Predicting Maturity of Coconut Fruit from Acoustic Signal with Applications of Deep Learning †

Farook Sattar

Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8P 5C2, Canada; fsattar@ieee.org
† Presented at the 2nd International Online Conference on Agriculture, 1–15 November 2023; Available online: https://iocag2023.sciforum.net/.

Abstract: This paper aims to develop an effective AI-driven method to predict the maturity level of coconut (*Cocos nucifera*) fruits using acoustic signals. The proposed sound-based autonomous approach exploits various deep learning models, including customized CNN pretrained networks, i.e., the ResNet50, InceptionV3, and MobileNetV2, models for maturity level classification of the coconuts. The proposed study also demonstrates the effectiveness of various deep learning models to automatically predict the maturity of coconuts into three classes, i.e., premature, mature, and overmature coconuts, for inspecting the coconut fruits by using a small amount of input acoustic data. We use an open-access dataset containing a total of 122 raw acoustic signals, which is the result of knocking 122 coconut samples. The results achieved by the proposed method for coconut maturity prediction are found to be promising, which enables producers to accurately determine the yield and product quality.

Keywords: coconut fruit; maturity levels; fruit quality; deep learning; artificial intelligence; acoustic signals; prediction

1. Introduction

Currently, a lot of attention is paid to evaluating and classifying horticulture crops, particularly fruits. Maturity prediction takes a major step in deciding the value of the coconut (*Cocos nucifera*) as it has a direct link to the quality of the product. Sound-based deep learning to predict the maturity of a coconut can play an important role in a number of tropical countries like the Philippines, Indonesia, and India, which are producing and exporting coconuts due to their high demand worldwide. Coconut has a delicious taste, many nutrients, and multiple uses.

It is essential to correctly measure the coconut maturity level since numerous benefits can be found in coconuts with different maturity levels. The type and food usage of the coconuts depend on the maturity level. Premature coconuts contain tender meat and water, whereas in mature coconuts, the meat is thick and there is less water. On the other hand, overmature coconuts have the hardest meat with the least coconut water or none at all. The maturity level then significantly affects the water composition and physicochemical characteristics of the coconuts. Moreover, premature coconuts can be used for refreshing drinks, while mature coconuts can be used for salads or desserts. Therefore, maturity prediction plays an important role in measuring the coconut value as well as the product quality.

Limited efforts have been carried out to develop an acoustical system for the maturity level prediction of coconut fruits. Developing such a system can largely help companies in mass coconut export to the global market, most remarkably with respect to time and money. In [1], a sound-based classification scheme is proposed for the prediction of maturity levels of coconuts into premature, mature, and overmature stages. A prototype tapping system

for coconuts is developed to record acoustic signals followed by spectral (FFT) feature extraction and classification using Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM) classifiers. Promising results are obtained with respect to classification accuracy and F-score, which had the best performances achieved by the RF classifier. Reference [2] addresses finding a robust deep learning scheme to automatically recognize the coconut maturity in different environmental conditions using real images taken from coconut farms and Google. Different CNN models have been used to recognize coconuts as tender coconut and mature coconut. The ResNet50 architecture is more effective in detecting maturity levels than the other used CNN models. In [3], a new Fuzzy Neural Network and Sperm Whale Optimization (FNN–SWO)-based adaptive model is proposed for predicting the maturity levels of coconut water as tender, mature, and very mature by using clustering and fuzzification of the input data. In the proposed FNN–SWO model, fuzzy rules are utilized for training and testing with an adaptive network and the SWO algorithm is applied to select the optimum weights of the fuzzy rules. The advantages of the present method over other methods are shown by the results of differentiating the maturity of the coconut water. Reference [4] presents a coconut palm disease and coconut maturity prediction method based on image processing and deep learning techniques using the images captured by a drone flying over a coconut farm. Both the results of coconut palm disease and coconut maturity prediction are improved by the CNN models with Densenet and ResNet architectures, respectively, comparing the results obtained by the SVM model as a classifier. In [5], an improved method is developed to detect maturity levels of coconuts in complex backgrounds using a faster region-based convolutional neural network (Faster R–CNN) model. The coconut and mature coconut images are collected from coconut farms and then augmented through rotation and color transformation to train the model. The performance of the Faster R-CNN model is found to be better than that of the methods based on other object detectors such as SSD (single shot detector), YOLOV3 (you only look once), and R-FCN (region-based fully convolutional networks) for the real-time coconut images.

