Data Science and Engineering (DSE) Record, Volume 4, issue 1.

Durian ripeness prediction through knocking sound

Warut Sanwibhuk [1] and Pruet Boonma [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
warut_k@cmu.ac.th

Abstract. This independent research aimed to develop a model for predicting the ripeness level of durians based on knocking sounds. The ripeness levels were categorized into 3 levels are raw, unripe and ripe. Each level have unique sound responsed when knocking. To achieve the highest accuracy in model development, the researcher have compared the data feature extraction with 3 methods: Mel-Spectrogram, Short-time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCC). In the model development phase, The Convolutional Neural Network (CNN) algorithm was selected bulding the prediction model. To evaluate the performance of the developed model, accuracy and F1 Score was measured by comparing along to data feature extraction 3 methods. It was found that the Short-time Fourier Transform (STFT) method yielded the highest both accuracy value and F1 Score. The resulted by training dataset give 99% of accuracy value and 94% of F1 Score. Along with the blind testing data give 99% of accuracy value and 92% of F1 Score.

Keywords: Sound Classification, STFT, MFCC, Convolutional Neural Network (CNN), Durain

1 Introduction

Durian is a crucial economic fruit and the number one exported fruit in Thailand. In the year 2021, the export value reached 109,205 million Baht [12]. One of the challenges in exporting durian to foreign destinations is the extended transportation time, which can take several days. Studies reveal that the weight of durian decreases by 4% a day after harvesting and can decrease by up to 27% on the fifth day [11]. This necessitates careful planning for durian farmers and exporters to ensure that the fruit reaches its destination at the expected ripeness level with minimal damage and weight loss. As a result of this export challenge, a significant occupation has emerged – durian sorter experts. The increasing demand for durian consumption has led to a short-age of experts, and the high cost of hiring these experts has become apparent. According to interviews, the average monthly income for durian quality inspectors ranges from 50,000 to 100,000 Thai Baht [13]. The study suggests that both external physical characteristics and knocking sounds can be used to assess durian ripeness. Knockinng the durian shell to listen for the sound of hollowness and density is identified as an effective method for durian ripeness classification.

The study further explores the classification of durian ripeness levels based on knocking sounds. Generally, 3 ripeness levels can be distinguished: raw, unripe, and ripe. Raw durians produce a dense sound when tapped, while ripe durians produce a resonant sound due to water excretion and the transformation of starch into sugar, causing the flesh to contract within the shell and creating a hollow space, resulting in a resonant sound. To address these challenges, researchers collaborated with experts to develop a model for predicting durian ripeness levels based on tapping sounds.

2 Literature Review

2.1 Analyzing Knocking Sounds to Evaluate the Ripeness Quality of Watermelons through Sound Frequencies [14]

This research focuses on the analysis of tapping sounds to assess the ripeness of watermelons using sound frequency. It leverages local wisdom to evaluate or predict the rawness or ripeness of fruits by tapping without damaging the fruit. The experimental sample for this research consists of 200 watermelons of the eating-recommended variety, weighing 2-4 kilograms each. The tapping was conducted, and sound recordings were made using a smartphone. The recorded sounds were then processed using Fast Fourier Transform to determine the frequency range. Subse-quently, the sound was analyzed using Probability Density Function, and Bayesian decision theory was employed to predict ripeness.

In the experimental phase, it was found that a frequency range (Fmax) of 32 hertz is suitable for predicting the ripeness of watermelons. The overall accuracy of the prediction was 73.5%, with individual accuracies of 77% for ripe watermelons and 70% for raw ones.

2.2 Durian Ripeness Classification from the Knocking Sounds Using CNN [3]

Another aspect of the research involves tapping sounds of durians to predict their ripeness levels, categorizing them as ripe, raw, or half-ripe. A sample of 30 durians was used for experimentation. Sound recordings were made using a smartphone, and the files were segmented into 3-second intervals. The next step involved feature extraction using Mel-frequency cepstral coefficient spectrogram (MFCC), and a Convolutional Neural Network (CNN) algorithm was employed to build a model for prediction. The results showed an accuracy of 90.78% for the validation set and 89.74% for the testing set.

3 Materials and Method

3.1 Define the definition and characteristics of durian knocking sounds.

Define the pronunciation of sound when knocking on the durian into 3 level. Each level have dirrence ripness level as shown in table 1.

Table 1. Showns definition and characteristics of durian knocking sounds.

Pronunciation (TH)	Rep Level (Class)	Description
Pǣk	Raw	Strong sound
Pæa	Unripe	Semi Strong sound
Phú	Rip	Resonant sound

The sound of knocking on durian falls into the category of impact or collision sounds. It is a sound that occurs at a specific moment and ends within 1 second, as exemplified by the waveform of various sound levels as shown in figtue 1.

