Current Applied Science and Technology Vol. 22 No. 1 (January-February 2022)

Research article

Acoustic Sensing for Quality Edible Evaluation of Sriracha Pineapple Using Convolutional Neural Network

Suphachai Phawiakkharakun, Pinyo Taeprasartsit and Sunee Pongpinigpinyo*

Department of Computing, Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand

Received: 23 January 2021, Revised: 22 April 2021, Accepted: 18 June 2021

DOI.....

Keywords

acoustic sensing;

quality edible evaluation;

deep learning;

pineapple juiciness;

classification;

MFCC;

Mel-Spectrogram

Abstract

Sriracha pineapple is a pineapple of the Smooth Cayenne variety and is one of Thailand's favorite tropical fruits. It is not only one of the most favorite agricultural products for the pineapple processing and canning industries but it is also widely consumed as a fresh fruit. Many exporters and consumers prefer to measure juiciness before processing food products and consuming, respectively. The traditional method used by pineapple sellers or farmers for fruit juiciness classification is to tap the pineapple using force impulse techniques with a rubber-tipped stick or a person's middle finger tapping. However, these traditional methods of classification require a lot of expertise and experience. Thus, the inspector's perception of accomplishing the classification process may be inclined to errors and uneven results. This paper proposes a combination of acoustic sensing and convolutional neural network (CNN). The tapping sound of 30 Sriracha pineapple samples using force impulse techniques was recorded on smartphone. The tapping sounds were processed into a juiciness classification system of three classes: (1) Juiciness 1, (2) Juiciness 2, and (3) Juiciness 3. The system involved a combination of acoustic sensing and CNN to compare the results between Mel Frequency Cepstral Coefficient (MFCC) and Mel-Spectrogram features extraction, with replication of the same CNN model, to evaluate the pineapple's edibility from its juiciness level. Experimental results showed that MFCC combined with CNN performed the best, with an accuracy of 96.67% and F1-score of 0.97. It outperformed the Mel-spectrogram combined with CNN.

*Corresponding author: Tel.: (+66) 861777572
E-mail: pongpinigpinyo_s@su.ac.th

1. Introduction

Ananas comosus, commonly known as pineapple, is one of the most popular tropical fruits. It contains many vitamins and minerals [1]. Pineapples can also be processed into a range of products including pineapple juice, canned pineapple, pineapple stir, and they can be eaten fresh or frozen. In addition, pineapple can be consumed as a supplementary nutritional fruit for good health [1].

Sriracha Pineapple is a Pattavia variety of pineapple grown in Chonburi Province. It is in a group of Smooth Cayenne [2]. At its best, it is one of the most popular types of pineapple for the processing and canning industry, and it is a most delightful fresh fruit to eat. Many canned pineapple industries and consumers prefer to know the juiciness of pineapples before manufacturing products and consuming them, respectively. It is generally known that the more juiciness means more sweetness. Consequently, the canned pineapple industries and consumers want to be able to accurately classify pineapples according to their level of ripeness and juiciness.

As previously mentioned, both sellers and pineapple agriculturists rely on experts to classify the fruit juiciness by tapping on the pineapple. The tapping method is one of the traditional methods. One non-destructive pineapple juiciness measurement is based on force impulse techniques and involves tapping on the pineapple surface with a rubber-tipped stick or with the person's middle finger. However, it can be difficult even for an expert who has years of experience in pineapple quality evaluation to predict the taste quality of pineapple. The flavor of pineapple is almost totally dependent on the sweetness, as measured by the juiciness percentage, rather than on the visual ripeness of the pineapple. Thus, the traditional method for pineapple quality evaluation may produce errors and unevenness as a result of inspector subjectivity.

Several researchers have studied how to classify the maturity of agricultural products such as cacao [3], pineapple [1], pineapple ripeness [4], durian ripeness [5], and described optimal pineapple harvesting [6]. Furthermore, researchers have adapted the structure of pre-trained convolutional neural networks (CNNs), using transfer learning, to pre-train AlexNet and VGGNet networks for apple mealiness detection [7].

This paper proposes a combination of acoustic impulse sensing (a non-destructive evaluation method) and convolutional neural networks (CNNs), which is applied in the classification of Sriracha pineapple juiciness in order to determine the eating quality of pineapples. Sriracha pineapple tapping sounds were recorded. The acoustic sound coming from tapping on the pineapple surface was processed. The pineapple juiciness depends significantly on the soundwave resonance levels. So, in this research, the juiciness was classified into three classes: Juiciness 1, Juiciness 2, and Juiciness 3, respectively. Juiciness 1 was defined as the echo imparted as flatness sound when the pineapple was exceptionally juicy and sweet with a little sour taste. Juiciness 2 was defined as the echo imparted as dullness when the fruit was slightly juicy and sweet with a slight sour taste. Juiciness 3 was defined as the echo imparted as tympany sound when the pineapple was slightly little juicy and sweet and had a rather sour taste. Our method can be divided into three parts: (1) preparation of the sound dataset, (2) extraction of a feature in the sound, and (3) training of a model.

