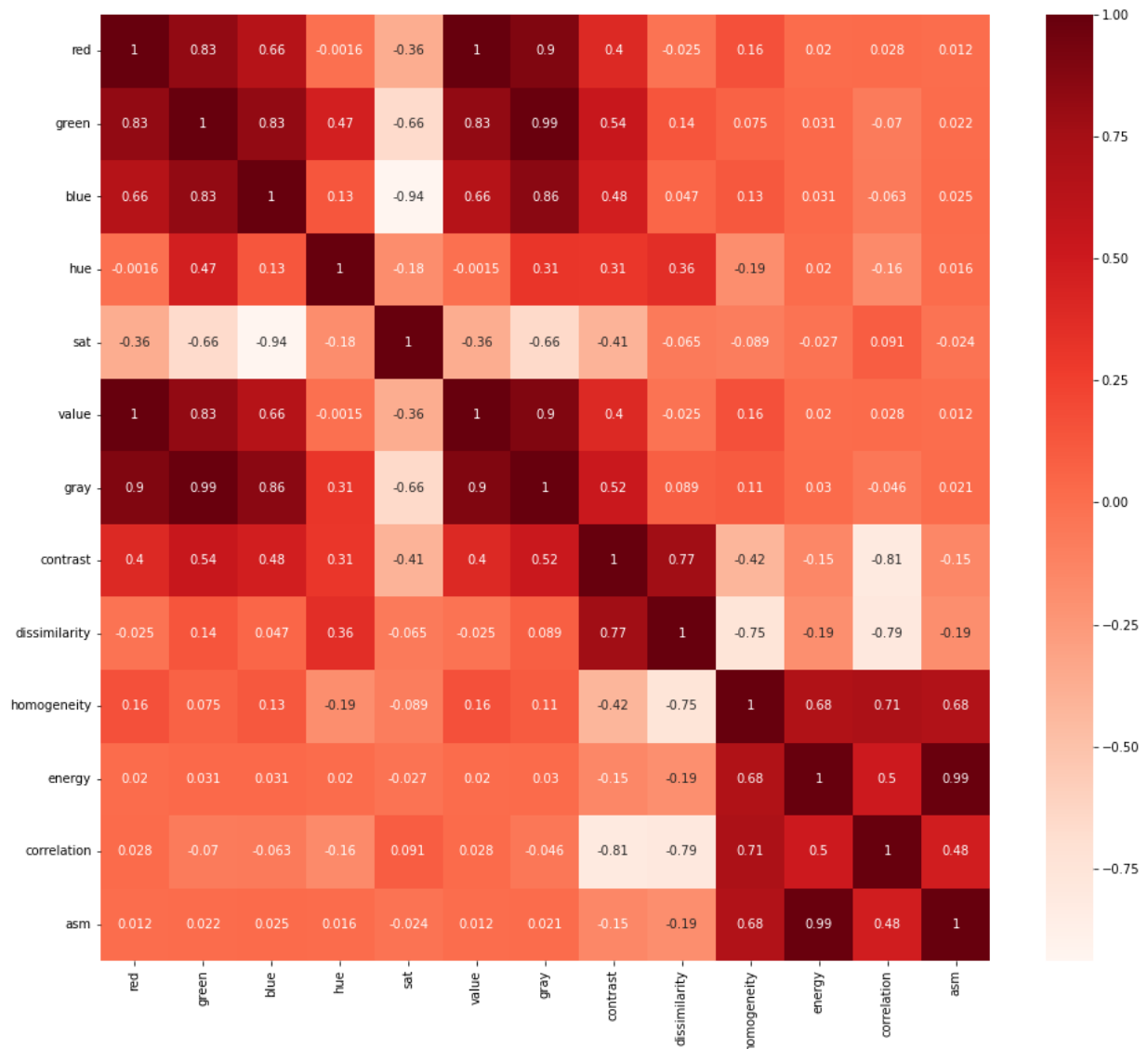


Figure 22. Pairwise Correlation Matrix Between Features from Color Segmentation Dataset.

