*American Journal of Sciences and Engineering Research*

*E-ISSN -2348 – 703X, Volume 8, Issue 2, 2025*

---

# Neural Network Based Classification of Agro-Products using Infrared Images

**Zeeshan Ali Haq[1], Zainul Abdin Jaffery[1], Shabana Mehfuz[1]**

[1]*Dept. Electrical Engineering, Jamia Millia Islamia, New Delhi, India.*

***ABSTRACT:*** *Quality assessment of agricultural products is essential in ensuring sustainability of agricultural based industries. It also ensures that sufficient information is available for customers and consumers to decide the product to obtain as per their requirement. In this paper a novel approach of pre-processing based on Variational Mode Decomposed thermograms of potatoes and guavas are presented. Quality assessment of potatoes and maturlty level grading of guava is performed using Convolutional Neural Network based architecture. This approach involves extracting Variational Mode Decomposed modes of defective and wholesome potatoes. Short time Fourier Transform of all the modes are performed to obtain the spectrum image of the thermograms. These spectrum images are used as an input to pre-trained convolutional neural network, AlexNet and LeNet. The thermograms of potatoes and guavas are generated using thermal scanner and a data set consisting of 3200 and 4166 thermograms in constituted respectively. It was observed that LeNet based convolutional neural network performed better as compared to AlexNet for classifying thermographs of potatoes and guavas based on Variational mode decomposition technique of pre-processing of images.*

***Keywords:*** *Convolutional Neural Network; Infrared Imaging; Quality Assessment; Thermograms; Variational Mode Decomposition;*

## I. INTRODUCTION

Global population is increasing at a very rapid rate and this causes a burden over the agricultural industry to provide food supplies to each person. Consumers often obtain raw fruits and vegetables, or some by products including pickles, jams, and other processed food items manufactured by the food processing industries. It also becomes essential to make sure that the food item that is reaching the consumer is not defective, bruised or diseased. If not ensured, it may impact negatively to the health of the consumers. Segregating defective food items from wholesome items, or differentiating one type of cultivars of items from another, or classifying these food items on the basis of their nutrient content are some of the applications where vision based approach is an important aspect.

Traditionally, these process were manual in nature. However, due to several drawbacks of manual nature of segregation, detection, and classification process, there is an impending need to design a system which is able to automatically do these processes. For decades, researchers are making an effort to develop a vision based automated system for quality assessment of agricultural based products. In this paper, an effective technique is used for segregation of defective potatoes from the wholesome potatoes. To make the system scalable to different agricultural items, the same model is also considered for classifying guavas into three categories including mature, half-mature, and immature.

In this paper, two agricultural items are considered including potato which is abundantly available item and belongs to tuber family in vegetable categories. Also, guavas belonging to fruit category are considered which is a very rich source of minerals and fibers, and abundantly consumed globally. Thermal images of both these agricultural items are obtained using the thermal scanner, TiS45 by FLUKE Corporation, USA. These thermograms thus obtained are stored in a dataset. For pre-processing of these thermograms, a novel signal decomposition technique, Variational Mode Decomposition (VMD) is employed. This technique generates several modes of the

 Received-16-03-2025, Accepted- 02-04-2025

input thermograms. Subsequently, a Short Time Fourier Transform (STFT) of each of these modes are obtained and combined together to generate the pre-processed images which are used as an input to the Convolutional Neural Network (CNN). In this study, two pre-trained CNN based models including AlexNet and LeNet are employed for classification process. The results of these two models are compared among themselves. Also, images classification is performed without using the VMD based pre-processing technique and these results are further compared to observe the impact of VMD based pre-processing technique on the classification process.

This paper is structured in five section. The first section is the introduction to the problem statement and motivation behind undertaking this study. The second section presents the brief a quick overview towards other research performed using similar methodologies. The third section presents the detailed methodology and algorithms used to achieve the results. Fourth section discusses the results obtained in the third section. The paper is finally concluded in the fifth section followed by references used in this study.