2. Materials and Methods

2.1 Proliferation of deep learning in acoustic sensing

A number of related studies have been concerned with pineapple classification. Azman and Ismail [6] developed a smart, intelligent system indicating the maturity of pineapples for optimal harvest. The maturity levels were divided into three groups, which were "unripe", "partial ripe", and "fully ripe", based on the skin color of the pineapple peel. The dataset was sized to 200x200 pixels and

was allocated according to the three maturity levels of the pineapple. The project proposes a convolution neural network (CNN) for the classification of pineapple images. The experimental results showed that the model's indication of "unripe" and "fully ripe" levels achieved 100% classification accuracy, and the "partially ripe" group achieved 82% classification accuracy. Dittakan *et al*. [1] constructed an automated system for grading pineapples by a non-destructive process. Their process detected pineapple patterns and used pineapple peel texture analysis to separate pineapples into two groups, "Keaw 1" and "Keaw 2". In this research, local binary pattern (LBP) was utilized to detect vital information on pineapple texture images. The model gave the best results with AUC (Area under the ROC Curve) value of 0.979. Chaikaew *et al*. [4] applied a neural network for a pineapple sorting machine using the skin color of the pineapple. The pineapples were divided into three levels: "unripe pineapple", "partial ripe pineapple", and "fully ripe pineapple." The predictiont results showed accuracies of 79% for "unripe pineapple", 82% for "partially ripe pineapple", and 100% for "fully ripe pineapple". Sornsrivichai *et al*. [2] proposed methods using X-ray CT images and CT numbers that showed a significant correlation between the ripeness, translucency of flesh, and taste quality.

Previous studies of pineapple quality classification were mainly focused on ripeness classification and grading of pineapples on the basis of pineapple texture. Our research study is different from previous research in that it applies acoustic sensing to classify Sriracha pineapple edible quality by tapping on pineapple surface and utilizing the resultant echoes. The different sounds depend on the juiciness level of each pineapple.

A number of researchers have applied acoustic sensing and convolution neural networks for detection and classification. A new, non-destructive method for detecting mealiness in Red delicious apple cultivars was proposed [7]. It illustrated the use of acoustic signals and a deep learning technique for mealy and non-mealy detection from the sample apple dataset. It used an impact response technique to record the impact sound between a plastic ball and the apple. The audio sound was recorded and was transformed into a spectrum. The spectrum images were imported into a pre-trained convolutional neural network. The famous pre-trained models were AlexNet and VGGNet which were fine-tuned and utilized as classifiers. The results showed that Alexnet and VGGNet achieved mealy and non-mealy detection results with 91.11% and 86.94% accuracy, respectively.

Kharamat *et al*. [5] proposed a durian ripeness classification by knocking sound and used Mel Frequency Cepstral Coefficients (MFCCs) for feature extraction. The dataset consisted in 900 files divided into three classes: (1) 300 "ripe" files, (2) 300 "mid- ripe" files, and (3) 300 "unripe" files. Each file was recorded in just one knock in 300 milliseconds, and the data was divided into three parts. The total data was separated into 20% for validation dataset, 70% for training dataset, and 10% for the test dataset. The researchers used CNN to help classify the sounds of durian into three groups: "ripe", "mid-ripe", and "unripe". The experimental results showed that the accuracy were 90.78% and 89.74% for validation data and testing data, respectively.

Caladcad *et al*. [8] studied coconut separation using sound for sorting and they developed a tapping system relying on both software and hardware to record coconut sound. The three most widely used machine learning tools were artificial neural network (ANN), support vector machine (SVM), and random forest (RF). The study consisted in 129 samples of coconuts. Each instance was classified into one of three groups according to their maturity level, "pre-mature", "mature", and "over-mature". The experimental results with all three machine learning methods showed at least 80% of accuracy. The RF models outperformed others, with accuracies of 90.98% and 83.48% for training and testing, respectively.

The sound of a knocked coconut was presented for the purpose of predicting coconut maturity using the Naive Bayes method [9]. The process consists in collecting the sound with a MAX9814 sensor device and then processing them with analog-to-digital conversion (ADC) and calculation of the sound signal frequency spectrum using Fast Fourier Transform. The Naive Bayes

method used for classifying coconut maturity level was applied to young, fairly mature and old coconuts. The method achieved a success rate of 80% for a total of 20 test samples.

Bueno *et al*. [3] presented research techniques for determining the ripeness of cacao. A total of 933 cacao samples were classified in the determination of the ripeness of cacao. Each of them was thumped five times in different locations. Each file had a duration of 1 second, producing 4665 cacao sound files, each with a sample rate of 16 kHz and 16-bit audio bit depth.

The Mel-Frequency Cepstral Coefficients spectrogram (MFCCs) was the input feature utilized to train the model using deep learning. Convolutional neural network (CNN) was used to classify cacao into two groups: "ripe" and "unripe". The results indicated accuracies of 97.50% for the training data and 97.13% for the validation data. Simultaneously, the overall accuracy mean was 97.46%. Further, Mel-frequency Cepstrum Coefficients (MFCC) were used to train a multi-layer perceptron (MLP) for the classification of watermelon ripeness into ripe and unripe categories. The method showed an accuracy of 77.25% [10].

2.2 System architecture

The primary system process of Sriracha pineapple juiciness classification is shown in Figure 1. The process consists in: (step 1) data preparation of 30 pineapple samples recorded from a mobile phone, (step 2) input of 1,200 audio waveforms and separations into three classes (Juiciness 1, Juiciness 2, and Juiciness 3), (step 3) audio preprocessing, which is the transformation from audio waveform (time-domain) to spectrum (frequency domain), (step 4) fine training configuration using deep learning methods, (step 5) an evaluation model with accuracy and F1-Score, and (step 6) visualization of the classification report with a confusion matrix.