Figure 22 shows the linear correlation values between each feature in a pairwise correlation matrix. When the correlation values of each characteristic are compared, the features with the

greatest correlation values emerge. According to the above correlation values, Red and Gray, Green and Gray, and Energy and ASM all have correlation values more than or equal to 0.9. After examining each correlation value, it was discovered that the greatest correlation values among the acquired attributes are Red, Green, and Energy. As a result, they were eliminated from the input variables in order to avoid overfitting and speed up the learning rate.

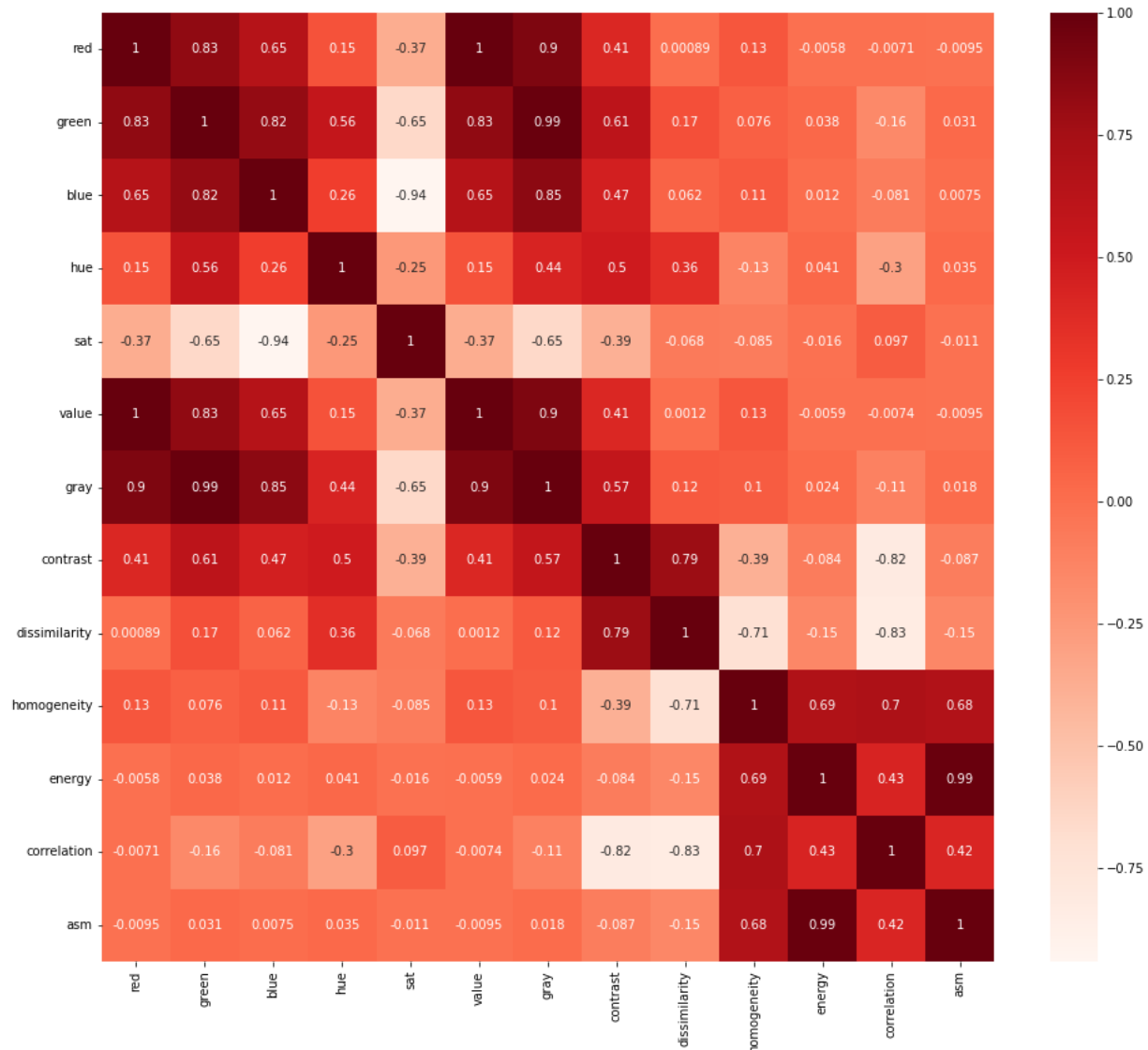


Figure 23. Pairwise Correlation Matrix Between Features from Marker-based Watershed Segmentation Dataset.