## II. LITERATURE REVIEW

Variational Mode Decomposition (VMD) and Short-Time Fourier Transform (STFT) are both powerful techniques used for image processing. VMD is a relatively new method that decomposes an image into a set of band-limited intrinsic mode functions (IMFs) through an optimization problem, while STFT provides a time-frequency representation of a signal by applying the Fourier transform to short overlapping segments of the signal. Both methods have their unique advantages and applications, which are explored in various research contexts. VMD is able to decompose an image into a quasi-orthogonal Intrinsic Mode Function (IMF) which makes it more robust is removal of noise from the image [1-2]. In [1] VMD has been applied for seismic time frequency analysis providing higher spatial and spectral resolution making it more sensitive to variations in the image. In [3] STFT is used for analyzing the time frequency characteristics of signals by segmenting the image into short windows and applying Fourier transform. The proposed short-time multivariate Variational Mode Decomposition (STMVMD) method effectively identifies instantaneous frequencies (IFs) of time-varying structures using output-only measurements, overcoming the limitations of traditional VMD methods that rely on narrowband assumptions and struggle with non-stationary signals containing closely-spaced wideband intrinsic mode functions (IMFs) [4].

In [5], an introduction is presented for a de-noising method utilizing VMD and the Cramer Von Mises statistic, demonstrating its effectiveness over traditional EMD in segregating noise and preserving signal integrity. In [6], a comprehensive analysis compares data-driven signal decomposition methods, including VMD, EMD, SST, and SSA, providing insights into their performance, advantages, and limitations in various applications. These are some of the research which gives a brief idea about the novelty of VMD and STFT based image processing techniques.

In [7], a review provides an extensive overview of machine learning applications in agriculture, highlighting various techniques used for quality assessment of agricultural products. In [8], a study presents a machine learning approach for seed quality assessment, utilizing image processing to analyze attributes like perimeter, area, and texture for classification. In [9], a review discusses the application of machine learning in analyzing hyperspectral images for non-destructive quality assessment of food products, emphasizing rapid and accurate evaluation methods. Author in [10], explores the integration of AI and machine learning in food quality control, detailing techniques like computer vision and deep learning for defect detection and consistency evaluation. Another review highlights AI applications in non-destructive quality inspection of food and agricultural products, focusing on expert systems, artificial neural networks, and fuzzy logic [11]. The article in [12] discusses current trends in machine learning applications for agricultural product quality assessment, discussing challenges and future directions in the field.

## III. METHODOLOGY

This section presents the detail discussion about different components used for completing the required study. A detail description of dataset acquisition is presented along with proper explanations of pre-processing techniques and classification models. The flow diagram of the methodology used in this study is given in figure 1.

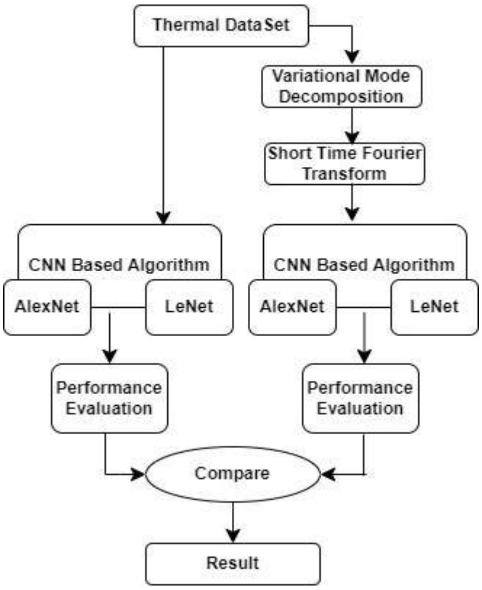


**Figure 1: Flow Diagram of Methodology**

3.1 **Data Set**

In this section detail description of dataset generation is presented. In this study, two agricultural items, potatoes and guavas are used to develop the dataset.