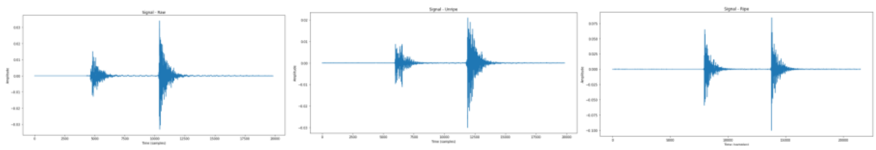


Figure 1. Showns sound signal by knocking on durian.

3.2 Data Collected

The population in this research consist of 300 durians for training dataset and 60 for blind testing dataset. Any level or class have equal of qantity of durain, see in table 2.

Table 2. Showns the population quantity of durians for experiemental.

Ripness Level	Training Dataset	Blind Testing Dataset
Raw	100	20
Unripe	100	20
Rip	100	20

During data collection, it was observed that experts knocking durians would knock twice in each round. The sound waves occur and end within 1 second. Consequently, the researchers designed the sound data collection using mobile phones to continuously record the knocking sound until the data collection is completed. Then, import the audio files into the sound processing program (WavePad Sound Editor) and proceed to cut and split the audio files into 1-second segments. Each file consists of the knocking sound occurring 2-4 times.

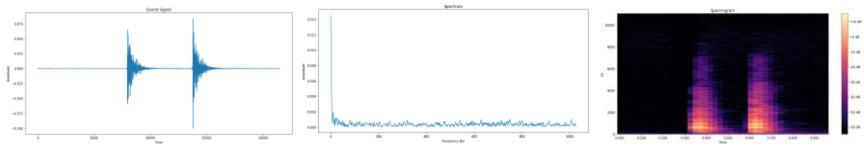
3.3 Feature extraction by using the Librosa package. [5][6][7]

This research involves importing data in the form of audio files. Therefore, it is necessary to process the audio data and extract features using a Python library called Librosa. This is done to study and compare the predictive performance of models. In this research, 3 methods of data feature extraction have been selected: Mel-Spectrogram, Short-time Fourier Transform (STFT), and Mel-Frequency Cepstral Coefficients (MFCC).

The original knocking sound of durians obtained from the audio recording is in the form of an analog signal, representing a continuous-time signal, as shown in Figure 2(a). The x-axis represents time, and the y-axis represents the amplitude, which is the height of the sound wave or the amplitude.

In other words, the original sound signal is a waveform that exists in the time domain. Therefore, when extracting data from the sound signal, it is necessary to transform it from the time domain to the frequency domain. To achieve this, a mathematical formula called Fourier Transform or Fast Fourier Transform is used to convert from the time domain to the frequency domain. The result, known as the spectrum, separates the signal into individual frequencies and their corresponding amplitudes, as depicted in Figure 2(b).

After obtaining the spectrum, it is typically transformed into a spectrogram for better understanding. The spectrogram, shown in Figure 2(c), illustrates the sound intensity at different frequencies over time. The y-axis is calculated on a Log Scale, the x-axis represents time, and the color is represented in decibels.



(a) (b) (c)

Figure 2. Showns the process of transform sound signal from time domain to frequency domain (Spectogram).

1) **Feature Extraction Method : Mel-Spectrogram [6] [10]**
   The Mel-Spectrogram data extraction method involves taking the signal of sound waves, transforming it into a spectrum, and calculating it into a spectrogram using Fast Fourier Transform (FFT). The frequencies are then converted onto the Mel Scale. A notable feature of this method is its ability to reduce noise or eliminate high-frequency sound waves.

2) **Feature Extraction Method : Short-time Fourier Transform (STFT) [7][8]**
The Short-Time Fourier Transform (STFT) involves mathematical calculations to analyze and extract features of sound signals for use in sound recognition systems. Instead of analyzing the entire signal, it analyzes specific time intervals. The short-time Fourier transform uses a window function to divide the signal into short time intervals, and then calculates the Fourier transform for each interval from the time domain to the frequency domain. This is done to observe changes in the signal over short time intervals, providing a detailed representation of the sound. The size of the window for division can be specified during the data extraction process.