In a pairwise correlation matrix, Figure 23 shows the linear correlation values between features. According to the matrix, the correlation values between Value and Green, Green and Gray, and Energy and ASM are larger than or equal to 0.9. Value, Green, and Energy have the highest correlation, as determined by comparing each two correlation values. Consequently, they were removed from the input data to avoid overfitting and to speed up the learning rate.

4.3 Peanut Mold Detection

During the course of the study, 1,725 peanut kernel samples were gathered, with 863 samples classified as mold-contaminated and 862 samples classified as non-contaminated. Two neural networks were created using data from color segmentation and marker-based watershed segmentation, and both were trained for 20,000 iterations. The log loss of each neural network as the number of iterations grows is shown in the graphs below (Figures 24 and 25). The log loss or error rate indicates how close the prediction probability is to the actual label, whereas accuracy merely indicates if the prediction is correct, whether it is more than or less than the binary threshold of 0.5. This means that log loss provides a more accurate assessment of the neural network model's performance.

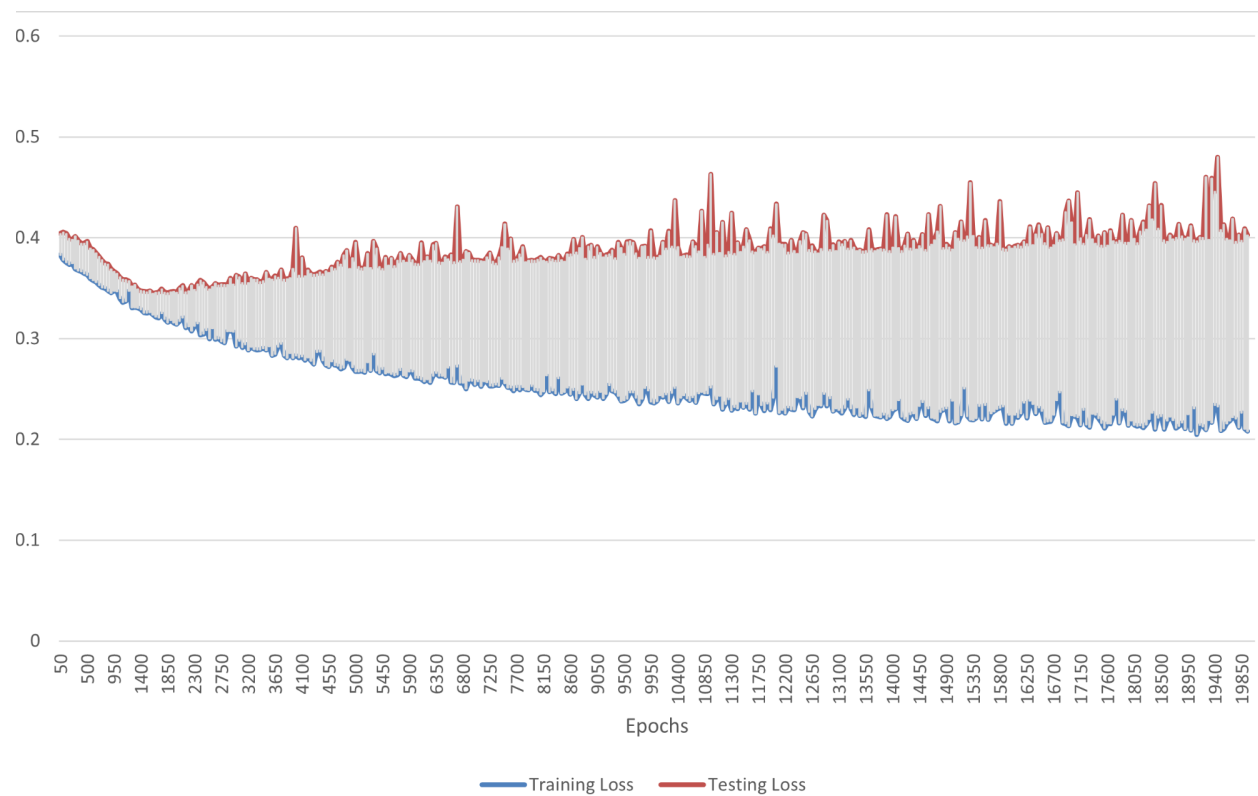


Figure 24. Log Loss of Peanut Mold Detection Model (Color Segmentation): Training vs Testing.