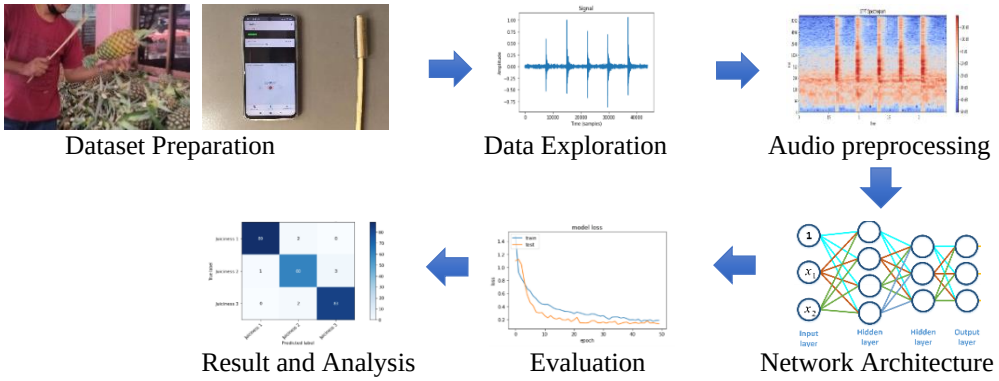


Figure 1. The primary system process of Sriracha pineapple juiciness classification

2.2.1 Dataset preparation

The pineapple sound dataset was prepared with 30 samples of Sriracha pineapples. The tapping sound on the pineapple surface was recorded with a mobile phone in a real environment. The tapping on the pineapple and recording of the sound was an example of an impact of force techniques which is the same as traditional use of a rubber-tipped stick or a person's middle finger to tap on the pineapple. This non-destructive recording method was a mimicking of the actions of sellers and farmers who traditionally tapped pineapple samples to classify the quality of pineapple juiciness. Pineapples were classified individually into three juiciness levels: "Juiciness 1", "Juiciness 2", and "Juiciness 3". Figure 2(a) illustrates a seller using a rubber-tipped stick and Figure 2(b) shows a

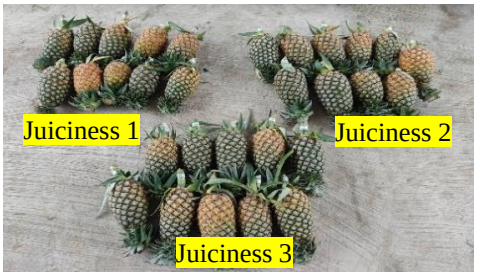
person using his middle finger, both of which are traditional methods to detect the pineapple juiciness. However, the inspector judging the fruit may be inclined to errors and unevenness as his or her decision involves some degree of individual or personal feeling.

A total of 30 pineapple samples were divided into three sets of equal size: Juiciness 1, Juiciness 2, and Juiciness 3 (Figure 3(a)). Each set has 10 pineapples of each juiciness class level and Figure 3(b) shows the juiciness levels of pineapples.

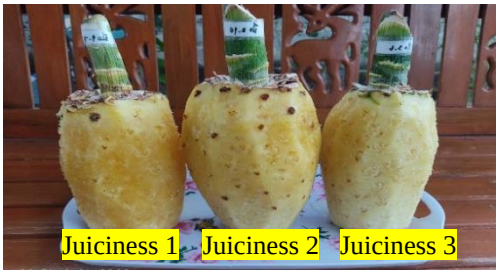
(a) using a rubber-tipped stick

(b) using a person's middle finger

Figure 2. The pineapple juiciness process quality evaluation



(a) Pineapples samples



(b) Juiciness level

Figure 3. The pre-classified harvested pineapple fruits

In the data preparation and collection, the sound of tapping on pineapple was prepared under real environment conditions. So, the sound datasets were collected under time-varying. The acoustic signal was processed using a rubber-tipped stick as shown in Figure 4(a), which was an impact response technique. This determined the sample rate at 44,100 Hertz (Hz) and bit-depth at 16-bits per sample (mono audio). Recording was done by smartphone microphone with Motiv audio software (Figure 4(b)). Each pineapple had 40 audio waveform files, each of which contained data of five taps. Therefore, the tapping sounds were recorded in a total of 1,200 audio waveform files and were labelled into three classes: (1) 400 audio waveform files for Juiciness 1, (2) 400 audio waveform files for Juiciness 2, and (3) 400 audio waveform files for Juiciness 3. The audio files were labelled based on the expertise of the experienced pineapple sellers and farmers. To conduct the experiment, the dataset was split randomly into 80% for the training dataset and 20% for the validation dataset.

(a) The rubber-tipped stick (b) Motiv audio software

Figure 4. The acoustic signal recording tool

2.2.2 Data exploration

Our dataset consisted of 1,200 audio wave files. The audio wave datasets that had a duration of less than 3 seconds numbered 1,098 files (91.5% of the total dataset). The audio wave datasets that had a duration of more than 3 seconds numbered 102 files (8.5% of the the total dataset). As shown in Figure 5, the audio samples had a range oft durations. In order to present an audio spectrogram with fixed input size to the convolutional neural networks (CNN) with a dataset of varying durations, we used zero-padding that filled up space with zeros. This method is one of the most widely used technique, and it does not affect the filters' capability to recognize patterns.