Dataset – I

The first dataset is developed consisting of thermal images of potatoes. Two qualities of potatoes including defective and wholesome are used. 200 samples of potatoes are used for image acquisition. Of these 200 samples, 80 samples belong to defective potatoes and remaining 120 samples consist of wholesome qualities. For thermal image acquisition, thermal scanner was employed with model ID TiS45 by FLUKE Corporation, USA. The sample was placed on a stationary mount. The lens of the camera is placed at a distance of 25 cm to acquire sharp images from four different direction including front, left, right, and top. Therefore, from a single sample of potato, four thermal images are acquired. This resulted in 800 thermal images from 200 samples of potatoes. To further enrich the dataset and make the model more robust towards classification, image augmentation technique is used. Three different image augmentation technique is employed resulting in dataset consisting of 3200 thermal images. Of these, 1280 thermal images belongs to defective samples and remaining 1920 thermal images belong to wholesome quality of potatoes. A sample dataset is presented in figure 2.

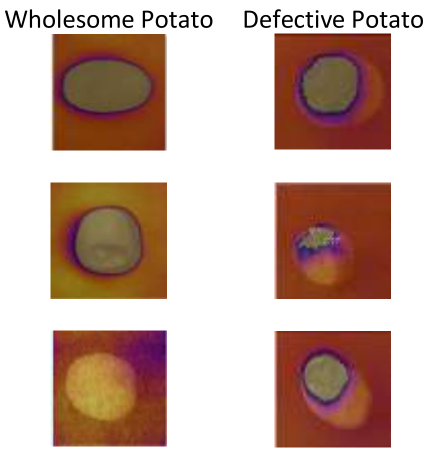


**Figure 2: Sample Dataset Consisting of Defective and Wholesome Quality of Potatoes**

Dataset – II

Thermal images of guava are acquired from the online available dataset at Mendeley [13]. This dataset consist of three categories of guava based on its maturity level including mature, half-mature, and immature. From the online available dataset 353, 326, and 362 thermal images of mature, half-mature, and immature are used to develop the dataset used in this study. The thermal scanner used to capture the thermal images is FLIR One Pro. Active thermography is used in this study, where a stream of hot air at an angle of 45° is used to increase the temperature of the object. Once the hot air stream is removed, the temperature of the object starts to fall. It is this falling temperature gradient which is captured by the thermal scanner to generate the thermograph. Once the thermal images are acquired from the dataset, using the three image augmentation technique, the size of the dataset is further increased. This resulted in 1414, 1304, and 1448 thermal images of mature, half-mature, and immature categories of guava. Thus, the total size of the dataset is 4166 thermal images. A sample of the dataset is presented in figure 3.

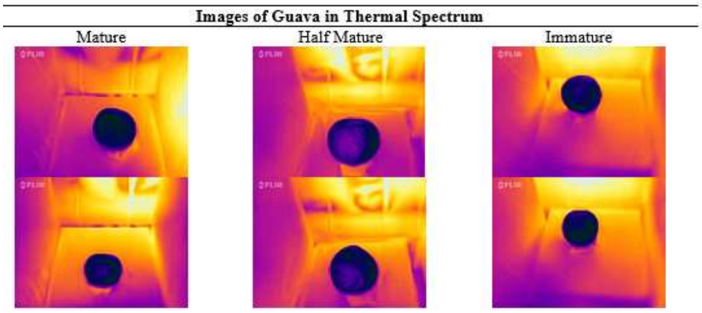


**Figure 3: Sample Dataset Consisting of Three Grade of Maturity of Guava Including Mature, Half-Mature and Immature.**

### 3.2 Variational Mode Decomposition

Variational Model Decomposition (VMD) is a novel image decomposition method which adaptively decomposes the image into several modes. Each mode is limited by certain central frequency and generates an image spectrum in frequency domain in a band limited condition. This technique employees optimized framework to generate non-overlapping. The mathematical relationship explain the complexity of VMD process is presented in equation 1.