Figure 24 shows a graphical illustration of the log loss rate on the neural network between training and testing using color segmentation data under 20,000 iterations of training. The prediction model obtained a minimum testing loss of 34.5 percent and a training log loss of 32.36 percent after 1650 iterations of training. Following this, the testing loss rose once again, before progressively decreasing due to overfitting. The neural network is guided to the optimal number of iterations it should be trained before overfitting by determining the minimum loss.

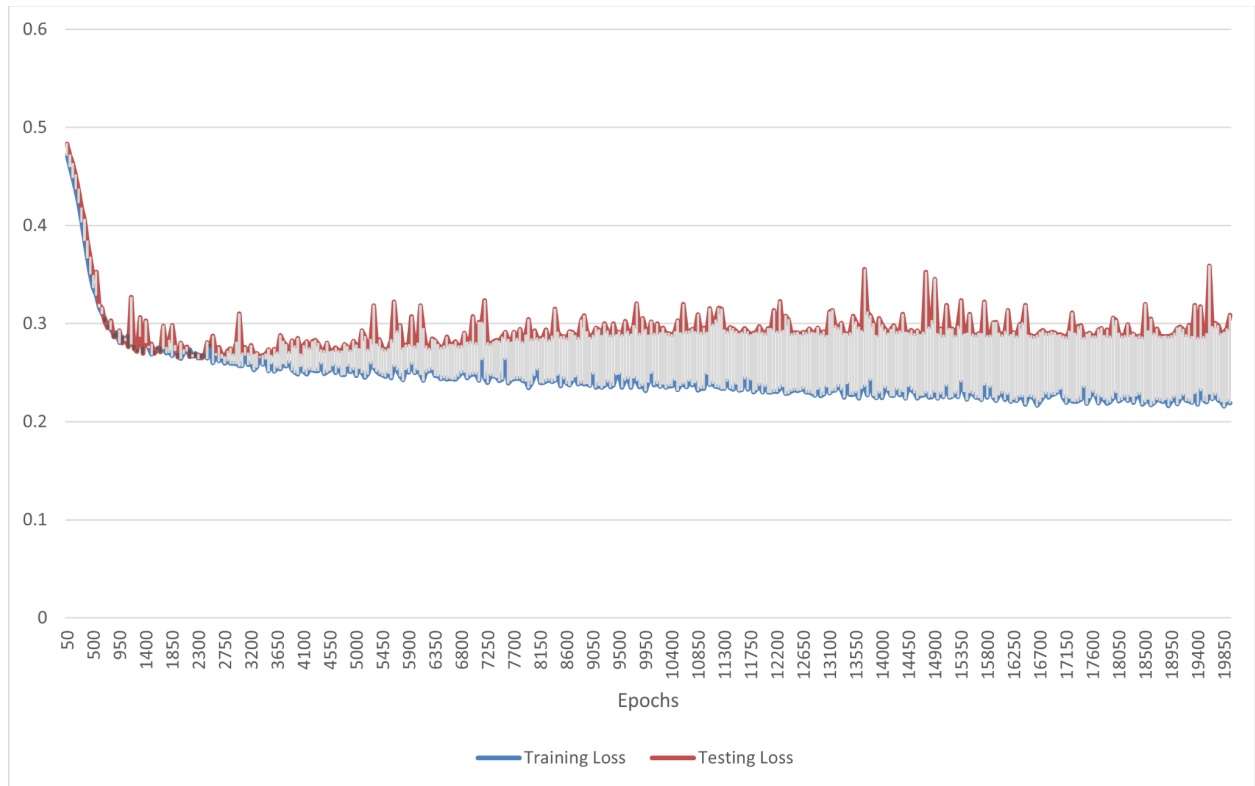


Figure 25. *Log Loss of Peanut Mold Detection Model (Marker-based Watershed Segmentation): Training vs Testing.*

Figure 25 depicts a graphical representation of the log loss rate on the neural network between training and testing using data from marker-based watershed segmentation under 20,000 iterations of training. The prediction model obtained its minimal testing loss of 26.5 percent and training loss of 26.78 percent after 2400 iterations of training. The similar result was seen, as the testing loss increased and continued to decrease owing to overfitting. This outcome will be used to determine how long the model should be trained before terminating in order to retain its minimum loss.