Reference [6] proposes a classification system to predict the maturity level of young coconuts, exploiting sound signatures or acoustic responses from sound vibrations collected by a vibration sensor. The model is trained and tested with 150 coconut samples. Eighteen sound signatures are generated for each coconut sample when the nine most significant signatures out of the eighteen are selected through PCA (Principal Component Analysis) for the training stage. Different classifiers such as SVM, KNN (K-Nearest Neighbor), ANN and DT (Decision Tree) are applied to the prediction models, where the ANN-based system gives the best performance. Following the results in [6], an ANN-based coconut maturity detector is designed and developed utilizing sound signatures, as presented in [7]. The prototype consists of a vibration sensor, vibration motor, RPi 3b+ microcontroller, LCD, and battery while achieving good classification results. In [8], the work deploys fuzzy logic to develop a coconut maturity classification scheme considering color and sound for fuzzy inputs. The image color analysis is adopted to find the brown color percentage in the coconut shell, and sound spectral analysis is used to relate the shell hardness and meat condition. Finally, fuzzy inference is performed to evaluate the relationship of sound and color with the coconut maturity. In [9], several state-of-the-art DNN (deep neural network) models, including Xception, ResNet50V2, ResNet152V2, and MobileNetV2, are applied to detect the maturity levels of the coconuts. Among them, the performance of MobileNetV2 is found to be the most effective. Moreover, the model is implemented on an Android device so that it can be easy for the farmers to identify the coconut maturity during harvesting and in other applications.

Our aim is to develop an effective AI-driven system to predict the maturity level of the coconuts using acoustic signals. The aim of this paper is to show the benefits of various deep learning models to inspect the quality of the coconuts by automatically predicting their maturity level into three classes, i.e., premature, mature, and overmature, from a small amount of input acoustic data. The proposed sound-based autonomous

approach would eventually exploit various deep learning models, including customized CNN and pretrained CNNs, such as ResNet50, InceptionV3, and MobileNetV2, in coconut maturity prediction.

2. Materials and Methods

2.1. Data

We used an open-access dataset [10] consisting of a total of 122 raw acoustic signals, which were generated by knocking 122 coconut samples on their ridges. The type and number of data samples used are as follows (see Table 1):

Table 1. The sound (wav) files used.

Types of Coconuts	Number of Samples
Premature coconut	8
Mature coconut	36
Overmature coconut	78

The setup for the system used to collect the acoustic data is shown in Figure 1. In the signal acquisition process, acoustic foams are attached to the interior walls of the coconut-knocking chamber to reduce undesirable disturbances from the surrounding environment. Also, during the knocking process, the coconut fruits are placed in a fixed and stable position using a rigid hollow holder (see Figure 1). With a knocking ball, the coconuts are uniformly knocked on their side ridges rather than their top and bottom ridges, where the coconut samples tend to have uneven thickness. Unlike any visual-based system, this audio-based system does not need to be adapted to different coconut varieties (having different skin, size, and density).
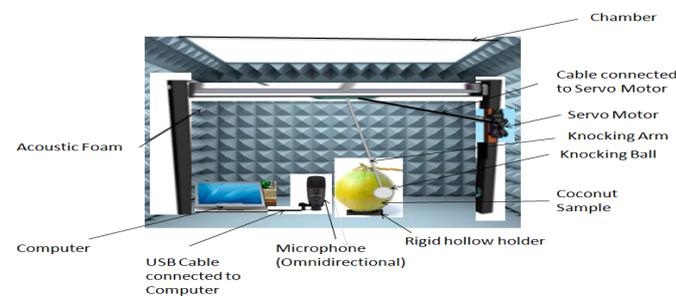

Figure 1. The system with coconut fruit and signal acquisition [1].

2.2. Proposed Scheme

We propose a hybrid scheme consisting of wavelet scattering transform (WST) followed by deep neural networks (DNN), as depicted in Figure 2. Different deep learning models, including customized CNN (Convolutional Neural Networks), pretrained CNN including ResNet50, InceptionV3, and MobileNetV2, are used for DNN.
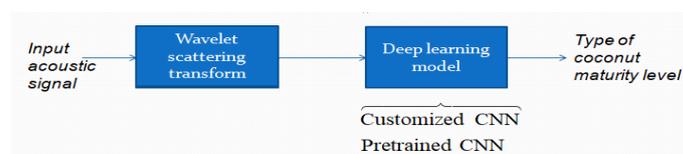

Figure 2. The overall flow diagram of the proposed scheme.

