International Journal of Autonomous Robotics and Intelligent Systems
Vol. 2, No. 1, June 2026 [95-107]

Non-Destructive Ultrasound Disorder and Disease Diagnostics for Harumanis Mango using Machine Learning and Deep Learning

Muhammad Zunnurrainie Zulkifli[1*], Fathinul Syahir Ahmad Sa'ad[1], Muhamad Khairul Ali Hassan[1], Sukhairi Sudin[1], Muhammad Juhairi Aziz Safar[1], Mohd Firdaus Ibrahim[2], Alvin Tan Bon Hui[1] and Muhammad Azri Aiman bin Azahari[1]

[1]Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
[2]Faculty of Mechanical Engineering & Technology, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia

Received 12 May 2026, Revised 10 June 2026, Accepted 29 June 2026

ABSTRACT

This study presents an innovative approach in post-harvest internal qualities diagnostics for Harumanis mango using ultrasound transducer combined with machine learning and deep learning techniques. Harumanis mango is highly susceptible to disorder and diseases. Conventional methods such as visual inspection, can be time-consuming and prone to inaccuracies. To address these challenges, this study investigates the use of ultrasound waves to capture internal structural variations in Harumanis mango. The acquired ultrasound data are subsequently processed using algorithms to classify the internal qualities of the fruit. One machine learning model and one deep learning model were trained and evaluated under a three-tier framework comprising held-out test accuracy, 5-fold cross-validation and prediction performance on 10 completely unseen mango samples. The Random Forest classifier achieved a test accuracy of 84%, a cross-validation accuracy of 82.4%. The One-Dimensional Convolutional Neural Network, trained achieved a test accuracy of 87% and a diseased-class recall of 92%. This research provides valuable insights into the advancement of precision agriculture, particularly in post-harvest quality assessment and internal qualities diagnostics. Future work may focus on refining model accuracy and adapting the system for field use under various environmental conditions.

Keywords: Ultrasound Mango Quality Assessment, Post-Harvest Non-Destructive Quality Diagnostics, Machine Learning and Deep Learning for Mango Quality Diagnostics

1. INTRODUCTION

Harumanis mango, native to Perlis, Malaysia, hold substantial economic and cultural significance. Renowned for their exceptional flavour, vibrant colour and distinctive aroma, the mango has become a symbol of Malaysian agricultural excellence. Their superior quality and unique sensory attributes have attracted considerable interest from consumers worldwide. They are often marketed as "Asia's best sweet and juicy fruit" and have a limited harvesting season from April to June each year [1]. In addition, Harumanis mango are deeply embedded in the cultural identity of Perlis, where they are frequently featured in local traditions, festivals and culinary practices. Annual production was recorded at approximately 9,219 metric tonnes in 2022 [2], and their short harvest window further increases their market value. Figure 1 shows an example of a Harumanis mango.

*zunnurrainiezulkifli@gmail.com

Figure 1. Harumanis mango

1.1 Disorder and Diseases in Harumanis Mango

Harumanis mango are highly susceptible to disorder and diseases, which can significantly affect the market value of the fruit. The most common disorder and diseases affecting Harumanis mango include Anthracnose and *Bactrocera Dorsalis* infestation, and the majority of the usual disorder is Insidious Fruit Rot (IFR) [3][4][5][6][7]. These disorder or diseases can alter the natural sweetness of the mango flesh, resulting in undesirable off-flavours. The financial impact of these disorder and diseases is severe. Anthracnose, for example, can cause yield losses of between 30 % and 100 % in orchards that are not properly managed [3]. IFR is particularly problematic because infected fruits show no external symptoms, yet suffer severe internal tissue breakdown, making them commercially worthless [4]. The presence of unhealthy fruit can harm the reputation of Harumanis mango and consequently reduce their market value as a key local agricultural product. Figure 2 shows the example of Anthracnose, while Figures 3 and 4 show examples of IFR and *Bactrocera Dorsalis* infestation, respectively.