$$\min_{\{u_k\},\{w_k\}} \left\{ \sum_{k=1}^{K} \left\| \nabla \left[ \left( \delta(x,y) + \frac{j}{\pi\sqrt{x^2+y^2}} \right) * u_k(x,y).\, e^{-j(w_{kx}x+w_{ky}y)} \right] \right\|_2^2 \right\} \quad (1)$$

Where f(x,y) is an input thermal image, $w_k(x,y)$ is the kth Mode in 2D, $w_k = w_{kx}, w_{ky}$is the center frequency in 2D, ∇ is the gradient operator, $\delta(x,y)$ is the 2D direc delta function, and * is the convolutional operator. In this study, the raw thermal image is divided into 6 modes.

### 3.3 Short Time Fourier Transform

Short Time Fourier Transform (STFT) is a technique where the signal is converted from time domain to frequency domain. It is based on the fundamentals of Fast Fourier Transform (FFT). In this study, after finding the six modes of the decomposed raw thermal image, STFT is computed of each mode to convert them to the frequency spectrum. The computed FFT of successive data-frame can be fused togather to obtain the time-frequency representation of the original signal. The mathematical representation of the STFT is given in equation 2.

$$STFT_f(u,v,\xi_x,\xi_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y).\,w(x-u,y-v).\,e^{-j2\pi(\xi_x x+\xi_x y)}\,dxdy \quad (2)$$

In the above equation, f(x,y) is the 2D decomposed thermal image, $w(x-u,y-v)$ is the window function centered around u,v, $\xi_x,\xi_y$ is the spatial frequency component along the respective axis, u,v are the spatial location where the window function is applied. The window function defines a particular region of the image that is being analyzed. This region is converted into frequency domain using Fourier kernel function given by $e^{-j2\pi(\xi_x x+\xi_x y)}$

### 3.4 Thermogram Spectrum Image

The thermal image obtained of the potatoes and guavas was decomposed into 6 VMD modes given by $w_k(x, y)$ where k is the number of mode which is 6 in this study. The Short Time Fourier Transform of kth mode is converted into an equivalent thermal image of size M x L where the values of M and L are given by

$$M = \frac{N+p}{2} + 1 \quad (3)$$

where p is the zero padding value and L is given by

$$L = \frac{N-w}{w-o} \quad (4)$$

Where w represents the window size and o represents the over lapping windows.

By stacking these 6 modes the spectrum size resulted in 6M x L. The window size of and 'p' = 10 was performed for all the thermal images in the dataset. Considering a 75% overlap, the value of 'o' = 35.

### 3.5 CNN based classification Model

Convolutional Neural Networks (CNNs) are a type of deep learning model used to efficiently and accurately classify structured data, particularly images. They draw inspiration from the human visual system and use hierarchical feature extraction to identify patterns, edges, textures, and higher-level features in raw input data. A CNN is made up of several layers, including convolutional, pooling, and fully connected layers. The convolutional layers apply filters to the input image, resulting in feature maps created by convolving small, localized regions. These filters learn to recognize low-level patterns like edges in early layers, as well as more complex structures like shapes and objects in deeper layers. Pooling layers, such as max pooling, down sample feature maps to reduce dimensionality while retaining essential information. Finally, fully connected layers combine these features to generate predictions. The classification process begins with feeding an image into a CNN, which extracts features using convolutional and pooling layers. Fully connected layers receive the extracted features and use a Softmax activation function to generate probabilities for each class. CNNs are widely used in a variety of classification tasks, such as image recognition (e.g., object detection), medical imaging (e.g., disease diagnosis), and natural language processing (text classification). CNNs' ability to learn features automatically from raw data has made them extremely useful in fields that require precise classification. In this study two pre-trained model including AlexNet and LeNet-5 are used for classifying potatoes into defective or wholesome category or grading f guava on the basis of their maturity level including mature, half-mature, and immature.