2.2.1. Wavelet Scattering Transform (WST)

The wavelet scattering transform (WST) provides a robust time–frequency (t-f) representation of the nonstationary and nonlinear signals [11]. The combination of wavelet

transform and scatterplot provides more comprehensive spectro-temporal characteristics of the signal. The aim here is to extract relevant features for predicting the maturity levels of the coconuts through WST. The WST involves the processing of an input signal by a number of connected layers, and each layer is identified by three successive operations, namely convolution, modulus as nonlinearity, and averaging as pooling. The advantages of WST are its invariance properties to signal translation and dilation. These properties play a vital role in the effective classification of the signals whose characteristics vary both with time and frequency. Moreover, there is no need for training of the networks with WST, unlike CNN. Another benefit of the WST is its exceptionally good performance even with limited data.

2.2.2. Customized CNN

WST is used to calculate the scattergram image ($n \times m$) as input to the customized CNN. The architecture of the customized CNN has three convolution blocks and one fully connected (FC) layer [12,13], where each convolution block contains one 1D convolution layer of size (3×1) and a batch normalization layer. A max-pooling (1×2) layer and stride(1,2) follow each convolutional layer. In the network, the three convolution blocks consist of 8, 16 and 32 filters, and an activation function of rectified linear unit (ReLU) is used in all layers. Finally, a fully connected layer with *C* hidden neurons is connected to a softmax layer to identify *C* number of classes.

2.2.3. Pretrained CNNs

- **ResNet50**
  ResNet50 is a convolutional neural network (CNN) based on a deep residual learning framework while training very deep networks with hundreds of layers [14]. This architecture introduces the idea of the Residual Network (ResNet) to address the problems of vanishing and exploding gradients. ResNet50 consists of a total of 50 layers (48 convolutional layers, 1 max pooling layer, and 1 average pooling layer), and the network structure is divided into 4 blocks when each block has a set of residual blocks. The residual blocks are constructed to preserve information from earlier layers and this enables the network to learn better representations for the input data. The main feature of ResNet is the skip connection to propagate information over layers to enable building deeper networks.

- **InceptionV3**
  InceptionV3 [15] is a CNN-based network introducing "inception module", which is composed of a concatenated layer with 1×1, 3×3, and 5×5 convolutions. The InceptionV3 model has a total of 42 layers and consists of multiple layers of convolutional and pooling operations, followed by fully connected layers. One of the key features of InceptionV3 is its ability to scale to large datasets and to handle images of varying resolutions and sizes. The method has reduced the number of parameters accelerating the training rate. The other name of this network is GoogLeNet model. The advantages of InceptionV3 are as follows: factorization into smaller convolutions, grid size reduction, and use of auxiliary classifiers to tackle the vanishing gradient problem during the training of a very deep network [16].

- **MobileNetV2**
  MobileNetV2 [17] is a convolutional neural network consisting of 53 layers and uses depth-wise separable convolutions to reduce the model size and complexity. The computationally expensive convolutional layers are then replaced here by depthwise separable convolution, which is constructed by a 3×3 depthwise convolutional layer followed by a 1×1 convolutional layer. MobileNetV2 is a modified version of the MobileNetV1 network achieved by adding inverted residuals and linear bottleneck layers [18]. Both ReLU6 activation function and linear activation function are used in the MobileNetV2 [19]. The inclusion of a linear activation function is made possible in the MobileNetV2 network to reduce the information loss by considering an inverse

residual structure with a low-dimensional feature at the final output. Moreover, in contrast to the ResNet network, this network provides shortcut connections only when stride = 1 [20].

3. Results and Discussion

The results are presented in terms of accuracy and F1 score. The accuracy is calculated using the expression below:

$$\text{Accuracy} = \frac{\text{Number of correct predictions (TP + TN)}}{\text{Total number of predictions (TP + TN + FP + FN)}} \tag{1}$$

where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively. The F1 score is obtained as an average of F1 score$_i$, $i = 1, \cdots, n$, over the total number of classes $n = 3$ when F1 score$_i$ for the i-th class is obtained as

$$\text{F1 score}_i = \frac{2\text{TP}_i}{2\text{TP}_i + \text{FP}_i + \text{FN}_i} \tag{2}$$

where TP$_i$, TN$_i$, FP$_i$, FN$_i$ refer to true positive, true negative, false positive, and false negative for the i-th class.

Here, we used MATLAB (https://www.mathworks.com/products/matlab.html, accessd on 1 March 2024) software to develop different CNNs by considering several common models, namely customized CNN, ResNet50, InceptionV3, and MobileNetV2 for training, thereby producing the corresponding prediction results. The hyperparameters, which we set for training, are as follows: the batch size = 10, the learning rate = 0.0001, and the number of epoch = 6. The optimization algorithm of stochastic gradient descent (SGD) is applied here for training. It can be noted that we use the same hyperparameters and dataset to train each model five times to obtain fair simulation results. The results after training the customized CNN, ResNet50, InceptionV3, and MobileNetV2 models five times are finally averaged to obtain accuracies of 61.30%, 84.25%, 77.32%, and 73.12%, respectively, as depicted in Table 2. The results of the F1 score are also presented in Table 2.