3) **Feature Extraction Method : Mel-Frequency Cepstral Coefficients (MFCC) [5]**
In the feature extraction process of Mel-Frequency Cepstral Coefficients (MFCC), numerical values are computed to represent the characteristics of each sound signal. The result is a set of numerical values representing the sound signal. This method is beneficial for reducing unnecessary data while preserving essential information. This approach builds upon the Short-Time Fourier Transform (STFT) and Mel-Spectrogram methods. It starts by dividing the sound signal into short time intervals using a window function to analyze each signal in each time interval within a sub-frame. The spectrogram is then calculated, displaying the frequency of the signal at each time, and it is transformed from the normal frequency scale to the Mel frequency scale. The Mel-Frequency Spectrogram is then used to calculate the MFCC values, which are mathematical values that represent the specific characteristics of the sound in each frame.

3.4 Develop a model using Convolutional Neural Network (CNN) [8].

This step, the researchers choose to use a Convolutional Neural Network (CNN) to develop a model for classifying or predicting the ripeness level of durians based on knocking sounds. The process involves importing audio data, extracting features from the audio data using the three methods mentioned in section 3.3, designing the data layers for the sound recognition system, and evaluating the model's performance. The development steps are showns below figture.

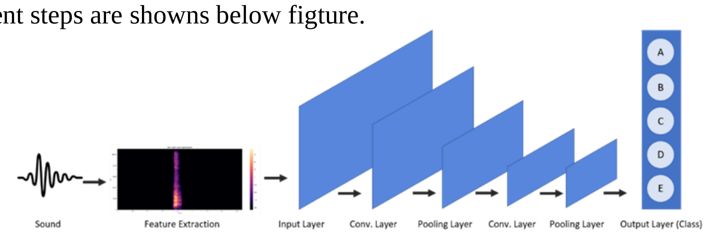


Figure 3. Showns the development steps of building prediction model.

4 Experimental Results

In evaluating the performance of the durian ripeness classification model in this experiment, the accuracy both of the training data and blind testing data will be measured. This involves comparing the accuracy and F1 Score values obtained from each of the three data extraction methods.

4.1 Training Data - Accuracy Result

According to training data set, the the accuracy value, validation accuracy value, loss value and validation loss value are showns in Table 3. There are comparing between 3 data feature extraction methods. The method STFT given hight accuracy than others with 99% when set epochs to 150 during traing process.

Table 3 . Showns accuracry result by training data set.

Feature Extraction Method	Scale	Epochs			
		30	50	100	150
Mel-Spectrogram	Accuracy	0.75	0.84	0.91	0.96
	Val Accuracy	0.69	0.75	0.75	0.78
	Loss	0.59	0.41	0.20	0.11
	Val Loss	0.97	1.20	2.04	2.43
STFT	Accuracy	0.96	0.96	0.98	0.99
	Val Accuracy	0.91	0.90	0.93	0.96
	Loss	0.07	0.00	0.00	0.03
	Val Loss	0.28	0.49	0.07	0.04
MFCC	Accuracy	0.90	0.96	0.96	0.97
	Val Accuracy	0.78	0.85	0.80	0.75
	Loss	0.22	0.01	0.00	0.16
	Val Loss	0.75	0.64	0.86	1.25

4.2 Training Data – F1 Score Result

The F1 Score result in table 4 indicates the correctness of prediction by any class. Upon observation from the table, it is noted that all three methods yield higher values for both the raw and ripe classes compared to the unripe class. The average value for the STFT method is the highest, reaching 94%.

Table 4. Showns F1 Score result by training data set.

Feature Extraction Method	F1 Score			
	Raw	Unripe	Rip	Average
Mel-Spectrogram	0.77	0.64	0.76	0.72
STFT	0.95	0.93	0.94	0.94
MFCC	0.87	0.75	0.93	0.85

4.3 Blind Test Data – Accuracy Result

Blind testing dataset contain 20 samples for each class or 60 samples in totally. According to blind testing dataset 5 times the accuracy result as shown in table 5. Method STFT reaching hightest average accuracy is 99%.

Table 5. Showns accuracry result by blind test data set.

Feature Extraction Method	Accuracy Experiment					Average Accuracy
	1	**2**	**3**	**4**	**5**	
Mel-Spectrogram	0.76	0.77	0.81	0.72	0.73	0.76
STFT	0.98	1.0	1.0	0.99	0.97	0.99
MFCC	0.63	0.73	0.65	0.61	0.74	0.67

4.4 Blind Test Data – F1 Score

The hightest average F1 Score belong to method STFT reaching 92%.

Table 6. Showns F1 Score result by blind testing data set.