Figure 2. Anthracnose disease

Figure 3. Insidious Fruit Rot (IFR) disorder

Figure 4. *Bactrocera Dorsalis* infestation disease

1.2 Internal Quality Diagnostics in Harumanis Mango

Internal qualities diagnosis in Harumanis mango is crucial for maintaining high-quality fruits and preventing economic losses. Conventional diagnostic methods for Harumanis mango rely heavily on individual expertise, which can lead to inconsistency and inaccuracies. Early detection of the previously mentioned disorder or diseases is essential for effective quality management; however, it often requires experienced inspectors, who may be costly and difficult to access for small-scale cultivators.

With ongoing technological advancements, there is growing interest in non-destructive approaches, such as ultrasound, for quality diagnosis. This study focuses on the use of ultrasound waves to capture the internal structural characteristics of Harumanis mango. Variations in the acquired ultrasound data may serve as indicators of disease or disorder presence. In the assessment, ultrasound waves are transmitted through the fruit, and the reflected signals are subsequently analysed. The underlying principle of the acoustic impulse response method is that an object's resonant frequency is related to its physical properties, including geometry, mass and the elastic modulus of its constituent material [7].

This study aims to develop an ultrasound-based diagnostic system for detecting disorder or diseases in Harumanis mango. The initial stage of the research involves the selection and calibration of an ultrasound transducer to capture internal abnormalities within the fruit. The ultrasound signals obtained from mango samples are then prepocessed.

After signal standardization, the data are used to train and refine deep learning models to determine whether the Harumanis mango are unhealthy. Two models, Random Forest and One-Dimensional Convolutional Neural Network (1D CNN) were evaluated to identify the most effective approach. Feature extraction was performed to identify disorder or disease related patterns in the ultrasound signals. The data were subsequently trained, tuned and validated to achieve an acceptable level of classification accuracy. The performance and accuracy of each model are discussed in the later sections of this paper.

1.3 Significance of the Study

This research makes several meaningful contributions to agricultural sensing and precision farming. First, it introduces a novel application of acoustic ultrasound signal analysis, paired with a broad comparative deep learning study, to the problem of Harumanis mango qualities diagnostics. Prior research combining ultrasound with machine learning for this specific crop and disease context is limited [8][9][10], and this study addresses that gap directly.

Second, the system is designed with practical deployment in mind. Using affordable, commercially available hardware assembled into a portable, weather-resistant device means the prototype is realistically usable by smallholder farmers in the field a group that has historically had limited access to precision agriculture technology.

Third, the study applies a more thorough model evaluation approach than is typical in comparable agricultural machine learning research [11]. Rather than reporting only a single test accuracy figure, each model is assessed across three levels: held-out test accuracy, k-fold cross-validation for robustness and performance on a separate batch of completely unseen samples that simulates real-world deployment. This approach gives a more honest and useful picture of how each model would actually perform in practice.

Finally, the research supports the broader goals of precision agriculture and sustainable food production in Malaysia. Early, non-invasive disorder and disease detection allows farmers to act quickly and precisely, limiting post-harvest losses, and helping to sustain the economic viability of Harumanis mango farming in Perlis [8][12].

2. METHODOLOGY

This study developed a non-destructive disorder or disease detection system for Harumanis mango by combining ultrasound signal acquisition device with deep learning and machine learning classification models. The overall workflow consisted of few main stages: hardware system design and construction, ultrasound signal acquisition at a field orchard, machine learning and deep learning model training and evaluation, and system deployment via an interactive dashboard. Figure 5 is the deployed device used in this research.

Figure 5. Ultrasound internal quality diagnostics device

2.1 Hardware System Design

Table 1 presents all the components used, along with their respective importance and functionality, which contributed to the success of this research.

Table 1 List of hardware components

Component	Specification	Function
Raspberry Pi 4	Quad-core, 4GB RAM	Central controller and data logger
Ultrasound transducer	40kHz, 100W	Emit and receive ultrasound signals
LM358 op-amp	Dual op-amp	Amplify weak reflected signals (gain 11×)
ADS1115 ADC	16-bit, 860 SPS	Analog-to-digital signal conversion
LM2596 buck converter	12 V to 5 V	Voltage regulation for Raspberry Pi
Signal module for ultrasound transducer	40 kHz / 100 W	Powerup and signal generator for ultrasound transducers
Enclosure	IP65-rated PVC box	Weatherproof housing for field use

2.2 Data Acquisition

Harumanis mango samples were collected directly from an established commercial orchard in Chuping, Perlis, with the full cooperation and authorisation of the orchard management. Experienced orchard personnel conducted an initial visual and physical classification of each fruit into two categories: healthy mangoes (Class 0), which were free from external symptoms such as discolouration, black spots or abnormal softness; and diseased or defective mangoes (Class 1), which exhibited visible signs of disease or disorder, including soft spots, lesions and black patches associated with conditions such as Anthracnose, IFR and *Bactrocera Dorsalis* infestation.