Table 2. The results for the prediction of coconut maturity level in terms of classification accuracy (%) and F1 score.

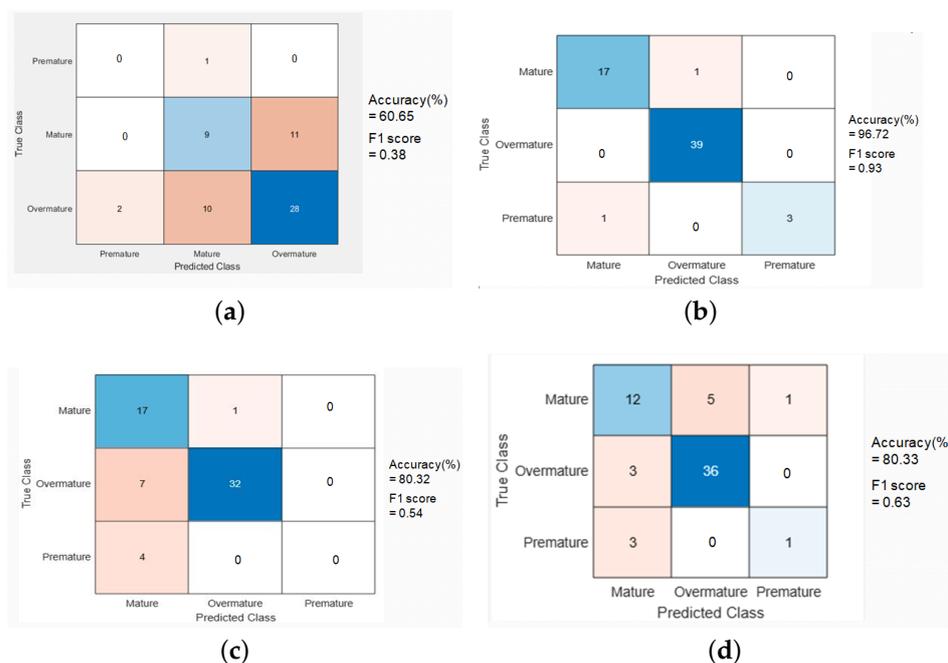| Method	Accuracy (%) (Mean ± Std)	F1 Score (Mean ± Std)
WST + Customized CNN	61.30 ± 2.48	0.29 ± 0.03
WST + Pretrained CNN	ResNet50: 84.25 ± 8.59 InceptionV3: 77.32 ± 10.28 MobileNetV2: 73.12 ± 6.30	0.74 ± 0.19 0.48 ± 0.12 0.53 ± 0.10

The confusion matrices are illustrated in Figure 3a–d.

Figure 3. The confusion matrices for (**a**) WST+customized CNN, (**b**) WST+ResNet50 (pretrained CNN), (**c**) WST+InceptionV3 (pretrained CNN), (**d**) WST+MobileNetV2 (pretrained CNN).

4. Conclusions

It is challenging to predict the maturity levels of the coconuts for harvesting without human intervention in a complex environment considering the similarities of the fruits and their backgrounds. In this paper, an acoustical hybrid method is proposed based on wavelet scattering transform (WST) and a deep learning model. The results of prediction for the coconut maturity stages by the present method are shown to be promising even with limited real acoustic data. This paper also shows the usefulness of deep learning in coconut fruit maturity prediction by evaluating different deep learning models, including customized CNN and pretrained CNN for acoustic signals. Compared with the results of customized CNN, the pretrained CNN models demonstrate higher accuracies and have a greater benefit for the acoustical data tested in this study. The method based on WST and ResNet50 provides the best performance among all the used DNN models in terms of classification accuracy ($\approx$85%) and F1 score ($\approx$0.75) with a small acoustic dataset. The proposed scheme can be tested to predict the quality of other types of fruits. The performance can be improved by further analyzing the deep learning models used in our hybrid method. Further, different datasets can be utilized for a more detailed evaluation of the proposed scheme.