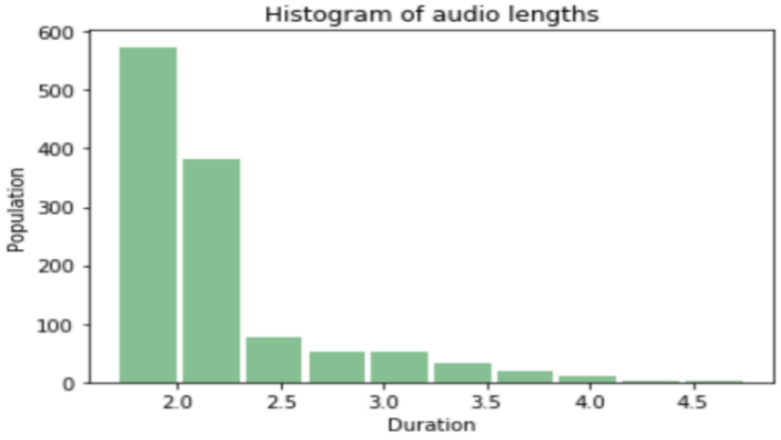
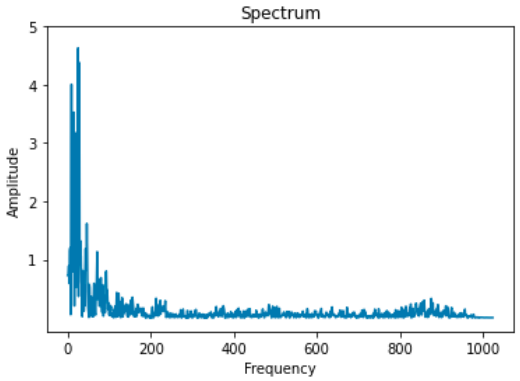
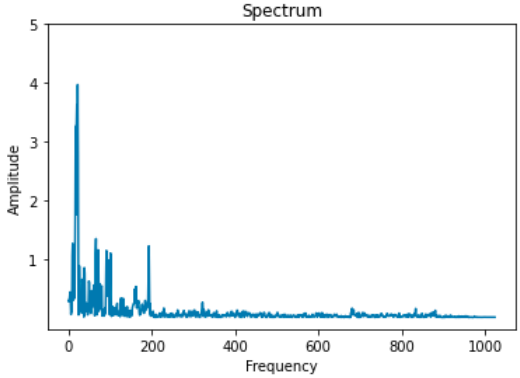


Figure 5. The lengths of the pineapple tapping sounds

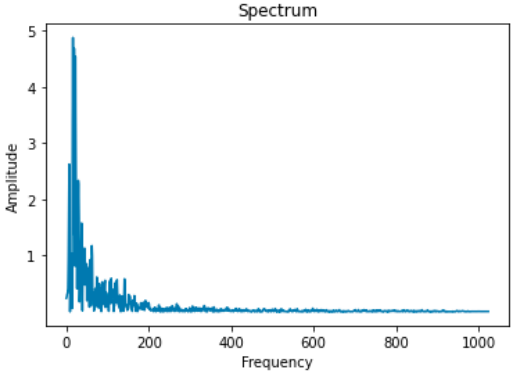
Figure 6 shows the visualization of the sampling frequency spectrum of three pineapple juiciness classes using python programming language, and the librosa python package for audio analysis. Fast Fourier Transform (FFT) was also applied to transform the audio waveform (time-domain) into spectrum (frequency domain). FFT is a widely used and suitable technique for digital audio frequency transformation [8]. The three spectrums, shown in Figure 6(a)-6(c), present the frequency modes of the three classes of pineapple juiciness. The presence of frequency mode varies depending on the juiciness class. For example, Figure 6(a) shows the frequency of juiciness 1, which is exceptionally more juicy and sweeter than the other two classes. Therefore, the magnitudes of the frequency differed (Figure 6(b) and Figure 6(c)), which showed slightly juicy and slightly little juicy, respectively.



(a) frequency of Juiciness 1



(b) frequency of Juiciness 2



(c) frequency of Juiciness 3

Figure 6. The wave sound and magnitude for the frequency of three juiciness acoustic signal samples

2.2.3 Audio preprocessing

The use of Mel-Spectrogram and Mel-Frequency Cepstral Coefficients (MFCC) is a popular method adopted for the sound recognition and visual representation process [11]. Mel-Spectrogram is computed by applying a Fourier transform to analyze the frequency content of a signal and to convert it to the mel-scale, while MFCCs are calculated with a discrete cosine transform (DCT) into a mel-frequency spectrogram. The main difference between the two extraction features is that the mel-spectrogram adopts a linear space-frequency scale while the MFCC use a quasi-logarithmic space-frequency scale [11]. Our experiment also involves a comparison of the results of the use of MFCC and Mel-Spectrogram as audio features. The librosa python library was used for the time-frequency transformation. Technically, we set the parameters: window size and hop length were 2,048 and 512, respectively. Figure 7 illustrates the procedure of feature extraction. The steps are explained as follows. The audio waveform processed into a module computing Short-time Fourier Transform (STFT) to construct the frequency domain and to generate Mel-Spectrogram. The Mel-Scale value was set to 40 as it was the number of audio waveform files. The sample value was normalized between -1 and +1. For the MFCC, we applied the Discrete Cosine Transform (DCT) to generate the Mel-frequencies from the logarithm of Mel-Spectrogram features.

The spectrogram image features for sound classification were represented by converting into vectors. So, MFCC and Mel-Spectrogram data were loaded into a Numpy float32 array and the shape of MFCC (1200, 40, 205) was 1,200 samples with 40 MFCC coefficients and 205 frames. Likewise, the shape of Mel-spectrogram was 1,200 samples with coefficients of 205 frames and 40 Mel brands. Both MFCC and Mel-Spectrogram had a scale between -1 to +1 as a result of the normalization made by the librosa python library. Next, we split data by applying a randomizing index between training and test data sets, so they contained 80 and 20 percent of the whole data, respectively. Lastly, we made a reshaping to fit the network input dimension. This consisted in a row, column, and one channel that were ready to feed it into the Network.