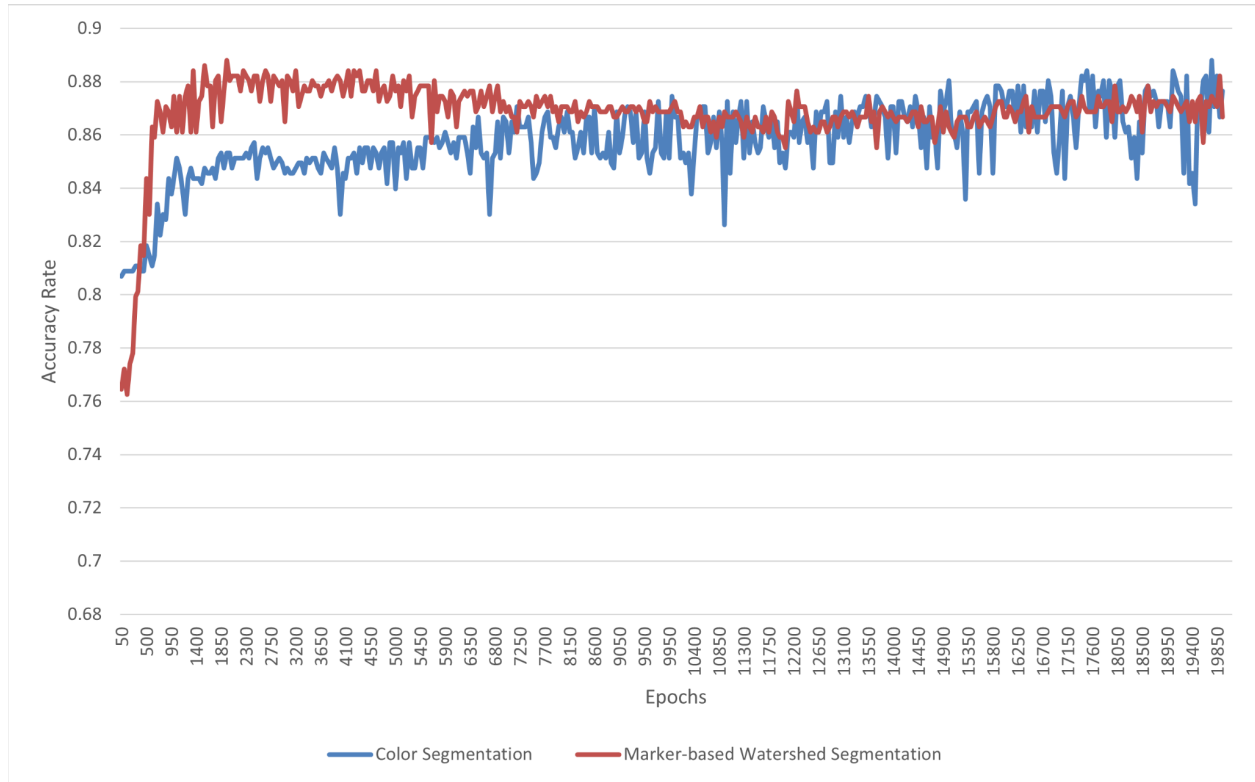


Figure 26. Accuracy of Peanut Mold Detection Models: Color Segmentation vs Marker-based Watershed Segmentation.

Figure 26 depicts the accuracy rates of the two neural networks utilizing color segmentation data and marker-based watershed segmentation after 20,000 training iterations. Color segmentation and marker-based watershed segmentation prediction models are identical in terms of peak accuracy rate, with an accuracy of 88.8 percent. However, the model employing watershed segmentation data achieved its maximal accuracy significantly faster, at 2000 iterations, than the model using color segmentation data, which took 19800 iterations.