References

1. Caladcada, J.A.; Cabahug, S.; Catamcoa, M.R.; Villaceran, P.E.; Cosgafa, L.; Cabizares, k.N.; Hermosilla, M.; Piedad, E.J. Determining Philippine coconut maturity level using machine learning algorithms based on acoustic signal. *Comput. Electron. Agric.* **2020**, *172*, 105327. [CrossRef]
2. Subramanian, P.; Sankar, T.S. Coconut Maturity Recognition Using Convolutional Neural Network. In *Computer Vision and Machine Learning in Agriculture*; Uddin, M.S., Bansal, J.C., Eds.; Algorithms for Intelligent Systems; Springer: Berlin/Heidelberg, Germany, 2002; Volume 2. [CrossRef]
3. El-Shafeiy, E.; Abohany, A.A.; Elmessery, W.M.; El-Mageed, A.A.A. Estimation of coconut maturity based on fuzzy neural network and sperm whale optimization. *Neural Comput. Appl.* **2023**, *35*, 19541–19564. [CrossRef]
4. Selshia, A. Coconut Palm Disease And Coconut Maturity Prediction Using Image Processing And Deep Learning. *Int. J. Creat. Res. Thoughts (IJCRT)* **2023**, *11*, 260–265.
5. Parvathi, S.; Selvi, S.T. Detection of maturity stages of coconuts in complex background using Faster R-CNN model. *Biosyst. Eng.* **2021**, *202*, 119–132. [CrossRef]
6. Fadchar, N.A.; Cruz, J.C.D. A Non-Destructive Approach of Young Coconut Maturity Detection using Acoustic Vibration and Neural Network. In Proceedings of the 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, 28–29 February 2020.
7. Fadchar, N.A.; Cruz, J.C.D. Design and Development of a Neural Network—Based Coconut Maturity Detector Using Sound Signatures. In Proceedings of the 7th IEEE International Conference on Industrial Engineering and Applications (ICIEA), Bangkok, Thailand, 16–21 April 2020.
8. Javel, I.M.; Bandala, A.A.; Salvador, R.C., Jr.; Bedruz, R.A.R.; Dadios, E.P.; Vicerra, R.R.P. Coconut Fruit Maturity Classification Using Fuzzy Logic. In Proceedings of the 10th IEEE International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 29 November–2 December 2018.
9. Varur, S.; Mainale, S.; Korishetty, S.; Shanbhag, A.; Kulkarni, U.; Meena, S.M. Classification of Maturity Stages of Coconuts using Deep Learning on Embedded Platforms. In Proceedings of the 3rd IEEE International Conference on Smart Data Intelligence (ICSMDI), Trichy, India, 30–31 March 2023.
10. Caladcada, J.A.; Piedadb, E.J. Acoustic dataset of coconut (Cocos nucifera) based on tapping system. *Data Brief* **2023**, *47*, 1–7. [CrossRef]
11. Parmar, S.; Paunwala, C. A novel and efficient Wavelet Scattering Transform approach for primitive-stage dyslexia-detection using electroencephalogram signals. *Healthc. Anal.* **2023**, *3*, 100194. [CrossRef]
12. Tanveer, M.H.; Zhu, H.; Ahmed, W.; Thomas, A.; Imran, B.M.; Salman, M. Mel-Spectrogram and Deep CNN Based Representation Learning from Bio-Sonar Implementation on UAVs. In Proceedings of the IEEE International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China, 8–10 January 2021.
13. Sattar, F. A New Acoustical Autonomous Method for Identifying Endangered Whale Calls: A Case Study of Blue Whale and Fin Whale. *Sensors* **2023**, *23*, 3048. [CrossRef] [PubMed]
14. Wisdom, M.L. 2023. Available online: https://wisdomml.in/understanding-resnet-50-in-depth-architecture-skip-connections-and-advantages-over-other-networks/ (accessed on 15 August 2023).
15. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
16. Srinivas, K.; Gagana Sri, R.; Pravallika, K.; Nishitha, K.; Polamuri,S.R. COVID-19 prediction based on hybrid Inception V3 with VGG16 using chest X-ray images. *Multimed Tools Appl.* **2023**, 1–18. [CrossRef]
17. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
18. Sandler, M.; Howard, A., Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
19. Yong, L.; Ma, L.; Sun, D.; Du, L. Application of MobileNetV2 to waste classification. *PLoS ONE* **2023**, *18*, e0282336. [CrossRef] [PubMed]
20. Liu, X.; Wu, Z.Z.; Wu, Z.J.; Zou, L.; Xu, L.X.; Wang, X.F. Lightweight Neural Network Based Garbage Image Classification Using a Deep Mutual Learning. In Proceedings of the International Symposium on Parallel Architectures, Algorithms and Programming, Shenzhen, China, 28–30 December 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 212–223.