2.2.4 Audio CNN architecture

A convolutional neural network (CNN) is used after the input sound signal has been converted into an image [3]. In this research, the feature vector of the spectrogram image is assigned to different input nodes. Our neural network architecture comprises three convolutional 2-dimensional layers interleaving with two max-pooling operations, one flatten layer, and two fully connected layers. The activation function parameter is a rectified linear unit (ReLU). For convolutional 2-dimensional layers, two layers are situated before the max pooling layer, which is a common technique in CNN to reduce the dimensions of input data.

The algorithms in this research are implemented using Keras and TensorFlow to train and evaluate the model. The model includes three convolutional 2-dimensional layers. At the end of the last convolutional 2-dimensional layers, the output data is fed into a flatten layer and then into a fully connected layers. The final layer containing the softmax output provides classification probabilities for the input data. The output results of the model indicate the three classes of juiciness level of pineapples. The CNN architecture for audio classification is presented in Figure 8.

Adam's optimization consists in a fine-tuning of the model and is compatible with a 0.00006 learning rate. Moreover, during model training, the dropout method is a regularization technique that randomly cuts the neural networks for reducing overfitting in the convolutional neural network (CNN). We determined the dropout values in convolution 2-dimensional layers 2 and 3, and in the fully connected layer were 25% and 50%, respectively. Batch normalization methods are used in convolution 2-dimensional layers for normalizing the output to increase the stability of the model and reduce overfitting of the neural network. The detailed network architecture is proposed in Table 1.

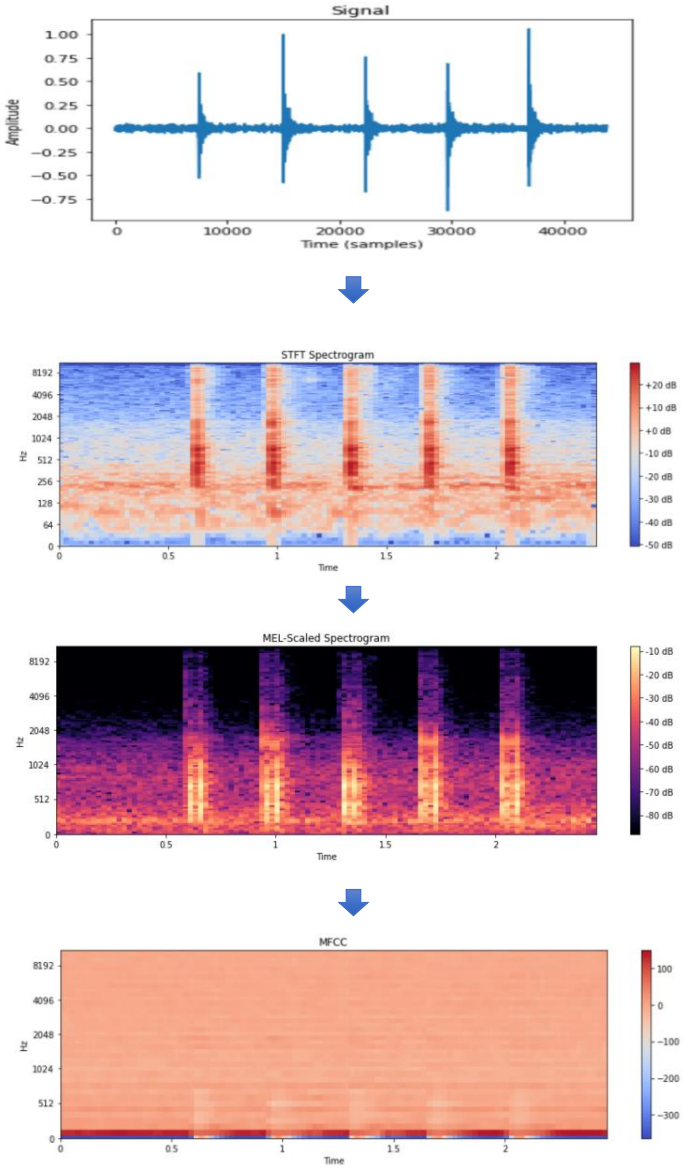


Figure 7. Procedure of feature extraction for MFCC and Mel-Spectrogram

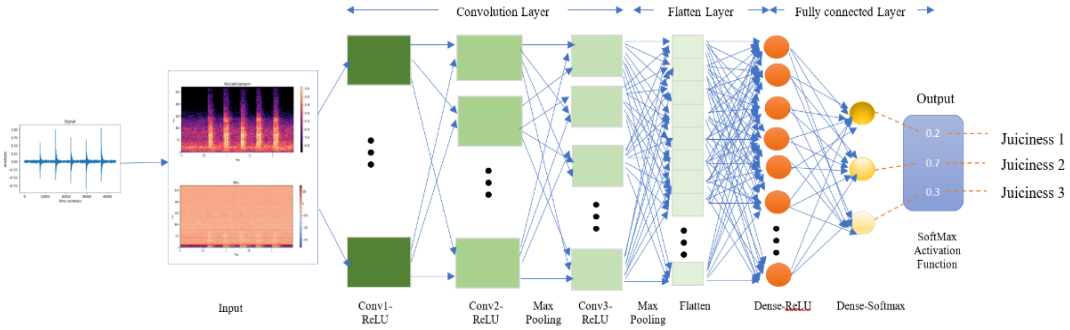


Figure 8. CNN architecture for audio classification

Table 1. The detailed network architecture of the model

Layer	Network architecture
Input layer	Input shape = 40x205x1
Conv2D #1	Filters = 32; Kernel = (3,3); Padding = 'same'; Activation function = ReLU; Batch Normalization.
Conv2D #2	Filters = 64; Kernel = (3,3); Padding= 'same'; Activation function = ReLU; Batch Normalization; Max pooling = (2,2); Dropout = 0.25.
Conv2D #3	Filters = 128; Kernel = (3,3); Padding= 'same'; Activation function = ReLU; Batch Normalization; Max pooling = (2,2); Dropout = 0.25.
Flatten layer	Flatten.
Dense layer	Dense = 128; Activation function = ReLU; Dropout = 0.50.
Output layer	Dense = 3; Activation function = Softmax.