AlexNet is a complex CNN-based architecture with eight layers. These eight layers comprise five convolutional layers for feature extraction and three fully connected layers for classification. It employs ReLU activation functions, which speed up training by introducing non-linearity while mitigating the vanishing gradient problem. The block diagram of architecture of AlexNet is given in figure 4(a). LeNet is one of the first CNN-based architectures designed for classifying objects into different categories. The architecture is simple and efficient, with seven layers: convolutional layers, subsampling (average pooling) layers, and fully connected layers. Figure 4 (b) shows a block diagram of the architecture. Computationally, AlexNet is much more complex as compared to LeNet based classification architecture.

| LeNet [14] | AlexNet [15] |
|---|---|
| Input Image | Input |
| Conv2D | Conv2D |
| Average Pooling2D | MaxPooling2D |
| Conv2D | BatchNorm |
| Average Pooling2D | Conv2D |
| Dense Layer | MaxPooling2D |
| Dense Layer | BatchNorm |
| Dense Layer | Conv2D |
| | MaxPooling2D |
| | BatchNorm |
| | Conv2D |
| | BatchNorm |
| | Conv2D |
| | BatchNorm |
| | Conv2D |
| | MaxPooling2D |
| | BatchNorm |
| | Flatten |
| | Dense |
| | Dropout |
| | BatchNorm |
| | Dense |
| | Dropout |
| | BatchNorm |
| | Dense |
| | Dropout |
| | BatchNorm |
| | Dense |

**Figure 4: block diagram of architecture of (a) LeNet and (b) AlexNet**

### 3.6 Performance Evaluation Parameters

To compare the different classification models, a confusion matrix is obtained consisting of four characteristic parameters including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on these parameters, performance evaluation parameters are evaluated including Accuracy, Precision, Recall, F-1 Score. Using the evaluated value of accuracy, the error value in terms of Root Mean Square Error (RMSE) is calculated giving a normalized measure of the classification error. It is determined based on error rate which is the difference between unity and the accuracy. The mathematical relationship of these formula are presented in table 1.

**Table 1: Mathematical Relationship of Performance Evaluation Parameters**

| Performance Evaluation Parameter | Mathematical Relation |
|---|---|
| Accuracy | $A = \frac{TP + TN}{TP + TN + FP + FN}$ |
| Precision | $P = \frac{TP}{TP + FP}$ |
| Recall | $R = \frac{TP}{TP + FN}$ |
| F-1 Score | $F - 1\ Score = \frac{2P.R}{P + R}$ |

## IV. METHODOLOGY

In this section, a detail discussion of simulated results of different architecture is presented and their interpretations is presented. In this study there are four different situations that are analyzed to assess the effectiveness of the proposed image pre-processing technique employed for classifying and grading potatoes and guavas respectively as per the requirements. The first situation is the classification of potatoes into different quality based categories using CNN based architecture. In this case, the input is the raw unprocessed thermal images of potatoes. The second case is the grading of guava into three categories based on their maturity level including mature, half-mature, and immature. In this case also, the grading is performed using CNN based architecture. The third case is similar to the first case, however, instead of using unprocessed thermal images of potatoes, VMD and STFT based pre-processing technique is employed. It is this pre-processed thermal images that are used by the AlexNet and LeNet which are based on CNN based architecture. In the fourth case, which is similar to second case, grading of guava is performed where instead of unprocessed thermal images, VMD and STFT are used to process the image. These processed images are used as an input by AlexNet and LeNet architectures separately to grade guava into their respective classes.