Table 3. *Comparison of Accuracy and Loss Between Image Segmentation Methods*

| Segmentation Method | Minimum Loss (%) | No. of Iterations at Minimum Loss | Accuracy at Minimum Loss (%) | Peak Accuracy (%) | No. of Iterations at Peak Accuracy | Loss at Peak Accuracy (%) |
|-------------------------------------|-------------------------|--|-------------------------------------|--------------------------|---|----------------------------------|
| Color Segmentation | 34.50 | 1650 | 84.56 | 88.80 | 19800 | 39.63 |
| Marker-based Watershed Segmentation | 26.50 | 2400 | 88.03 | 88.80 | 2000 | 26.57 |

Table 3 compares the performance of both neural networks under 20,000 training iterations, one utilizing color segmentation data and the other using marker-based watershed segmentation.

The model trained using marker-based watershed segmentation data produced a lower minimum error rate of 26.5 percent than the model trained with color segmentation data, which had an error rate of 34.5 percent. As a result, the watershed neural network's accuracy at its minimal loss is considerably larger. Furthermore, the color segmentation model achieved its minimum loss after 1650 training cycles, but the watershed model reached its minimum loss after 2400 iterations.

Both neural networks, on the other hand, attained the same peak accuracy rate (88.80%). The color segmentation model, however, has a substantially greater error rate (39.63%) than the watershed error rate (26.57%). Moreover, the color segmentation model involves 19800 training iterations to obtain optimal accuracy, but the watershed model only required 2000 training iterations. Inconsistency between accuracy and loss indicates that while the correct predictions are increasing, the wrong predictions are straying further away from the actual values due to overfitting.

4.4 Validation of Samples

Upon consultation with Prof. Mandap, it was known that it is possible for peanut seeds to not be contaminated even after exposing them to other mold-contaminated peanut seeds. To further validate the mold contamination present on the peanut seeds, they were examined one-by-one following the guidelines as stated in the methodology (3.2). Thereafter, the neural network predictions were compared with the actual labels. The tables below (Tables 4 and 5) show the confusion matrix between actual and predicted labels. The actual accuracy rates of both neural networks were then calculated using the following formula:

$$\text{Accuracy} = \frac{\text{True Negatives} + \text{True Positives}}{\text{Positives} + \text{Negatives}}$$

Table 4. Confusion Matrix of Peanut Mold Detection Model Using Color Segmentation Data.

| | | Predicted | |
|--------|-------------------|------------------|-------------------|
| | | Non-Contaminated | Mold-Contaminated |
| Actual | Non-Contaminated | 768 | 118 |
| | Mold-Contaminated | 125 | 714 |

The findings in Table 4 were obtained by comparing the hand inspection labels to those predicted by the neural network, resulting in a confusion matrix. The peanut mold contamination detection model was re-trained up to 1700 times based on the number of iterations with the lowest error rate (Table 3).

There were 768 samples that were identified and correctly predicted as non-contaminated (true negatives), 118 samples that were erroneously predicted as mold-contaminated (false positives), 125 samples that were erroneously predicted as non-contaminated (false negatives), and 714 samples that were labeled and accurately predicted as mold-contaminated (true positives).

The actual accuracy of the model was computed at **85.91%** based on the findings displayed on the confusion matrix.

Table 5. Confusion Matrix of Peanut Mold Detection Model Using Marker-based Watershed Segmentation Data.

| | | Predicted | |
|--------|-------------------|------------------|-------------------|
| | | Non-Contaminated | Mold-Contaminated |
| Actual | Non-Contaminated | 779 | 107 |
| | Mold-Contaminated | 77 | 762 |

Table 5 shows the results of comparing hand inspection labels to those predicted by the neural network as presented in a confusion matrix. Based on the number of iterations with the lowest error rate, the peanut mold contamination detection model was re-trained up to 2500 iterations (Table 3).

There were 779 samples labeled and correctly predicted as non-contaminated (true negatives), 107 samples incorrectly predicted as mold contaminated (false positives), 77 samples incorrectly predicted as non-contaminated (false negatives), and 762 samples labeled and correctly predicted as mold-contaminated (true positives).

Based on the data reported on the confusion matrix, the model's actual accuracy was calculated to be **89.33%**.