2.2.5 Fine-tune tranning configuration

To find out the fine-tuning parameters for training the network, a determined number of batch sizes, 8, 16, and 32, and different epochs, 30, 40, and 50, were utilized to train and validate the CNN. We set hyperparameters for the model. Expressly, we set parameters for the spectrum image's visualization. The value of windows size was equal to 2,048 and the hop size was equal to 512. The image size of the convolutional neural network model was 40 x 205.

In this case, the neural training network using CNN combination with MFCC was compared with Mel-Spectrogram. We used Adam's optimization with the initial learning rate of $6x10^{-5}$. All the dataset was split randomly into 80% for the training dataset and 20% for the validation dataset. The model training was repeated five times. Therefore, the dataset randomly changed every time and the average validation accuracy (VA) as calculated in equation (1), error (loss), and the standard deviation (SD), were computed and recorded.

Table 2 shows the results of the comparison of the fine-tuning hyperparameters used for training the CNN+MFCC and CNN+Mel-Spectrogram. For training the CNN+MFCC, after fine training configuration, and repetition of model training five times, we recorded the validation

Table 2. Results of fine-tuning hyperparameters for the CNN combined with MFCC and Mel-Spectrogram (the validation values±SD)

Epochs	30		40		50	
Batch-sizes	**VA (%)±SD**	**Loss±SD**	**VA (%)±SD**	**Loss±SD**	**VA (%)±SD**	**Loss±SD**
CNN+MFCC						
8	**96.58 ±0.54**	**0.11 ±0.02**	95.50 ±1.23	0.16 ±0.05	96.17 ±1.08	0.15 ±0.04
16	93.42 ±1.99	0.20 ±0.05	94.58 ±1.41	0.16 ±0.04	94.92 ±1.54	0.15 ±0.03
32	93.25 ±2.15	0.21 ±0.07	94.75 ±0.48	0.16 ±0.03	95.83 ±1.06	0.14 ±0.01
CNN+Mel-Spectrogram						
8	94.83 ±0.96	0.15 ±0.04	94.92 ±1.19	0.15 ±0.03	**96.33±0.68**	**0.13±0.02**
16	93.42 ±2.51	0.19 ±0.04	92.42 ±1.26	0.21 ±0.04	93.50 ±3.30	0.20 ±0.07
32	91.33 ±1.70	0.25 ±0.04	89.42 ±2.74	0.28 ±0.08	91.75 ±2.75	0.23 ±0.06

VA= Validation accuracy, Loss = the error of the model, SD = standard deviation

accuracy (VA), the error of the model (loss), and a computed average of VA and loss, and standard deviation (SD). For CNN+MFCC, the results showed that setting a batch size and epochs values of 8 and 30, respectively, produced a validation accuracy equal to 96.58%.

For the CNN+Mel-Spectrogram, we set parameters with 8 batch sizes and 50 epochs and obtained a validation accuracy equal to 96.33%. The CNN+Mel-Spectrogram gave the lowest error and standard deviation of models each time.

2.3 Evaluation

The optimal network with optimal hyperparameters was selected and run on the dataset, which had been split randomly into 80% for the training dataset and 20% for the testing dataset. Validation accuracy and error (loss) were reported. The model performance evaluation uses various criteria including accuracy, precision, recall, and F1-Score to compare the results of both models on the test dataset via the confusion matrix. The accuracy is the number of correctly classified pineapple juiciness samples over the total number of pineapple juiciness samples (expressed as a percentage) as shown in equation (1).

$$\text{Accuracy} = \frac{number\ of\ correctly\ classified\ samples}{total\ number\ of\ samples} * 100 \quad (1)$$

The second measure, F1-Score, as shown in equation (2), is the weighted average of Precision and Recall. Therefore, the score taken for precision is (TP/(TP+FP)) and recall (TP/(TP+FN)) where TP (True Positive), FP (False Positive), and FN (False Negative) refer to terms used in the confusion matrix.

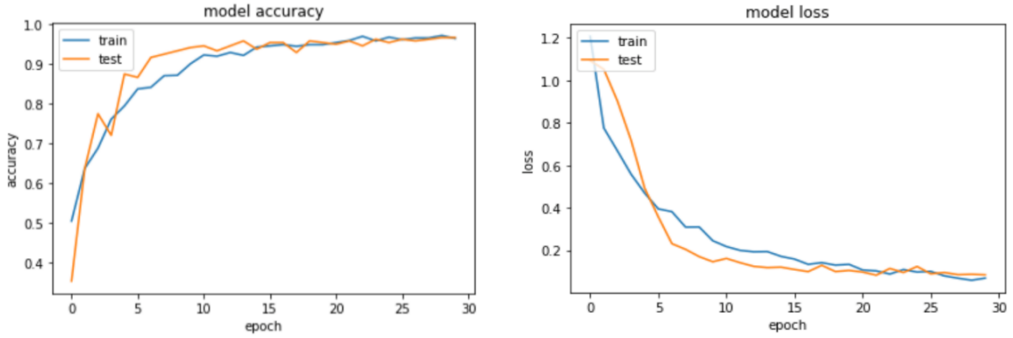
$$F1 = 2 \times \frac{Precision \; x \; Recall}{Precision+Recall} \quad (2)$$

Furthermore, a confusion matrix is a specific table layout that illustrates a classification model performance on a set of test data.