Each classified mango was placed securely on a stable testing platform. The ultrasound transducer was positioned centrally on the fruit surface, and the Raspberry Pi was activated to emit 40 kHz pulses. The reflected signals were amplified using the amplifier, digitised by the ADS1115, and recorded by the Raspberry Pi. Each captured signal consisted of 1,000 samples representing the acoustic response over a 300 ms acquisition window.

40 kHz ultrasonic system is suitable for non-destructive diagnostic of mangoes because it can penetrate the full fruit diameter under proper coupling conditions. It can penetrate of maximum of 12 cm throughout the fruit. The receiver detects internal disorders that produce in the ultrasound signal. 40 kHz ultrasound has limited spatial resolution and may not reliably detect small or early-stage disease but it is very suitable to detect the disorder and disease for this research. The signal will then will conditioned using the op-amp.

A total of 500 mango samples were collected, comprising 250 healthy and 250 unhealthy samples, thereby forming a balanced dataset. Each signal was stored in three formats to support different model types: raw waveform arrays (.npy) for 1D CNN input and extracted numerical features (.csv) for Random Forest input.

2.3 Data Preprocessing and Feature Extraction

Raw ultrasound signals were preprocessed before feature extraction to improve signal consistency and reduce measurement noise. Each acquired waveform was first padded or clipped to a fixed length of 1,000 samples to ensure uniform input dimensions for both machine learning and deep learning models. A moving average filter was then applied to suppress short-duration high-frequency fluctuations introduced by the ADC readings, electronic noise and minor instability during signal acquisition. This filtering step preserved the main acoustic response pattern while reducing random voltage spikes that could affect noise-sensitive features such as zero crossing rate, peak count, entropy and slope changes.

After filtering, the signals were scaled to a standard range to minimise amplitude variation caused by differences in mango size, surface contact pressure and transducer positioning. This normalization step ensured that the classification models learned disease-related acoustic characteristics rather than measurement artefacts. The preprocessed signals were then used for feature extraction and as input to the 1D CNN model. The summary of the stages of data preprocessing is shown in Table 2.

Table 2 Summary of signal preprocessing stages

Preprocessing Stage	Description	Effect on Signal
Signal padding / clipping	Standardizes all signals to 1,000 samples	Ensures uniform input size for feature extraction and 1D CNN training
Moving average filtering	Reduces high-frequency noise from ADC and electronic components	Produces smoother waveform and clearer signal peaks
Signal scaling / normalization	Reduces amplitude variation caused by fruit size and transducer contact pressure	Improves consistency across samples
Feature extraction after preprocessing	Converts processed signals into numerical descriptors	Produces more reliable time-domain and frequency-domain features

Following preprocessing, twelve numerical features were extracted from each signal, covering both time-domain and frequency-domain characteristics. These features were compiled into a structured dataset (ultrasound_features.csv) containing 500 rows, 12 feature columns and one label column. Table 3 lists all extracted features and their physical significance.

Table 3 Extracted signal features

Feature	Description
Mean	Average voltage across the signal duration
Standard deviation	Spread or variability of the signal
Maximum	Peak voltage, typically linked to skin reflection
Minimum	Lowest voltage, related to absorption depth
Peak-to-peak	Total amplitude range of the signal
Skewness	Asymmetry of the signal shape
Kurtosis	Sharpness or flatness of signal peaks
Energy	Sum of squared amplitudes; reflects signal strength
Zero crossing rate	Frequency-related signal oscillation count
Entropy	Signal complexity; higher in defective samples
Peak count	Number of significant peaks; reduced in soft or decayed fruit
Slope changes	Number of derivative sign changes; indicates noise or roughness
FFT dominant Frequency	FFT dominant frequency

2.4 Machine Learning and Deep Learning Models

Machine learning and deep learning models were trained and evaluated for binary classification of mango health status. All models used an 80:20 stratified train-test split. Features were normalised using StandardScaler before training, and SelectKBest with an ANOVA F-test was applied to retain the top 10 most informative features for classical models. Hyperparameters for each model were optimised using GridSearchCV with 5-fold cross-validation.