5. Conclusion

Based on the results obtained from the peanut mold detection models, non-contaminated and mold-contaminated peanut seeds differ significantly in terms of gathered color and texture characteristics. Upon manual inspection, a wide variety of molds, including *Aspergillus* were indeed present on the peanut seeds after inducing mold growth. During the study, two prediction models were developed, one using color segmentation data and one using marker-based watershed segmentation data. The neural network using marker-based watershed segmentation data achieved an accuracy of 89.33% which is higher than the color segmentation neural network with an accuracy of 85.91%. As an outcome, these findings indicate that a computer vision system built using low-cost and easily available components, such as a smartphone camera, open-source software, and LED strip lighting, as demonstrated in the study, can efficiently detect mold contamination in peanut kernels.

6. Recommendation

Further research could include models that can identify the specific mold species present on peanut seeds, for example, the differences between *Aspergillus* species and other kinds of mold. Considering that the presence of spores on peanut seeds is always a possibility [19], it is also important to develop a method for determining the level of contamination. In addition to GLCM properties, other tactile features should be considered, such as wavelet transformation [5], edge-based texture granularity [28], or combinations thereof. As soon as high spec cameras become more available, they should also be used to improve the quality of image samples.

References

- [1] American Peanut Council, "The Peanut Industry." <https://www.peanutsusa.com/about-apc/the-peanut-industry.html> (accessed May 27, 2022)
- [2] Agribusiness And Marketing Assistance Service, "Investment Guide For Peanut." 2021. [Online]. Available: <https://www.da.gov.ph/wp-content/uploads/2021/04/Investment-Guide-for-Peanut.pdf>
- [3] National Cancer Institute, "Aflatoxins - Cancer-Causing Substances - NCI," Mar. 20, 2015. <https://www.cancer.gov/about-cancer/causes-prevention/risk/substances/aflatoxins> (accessed May 27, 2022).
- [4] M. Balendres, P. Karlovsky, and C. Cumagun, "Mycotoxigenic Fungi and Mycotoxins in Agricultural Crop Commodities in the Philippines: A Review," *Foods*, vol. 8, no. 7, p. 249, Jul. 2019, doi: 10.3390/foods8070249.
- [5] X. Qi, J. Jiang, X. Cui, and D. Yuan, "Moldy Peanut Kernel Identification Using Wavelet Spectral Features Extracted from Hyperspectral Images," *Food Anal. Methods*, vol. 13, no. 2, pp. 445–456, Feb. 2020, doi: 10.1007/s12161-019-01670-w.
- [6] L. B. Abdulra'Uf, *Aflatoxin - Control, Analysis, Detection and Health Risks | IntechOpen*. 2017. Accessed: May 27, 2022. [Online]. Available: <https://www.intechopen.com/books/5956>
- [7] I. Lavkor and I. Var, "'The Control of Aflatoxin Contamination at Harvest, Drying, Pre-Storage and Storage Periods in Peanut: The New Approach,' in Aflatoxin-Control, Analysis, Detection and Health Risks," in *The Control of Aflatoxin Contamination at Harvest, Drying, Pre- Storage and Storage Periods in Peanut: The New Approach | IntechOpen*, 2017. Accessed: May 27, 2022. [Online]. Available: <https://www.intechopen.com/chapters/55153>
- [8] Agricultural Marketing Resource Center, "Peanut Profile," 2022. <https://www.agmrc.org/commodities-products/nuts/peanut-profile> (accessed May 27, 2022).
- [9] A. Suyantohadi and R. E. Masithoh, "Identify Toxin Contamination in Peanuts Using the Development of Machine Vision Based on Image Processing Technique," *KnE Life Sci.*, vol. 3, no. 3, p. 83, Jan. 2016, doi: 10.18502/kls.v3i3.404.
- [10] P. Ziyae, V. Farzand Ahmadi, P. Bazayr, and E. Cavallo, "Comparison of Different Image Processing Methods for Segregation of Peanut (*Arachis hypogaea* L.) Seeds Infected by Aflatoxin-Producing Fungi," *Agronomy*, vol. 11, no. 5, p. 873, Apr. 2021, doi: 10.3390/agronomy11050873.
- [11] E. Manhando, Y. Zhou, and F. Wang, "Early Detection of Mold-Contaminated Peanuts Using Machine Learning and Deep Features Based on Optical Coherence Tomography," *AgriEngineering*, vol. 3, no. 3, pp. 703–715, Sep. 2021, doi: 10.3390/agriengineering3030045.
- [12] N. V G, A. S. Kini, and A. S. Kini, "An intelligent classification model for peanut's varieties by color and texture features," *Int. J. Eng. Technol.*, vol. 7, no. 2.27, p. 250, Aug. 2018, doi: 10.14419/ijet.v7i2.27.12473.
- [13] Z. Han and Y. Zhao, "A Method of Detecting Peanut Cultivars and Quality Based on the Appearance Characteristic Recognition," in *2009 Second International Conference on Information and Computing Science*, Manchester, England, UK, 2009, pp. 21–24. doi: 10.1109/ICIC.2009.113.
- [14] J. Jiang, X. Qiao, and R. He, "Use of Near-Infrared hyperspectral images to identify moldy peanuts," *J. Food Eng.*, vol. 169, pp. 284–290, Jan. 2016, doi: 10.1016/j.jfoodeng.2015.09.013.
- [15] H.-S. Chuang and C.-H. Lin, "Automated grain sizing using mark-based watershed algorithm," in *2012 IEEE International Geoscience and Remote Sensing Symposium*,