Case I is the model for classifying potatoes into their respective categories based on their quality. Two qualities are considered including defective and wholesome. In this case, two CNN based classification models are used including AlexNet and LeNet. Five performance evaluation parameters are evaluated for each classification models and their determined values are presented in table 2. From this table it is observed that the performance of LeNet based architecture is slightly better as compared to AlexNet based classification model. The achieved accuracy for classifying for AlexNet and LeNet is 82.3% and 86% respectively. For other parameters used to evaluate the performance of the classification model, including precision, Recall, and F-1Score, the values obtained for LeNet is higher as compared to AlexNet. The observed values for LeNet for these parameters for wholesome class is 87.2%, 89.9% and 88.5% respectively.

**Table 2: Performance Evaluation of AlexNet and LeNet for Classifying Potatoes as Defective or Wholesome**

| Classification model | Quality Class of Potato | Performance Evaluation Parameters | | | |
|---|---|---|---|---|---|
| | | P | R | F-1 | A |
| AlexNet | Defective | 0.799 | 0.746 | 0.772 | 0.823 |
| | Wholesome | 0.838 | 0.875 | 0.856 | 0.823 |
| LeNet | Defective | 0.841 | 0.802 | 0.821 | 0.860 |
| | Wholesome | 0.872 | 0.899 | 0.885 | 0.860 |

In another case, the same dataset is used for classifying potatoes into their respective qualities. However, for pre-processing technique, the raw thermogram is first divided into six modes using Variational Mode Decomposition Technique. Once these modes are generated, Short Time Fourier Transform is determined for each mode. Thus, the six STFT obtained for each mode is combined to generate the processed image. These processed images are used as input to the AlexNet and LeNet based classification models. The performance of these models using VMD + STFT based processed images is presented in tabular form in table 3. As compared to values achieved in the previous model, values presented in table 2, the performance of the classification model is improved by almost 5%. The accuracy of AlexNet when using proposed pre-processing technique is 88.8% as compared to 82.3% when proposed pre-processing technique is not used. Also, for LeNet based classification model, the classification accuracy is increased from 86% to 82.6% by using the proposed pre-processing technique. As observed from table 3, the performance of LeNet is better as compared to AlexNet. The accuracy for LeNet based Architecture is much better as compared to AlexNet with 92.6% and 88.8% respectively. In other parameters also, LeNet out performs AlexNet for classifying potatoes into their respective classes based on quality such as defective and wholesome.

**Table 3: Performance Evaluation of VMD+STFT+CNN Based Architecture for Classifying Potatoes as Defective or Wholesome**

| Classification model | Quality Class of Potato | Performance Evaluation Parameters | | | |
|---|---|---|---|---|---|
| | | P | R | F-1 | A |
| AlexNet | Defective | 0.857 | 0.863 | 0.860 | 0.888 |
| | Wholesome | 0.909 | 0.905 | 0.907 | 0.888 |
| LeNet | Defective | 0.895 | 0.917 | 0.906 | 0.926 |
| | Wholesome | 0.946 | 0.931 | 0.939 | 0.926 |

In other scenario, the AlexNet and LeNet based architecture are used for grading guava based on their maturity levels. Three maturity levels are considered including mature, half-mature, and immature. In the first scenario of grading guava on the basis of maturity level, no preprocessing is performed on the thermographs. The raw thermographs are directly used to the categorized into their respective classes. For categorization, AlexNet and LeNet based classification architecture is considered. The achieved output results in terms of performance evaluation parameters is presented in table 4. It is observed that in AlexNet, the accuracy in categorizing guava into mature class is 88.2% which is increased to 90.8% while using LeNet based classification model. Also, the other parameters such as precision, recall, and F-1 Score, there is an increase in the achieved values for mature class grading while using AlexNet and LeNet based classification models. Value of precision increased from 80.7% to 83.6%, value for recall increased from 84% to 88.7%, and F-1 score increased from 82.3% to 86.1% respectively. The same degree of increment can be observed in other two classes such as half-mature and immature.