3. Results and Discussion

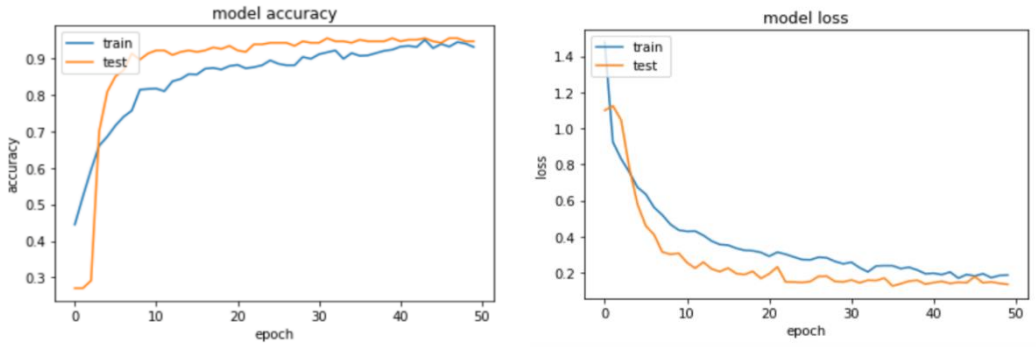
The procedure was implemented on Google Colab and with Python language. We applied LibROSA to extract audio features, and Keras and TensorFlow library to develop a classification model. The experiment used 'Sriracha' pineapples. A set of 30 pineapple samples was used to record tapping sound data and construct a 1,200 audio waveform files dataset. The dataset consisted in 400 juiciness 1 pineapple audio waveform files, 400 juiciness 2 pineapple audio waveform files, and 400 juiciness 3 pineapple audio waveform files.

To measure and present the performance of the proposed model, the results can be viewed in the graph presented to see the difference in model accuracy between the training dataset and the validation dataset. Figures 9 and 10 show the results with the proposed audio CNN combined with MFCC and Mel-Spectrogram feature extraction and setting softmax for the output function. The result shows that CNN combined with MFCC outperformed the CNN combined with Mel-Spectrogram. The results show that MFCC performed better for classification juiciness level. Table 3 illustrates the performance by adopting a model for classification of pineapple juiciness. The VA of test datasets for CNN+MFCC and CNN+Mel-Spectrogram reached 96.67% and 94.58%, respectively. Consequently, CNN+MFCC outperformed CNN+Mel-Spectrogram for both accuracy and loss.



(a) validation accuracy (b) loss

Figure 9. The results of CNN+MFCC



(a) validation accuracy (b) loss

Figure 10. The results of CNN+Mel-Spectrogram

Table 3. Performance of the optimal selected fine-tuning for the model CNN

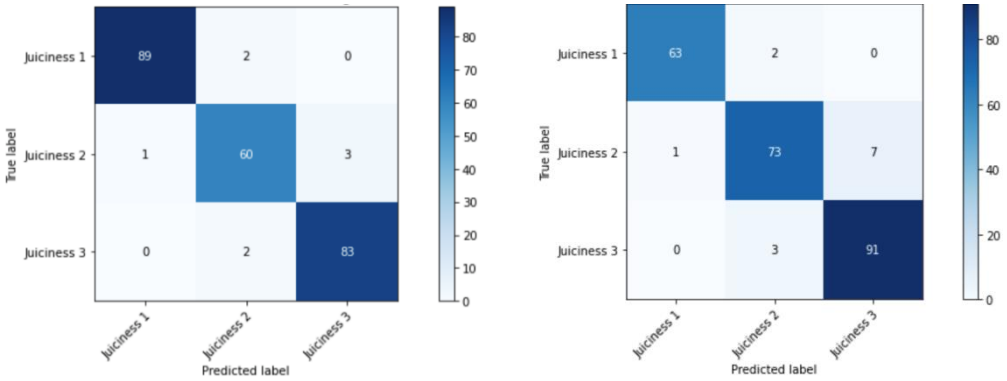
Model	Validation Accuracy (%)		Loss (The error of the model)	
	Training	Test	Training	Test
CNN+MFCC	99.7917	**96.6667**	0.0144	0.0844
CNN+Mel-Spectrogram	99.2708	**94.5833**	0.0300	0.1347

We present the F1-Score of each class in Table 4. It provides the comparison results for the two feature extraction methods. Table 4 shows the value of F1-Score of the three classes. The CNN+MFCC outperformed the CNN+Mel-Spectrogram in Juiciness 1, Juiciness 2 and Juiciness 3 and obtained 0.98, 094 and 0.97, respectively. A macro average is 0.96, and weighted average is 0.97 for the evaluation system.