- Munich, Germany, Jul. 2012, pp. 2332–2335. doi: 10.1109/IGARSS.2012.6351027.
- [16] S. Zivkovic, “#007 OpenCV projects – Image segmentation with Watershed algorithm,” *Data Hacker*, 2020.
<https://datahacker.rs/007-opencv-projects-image-segmentation-with-watershed-algorithm/>
 - [17] D. Khattab, H. M. Ebied, A. S. Hussein, and M. F. Tolba, “Color Image Segmentation Based on Different Color Space Models Using Automatic GrabCut,” *Sci. World J.*, vol. 2014, pp. 1–10, 2014, doi: 10.1155/2014/126025.
 - [18] C. Hansen, “Neural Networks: Feedforward and Backpropagation Explained,” *Machine Learning From Scratch*, Aug. 05, 2019.
<https://mlfromscratch.com/neural-networks-explained/> (accessed May 29, 2022).
 - [19] J. A. Mandap, “Consultation with Prof. Mandap,” May 25, 2022.
 - [20] R. Khan, F. Mohamad Ghazali, N. A. Mahyudin, and N. I. P. Samsudin, “Morphological Characterization and Determination of Aflatoxigenic and Non-Aflatoxigenic *Aspergillus flavus* Isolated from Sweet Corn Kernels and Soil in Malaysia,” *Agriculture*, vol. 10, no. 10, p. 450, Oct. 2020, doi: 10.3390/agriculture10100450.
 - [21] Students of Jennifer Talbot, “*Aspergillus Glaucus* - microbewiki.”
https://microbewiki.kenyon.edu/index.php/Aspergillus_Glaucus (accessed May 27, 2022).
 - [22] P. Hinton-Sheley, “What is *Aspergillus niger*?,” 2018.
<https://www.news-medical.net/life-sciences/What-is-Aspergillus-niger.aspx> (accessed May 27, 2022).
 - [23] M. A. Junaithal, . B. V., and . M. J., “Storage Fungi in Groundnut and the Associate Seed Quality Deterioration-A Review,” *Plant Pathol. J.*, vol. 12, no. 3, pp. 127–134, Jun. 2013, doi: 10.3923/ppj.2013.127.134.
 - [24] “*Alternaria Fungi* - Classifications, Characteristics and Pathogenesis,” *MicroscopeMaster*.
<https://www.microscopemaster.com/alternaria.html> (accessed May 30, 2022).
 - [25] Doctor Fungus, “*Curvularia Species*,” *Doctor Fungus*.
<https://drfungus.org/knowledge-base/curvularia-species/> (accessed May 30, 2022).
 - [26] M. J. Blaser, J. E. Bennett, and R. Dolin, *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. Elsevier, 2015. doi: 10.1016/C2012-1-00075-6.
 - [27] Vedantu, “*Rhizopus*,” *VEDANTU*. <https://www.vedantu.com/biology/rhizopus>,
<https://www.vedantu.com/biology/rhizopus>, <https://www.vedantu.com/biology/rhizopus>,
<https://www.vedantu.com/biology/rhizopus> (accessed May 30, 2022).
 - [28] H. Liang and D. S. Weller, “Edge-based texture granularity detection,” in *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 3563–3567. doi: 10.1109/ICIP.2016.7533023.