**Table 4: Performance Evaluation of AlexNet and LeNet for Grading Guava on The Basis of Maturity**

| Classification Model | Guava Maturity Level | Performance Evaluation Parameter | | | |
|---|---|---|---|---|---|
| | | P | R | F-1 | A |
| AlexNet | Mature | 0.807 | 0.840 | 0.823 | 0.882 |
| | Half-Mature | 0.810 | 0.724 | 0.765 | 0.844 |
| | Immature | 0.801 | 0.860 | 0.829 | 0.886 |
| LeNet | Mature | 0.836 | 0.887 | 0.861 | 0.908 |
| | Half-Mature | 0.846 | 0.756 | 0.798 | 0.866 |
| | Immature | 0.839 | 0.884 | 0.861 | 0.906 |

The values presented in table 4 shows the performance of classification model when the input thermographs are unprocessed. However, when the same raw images are pre-processed using Variational model decomposition technique and short time Fourier transform, the performance of AlexNet and LeNet based guava grading algorithm increases. It can be observed that while using unprocessed image, the accuracy of predicting mature class of guava using AlexNet based model is 88.2% which increases to 91.7% when VMD + STFT based pre-processing of thermograms are used. Similarly there is increase in values for precision, recall, and F-1 score for mature guava grading from 80.7% to 83.5%, 84% to 91.4%, and 82.3% to 87.3% respectively. These values are presented in table 5.

**Table 5: Performance Evaluation of VMD+STFT+CNN Based Architecture for Grading Guava on the Basis of Maturity**

| Classification Model | Guava Maturity Level | Performance Evaluation Parameter | | | |
|---|---|---|---|---|---|
| | | P | R | F-1 | A |
| AlexNet | Mature | 0.835 | 0.914 | 0.873 | 0.917 |
| | Half-Mature | 0.851 | 0.759 | 0.803 | 0.869 |
| | Immature | 0.851 | 0.873 | 0.862 | 0.905 |
| LeNet | Mature | 0.888 | 0.904 | 0.896 | 0.930 |
| | Half-Mature | 0.887 | 0.849 | 0.868 | 0.915 |
| | Immature | 0.891 | 0.912 | 0.901 | 0.932 |

A complete comparative analysis is presented in tabular manner in table 6 where a comparison is made in the performance of AlexNet and LeNet based classification models with and without using pre-processed images. It is observed that LeNet based classification model using proposed pre-processing model performs better for classifying potatoes into their respective quality classes and grading guava into their respective maturity classes.

**Table 6: Comparative Analysis of Different Architecture**

| | Classification Models | P | R | F-1 | A |
|---|---|---|---|---|---|
| Maturity Level Grading of Guava | VMD+STFT + AlexNet | 0.845 | 0.848 | 0.846 | 0.897 |
| | AlexNet | 0.806 | 0.808 | 0.805 | 0.868 |
| | VMD + STFT + LeNet | 0.888 | 0.888 | 0.888 | 0.925 |
| | LeNet | 0.84 | 0.842 | 0.841 | 0.893 |
| Quality Assessment of Potato | VMD+STFT + AlexNet | 0.883 | 0.884 | 0.883 | 0.888 |
| | AlexNet | 0.818 | 0.811 | 0.814 | 0.823 |
| | VMD + STFT + LeNet | 0.92 | 0.924 | 0.922 | 0.926 |
| | LeNet | 0.856 | 0.849 | 0.853 | 0.86 |