Feature Extraction Method	F1 Score			
	Raw	Unripe	Rip	Average
Mel-Spectrogram	0.77	0.57	0.71	0.67
STFT	0.94	0.85	0.96	0.92
MFCC	0.71	0.58	0.67	0.65

5 Conclustion

5.1 Summary

From the independent research conducted in this thesis on durian ripeness classification using knocking sounds, comparing the methods of extracting features from durian knocking sound signals affecting the classification model, it can be concluded that the Short Time Fourier Transform (STFT) method for sound data extraction performs the best in creating a model for durian ripeness classification. This conclusion is drawn based on accuracy measurements from both the trained data and blind test data, with an accuracy reaching up to 99%. This superior performance is observed compared to other methods of extracting features from sound signals. Through experimentation and observation, it was noted that accuracy values from training data were high, while accuracy from test data was relatively lower or had a significant gap between the two. In other words, in the early stages of training, the values were close, but they diverged as the training rounds increased. This implies that the extracted data may not be a good representative of the training population for the model to learn sound patterns effectively. In terms of F1 Score performance measurement, it was found that the average accuracy for the 3 classes was 94% for training data and 92% for blind testing dataset.

In terms of real-world application, it was discovered that artificial intelligence can effectively classify the ripeness levels of durians based on knocking sounds. This application has the potential to reduce costs by eliminating the need to hire experts for tapping and can increase the speed of durian sorting.

5.2 Suggestions

- Collecting sound data in diverse environments, such as areas with various ambient sounds like human activity, traffic, machinery, and a higher volume, to represent the actual population for building a model to classify the ripeness of durians. This model should be applicable in real-world environments at durian sorting centers.
- Increasing the granularity of classification responses for sorting durians, such as specifying ripeness levels like "consumable within 7 days," "consumable within 3 days," "consumable within n days," and "ready for consumption." This enhancement is based on recommendations from experts.
- Transforming the developed model into a durian- knocking device for user-friendly applications or integrating it into a conveyor-belt-based sorting system.

References

1. Abdul, Z. Kh., & Al-Talabani, A. K. (2022). Mel frequency cepstral coefficient and its applications: A Review. *IEEE Access*, *10*, 122136–122158. https://doi.org/10.1109/access.2022.3223444
2. Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, *65*(11), 1558–1564. https://doi.org/10.1109/proc.1977.10770
3. Kharamat, W., Wongsaisuwan, M., & Wattanamongkhol, N. (2020). Durian ripeness classification from the knocking sounds using convolutional neural network. *2020 8th International Electrical Engineering Congress (iEECON)*, 1–4. https://doi.org/10.1109/ieecon48109.2020.229571
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539
5. *Librosa Package – Mel Frequency Cepstral Coefficients*. librosa.feature.mfcc - librosa 0.10.1 documentation. (n.d.). https://librosa.org/doc/main/generated/librosa.feature.mfcc.html
6. *Librosa Package – Mel-Spectrogram*. librosa.feature.melspectrogram - librosa 0.10.1 documentation. (n.d.). https://librosa.org/doc/main/generated/librosa.feature.melspectrogram.html

7. *Librosa Package – Short-time Fourier transform*. librosa.stft - librosa 0.10.1 documentation. (n.d.). https://librosa.org/doc/main/generated/librosa.stft.html
8. McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and Music Signal Analysis in python. *Proceedings of the 14th Python in Science Conference*, 18–24. https://doi.org/10.25080/majora-7b98e3ed-003
9. Piczak, K. J. (2015). Environmental sound classification with Convolutional Neural Networks. *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. https://doi.org/10.1109/mlsp.2015.7324337
10. Roberts, L. (2022, August 17). *Understanding the mel spectrogram*. Medium. https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53
11. Sriyook, S., Siriatiwat, S., & Siriphanich, J. (1994). Durian fruit dehiscence—water status and ethylene. *HortScience*, *29*(10), 1195–1198. https://doi.org/10.21273/hortsci.29.10.1195
12. ฐานเศรษฐกิจดิจิทัล. (2022, May 7). *สถิติใหม่ "ทุเรียนไทย" ส่งออกจีน 1แสนล้านบาท ขยายตัว 68.4%*. thansettakij. https://www.thansettakij.com/economy/trade/524024
13. อมรินทร์ทีวี ช่อง 34 HD. (2019, May 13). *เปิดใจมือเคาะทุเรียนแยกผลสุก-ห่าม ทำรายได้หลักแสนต่อเดือน*. อมรินทร์ทีวี เอชดี ช่อง 34 - AMARIN TV HD. https://www.amarintv.com/news/detail/15054
14. เกษรินทร์ ชาวเกวียน, เสาวนีย์ สันติวงศ์, มนตรี โพธิโสโนทัย และ วรพล ลีลาเกียรติกุล . (2014). การวิเคราะห์เสียงเคาะเพื่อประเมินคุณภาพการสุกของผลไม้แตงโมด้วยความถี่เสียง. In *ECTI-CARD Proceedings 2014* (pp. 1–5). เชียงใหม่.