Table 4. The comparison results classification of the two feature extraction methods

Class	CNN+MFCC				CNN+Mel-Spectrogram			
	Precision	**Recall**	**F1-Score**	**No. of Samples**	**Precision**	**Recall**	**F1-Score**	**No. of Samples**
Juiciness1	0.99	0.98	0.98	91	0.98	0.97	0.98	65
Juiciness2	0.94	0.94	0.94	64	0.94	0.90	0.92	81
Juiciness3	0.97	0.98	0.97	85	0.93	0.97	0.95	94
Macro avg	0.96	0.96	**0.96**	240	0.95	0.95	**0.95**	240
Weighted average	0.97	0.97	**0.97**	240	0.95	0.95	**0.95**	240

The confusion matrices of Figure 11(a)-(b) visualize the two networks algorithms performance. Firstly, the CNN combined with the MFCC algorithm correctly classified 89 out of 91 (97.80%,) 60 out of 64 (93.75%), and 83 out of 85 (97.65%) of pineapples samples of juiciness 1, juiciness 2, and juiciness 3, respectively. The remaining 2, 4, and 2 pineapple samples were classified incorrectly, respectively. Lastly, the CNN combined with Mel-Spectrogram algorithm correctly classified 63 out of 65 (96.92%), 73 out of 81 (90.12%), and 91 out of 94 (96.81%) of pineapples samples of juiciness 1, juiciness 2, and juiciness 3, respectively. The remaining 2, 8, and 3 pineapple samples were classified incorrectly, respectively.



(a) CNN+MFCC (b) CNN+Mel-Spectrogram

Figure 11. Confusion matrices of test data

4. Conclusions

This research study provides a novel method for non-destructive quality evaluation method for the juiciness classification of Sriracha pineapples from audio waveform sounds. The model applies a combination of acoustic sensing and deep learning to compare the results between MFCC and Mel-Spectrogram feature extraction connected to the same convolutional neural network (CNN) model for the classification into three classes: Juiciness 1, Juiciness 2, and Juiciness 3. MFCC combined with CNN performed best, outperforming the Mel-spectrogram combined with CNN. The accuracy of our model was 96.67%. This research result suggests that the proposed method can be applied at a fresh pineapple market to help buyers decide which pineapples are of the best edible quality for them. The model can be further improved to enhance performance by using other extraction methods including the linear predictive cepstral coefficient (LPCC) and Perceptual Linear Prediction (PLP). Other deep learning models such as recurrent neural networks (RNNs), which allow for larger data samples and other collection methods, could also be applied in the future work.

5. Acknowledgements

This research was supported by the “Celebrations on the Auspicious Occasion of His Majesty the King’s 70 the Birthday Anniversary” Ph.D. Degree Scholarship, Project for Doctoral Studies in Thailand.

References

[1] Dittakan, K., Theera-Ampornpunt, N. and Boodliam, P., 2018. Non-destructive grading of Pattavia pineapple using texture analysis. *The 21st International Symposium on Wireless Personal Multimedia Communications.* Chiang Rai, Thailand, November 25-28, 2018, pp. 144-149.

[2] Sornsrivichai, J., Yantarasri, T. and Kalayanamitra, K., 2000. Nondestructive techniques for quality evaluation of pineapple fruits. *Acta Horticulturae (International Pineapple Symposium)*, 529, 337-341.

[3] Bueno, G.E., Valenzuela, K.A. and Arboleda, E.R., 2020. Maturity classification of cacao through spectrogram and convolutional neural network. *Jurnal Teknologi dan Sistem Komputer*, 8(3), 228-233.

[4] Chaikaew, A., Thanavanich, T., Duangtang, P., Sriwanna, K. and Jaikhang, W., 2019. Convolutional neural network for pineapple ripeness classification machine. *Proceeding of the 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Pattaya, Thailand, July 10-13, 2019, pp. 373-376.

[5] Kharamat, W., Wongsaisuwan, M. and Wattanamongkhol, N. 2020., Durian ripeness classification from the knocking sounds using convolutional neural network. *The 8th International Electrical Engineering Congress*. Chiang Mai, Thailand, March 4-6, 2020, pp. 1-4.

[6] Azman, A.A. and Ismail, F.S., 2017. Convolutional neural network for optimal pineapple harvesting. *Journal of Electrical Engineering,* 16(2), https://doi.org/10.11113/elektrika. v16n2.54

[7] Lashgari, M., Imanmehr, A., and Tavakoli, H., 2020. Fusion of acoustic sensing and deep learning techniques for apple mealiness detection. *Journal of Food Science and Technology*, 57(6), 2233-2240.

[8] Caladcad, J.A.A., Cabahug, S., Catamco, M.R., Villaceran, P.E., Cosgafa, L., Cabizares, K.N., Hermosilla, M. and Piedad, E., 2020. Determining Philippine coconut maturity level using machine learning algorithms based on acoustic signal. *Computer and Electronics in Agriculture*, 172, 105327, https://doi.org/10.1016/j.compag.2020.105327.

[9] Rahmawati, D., Haryanto, H. and Sakariya, F., 2019. The design of coconut maturity prediction device with acoustic frequency detection using Naïve Bayes method based microcontroller. *Journal of Electrical Engineering, Mechatronic and Computer Science*, 2(1), https://doi.org/10.26905/jeemecs.v2i1.2806.

[10] Baki, S.R.M.S., Z., M.A.M., Yassin, I.M., Hasliza, A.H. and Zabidi, A., 2010. Non-destructive classification of watermelon ripeness using Mel-Frequency Cepstrum Coefficients and Multilayer Perceptrons. *The 2010 International Joint Conference on Neural Networks*. Barcelona, Spain, July 18-23, 2010, pp. 1-6.

[11] Bui, K.N., Oh, H. and Yi, H., 2020. Traffic density classification using sound datasets: An empirical study on traffic flow at asymmetric roads. *IEEE Access*, 8, 125671-125679.