## V. CONCLUSIONS

In this study, two separate thermally generated agricultural products are used to generate the dataset. The two items include potato and guava. For potato, there were two categories for quality assessment including defective and wholesome. Whereas for guava there were three categories for grading them as per the maturity levels including mature, half-mature, and immature. For their classification, two CNN based pre-trained networks including AlexNet and LeNet was used. The entire study was divided into two parts. During the first part, unprocessed thermal images were used as an input to the AlexNet and LeNet based classification models and in the second part a proposed pre-processing techniques was used to process the images. The proposed pre-processing technique is the combination of Variational Mode Decomposition (VMD) which decomposed the thermal image into six modes. These decomposed images were converted into to frequency domain using Short Time Fourier Transform (STFT). The achieved six STFT converted thermal modes of the decomposed thermal image were fused together to generate the processed image. These processed images were used for classification using the same two CNN based pre-trained network including AlexNet and LeNet. It was observed that that LeNet based classification model performed better as compared to AlexNet based model. Also, the proposed technique for pre-processing the image improved the models performance. Considering the quality assessment of Potato into defective or wholesome, when using unprocessed image, the achieved accuracy for AlexNet and LeNet based model was 82.3% and 86% respectively. However, using the processed image using VMD and STFT, the accuracy for AlexNet and LeNet based classification model increased to 88.8% and 92.6% respectively. The same increase in model performance can be observed for grading of guava into three different maturity levels. Also, for other classification model performance evaluation parameters, the values were improved when VMD and STFT based processed images were considered to as an input image to the pre-trained CNN based classification models. Based on the finding of this study, it is concluded that pre-processing images using image decomposition technique improves the performance of classification models.

## VI. REFERENCES


1. Qin, Kangzhi, Yu Zhang, and Jinlong Li. "Application of variational mode decomposition method in Fourier transform profilometry." 9th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optical Test, Measurement Technology, and Equipment. Vol. 10839. SPIE, 2019.

2. Chen, Wei, and Yong Zhang. "A queued Variational Mode Decomposition method." Journal of the Franklin Institute (2024): 106965.
3. Liu, Shuaishuai, et al. "A novel instantaneous frequency estimation method for operational time-varying systems using short-time multivariate variational mode decomposition." Journal of Vibration and Control 29.17-18 (2023): 4046-4058.
4. Liu, Wei, Siyuan Cao, and Yangkang Chen. "Applications of variational mode decomposition in seismic time-frequency analysis." Geophysics 81.5 (2016): V365-V378.
5. Naveed, Khuram, et al. "A statistical approach to signal denoising based on data-driven multiscale representation." Digital Signal Processing 108 (2021): 102896.
6. Eriksen, Thomas. "Data-driven Signal Decomposition Approaches: A Comparative Analysis." arXiv preprint arXiv:2208.10874 (2022).
7. Benos, Lefteris, et al. "Machine learning in agriculture: A comprehensive updated review." Sensors 21.11 (2021): 3758.
8. Kini MG, Ramanath, and Rekha Bhandarkar. "Quality assessment of seed using supervised machine learning technique." Journal of The Institution of Engineers (India): Series B 104.4 (2023): 901-909.
9. Saha, Dhritiman, and Annamalai Manickavasagan. "Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review." Current Research in Food Science 4 (2021): 28-44.
10. Chhetri, Krishna Bahadur. "Applications of Artificial Intelligence and Machine Learning in Food Quality Control and Safety Assessment." Food Engineering Reviews 16.1 (2024): 1-21.
11. Ali, Maimunah Mohd, et al. "Quality inspection of food and agricultural products using artificial intelligence." Advances in Agricultural and Food Research Journal 2.2 (2021).
12. Araújo, Sara Oleiro, et al. "Machine learning applications in agriculture: current trends, challenges, and future perspectives." Agronomy 13.12 (2023): 2976.
13. Pathmanaban, P., B. K. Gnanavel, and Shanmuga Sundaram Anandan. "Comprehensive guava fruit data set: digital and thermal images for analysis and classification." Data in Brief 50 (2023): 109486.
14. LeCun Y, Bottou L, Bengio Y, Hafner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278–324
15. Krizhevsky A, Sutskever I, Hinton G.E. Imagenet classifcation with deep convolutional neural networks. In: Advances in neural information processing systems, 2012;pp. 1097–1105.