2.4.1 Random Forest

Random Forest builds multiple independent decision trees during training, each trained on a random bootstrap sample of the data and combines their predictions through majority voting. This approach reduces overfitting and handles non-linear relationships well without requiring

extensive preprocessing. Hyperparameter tuning explored combinations of tree count (n_estimators), tree depth (max_depth), and feature selection strategy (max_features).

2.4.2 One-Dimensional Convolutional Neural Network (1D CNN)

The 1D CNN operated directly on raw. npy signal arrays, eliminating the need for manual feature engineering. The architecture consisted of a Conv1D layer with 32 filters and kernel size 3, followed by a MaxPooling1D layer, a dropout layer, a fully connected dense layer and a sigmoid output. The model learned localised patterns in the ultrasound waveform such as echo edges and amplitude variations that distinguish healthy from defective internal tissue.

2.5 Model Evaluation and Deployment

Each model was evaluated using three criteria to assess both internal and real-world performance: 1) test accuracy on a held-out 20% test set, 2) mean k-fold cross-validation accuracy (k = 5) for robustness assessment, and 3) prediction accuracy on a separate batch of 10 completely unseen mango samples to simulate genuine field deployment. Performance was further reported using precision, recall, F1-score and confusion matrices for each model.

Following evaluation, all seven trained models were deployed in a Streamlit-based interactive web dashboard. The dashboard allows users to upload new signal files in .csv, .npy, or .png format, model selection and receive an immediate classification result along with the predicted probability and performance metrics including accuracy and confusion matrix visualisation. This deployment step validated the practical usability of the system beyond the laboratory setting.

3. RESULTS AND DISCUSSION

3.1 Ultrasound Signal and Signal Propagation Analysis

A total of 500 Harumanis mango samples were collected from the commercial orchard in Chuping, Perlis, comprising 250 healthy samples (Class 0) and 250 unhealthy samples (Class 1). The balanced class distribution ensured that model training was not biased toward either class, and that evaluation metrics including precision, recall and F1-score remained reliable and meaningful throughout the study. The dataset is verified by Lembaga Pemasaran Pertanian Persekutuan (FAMA) of Perlis.

Each recorded signal consisted of 1,000 samples captured over a 300 ms acquisition window, representing the full acoustic response of the mango tissue from the initial pulse emission through to signal decay. Analysis of the propagation behaviour revealed that ultrasound waves emitted at 40 kHz underwent five distinct interaction stages within the fruit: an entry reflection at the skin surface, internal wave penetration through the flesh, possible secondary reflections from the seed core, absorption and scattering within diseased or structurally irregular tissue or disorder, and the return of echo signals to the transducer. These stages are shown and summarised in Table 4.

Table 4 Ultrasound signal propagation stages in Harumanis mango

Stage	Description
Entry reflection	Wave struck the mango skin, generating the first voltage peak
Internal penetration	Wave travelled through the flesh, oscillations reflected tissue uniformity
Core reflection	Secondary peaks occurred from dense seed or core tissue
Absorption and scattering	Defective tissue caused greater signal attenuation and loss of clarity
Return echoes	Reflected waves captured by the transducer formed the recorded signal

A clear and consistent difference was observed between healthy and unhealthy signal profiles. Healthy mangoes produced signals with stable amplitude, well-defined peaks and lower entropy values, reflecting the acoustically uniform properties of intact internal tissue. Defective mangoes, by contrast, exhibited greater signal attenuation, irregular oscillations, reduced peak definition and elevated entropy all consistent with the disrupted internal structure caused by conditions such as IFR, Anthracnose or *Bactrosera Dorsalis* infestation. These measurable physical differences confirmed that acoustic ultrasound at 40 kHz was capable of distinguishing between the internal conditions of healthy and unhealthy Harumanis mango, providing a sound physical basis for the feature extraction for deep learning and machine learning classification pipeline applied in this study. The example of one of the signal comparisons between healthy and non-healthy Harumanis mango can be seen in Figure 6.

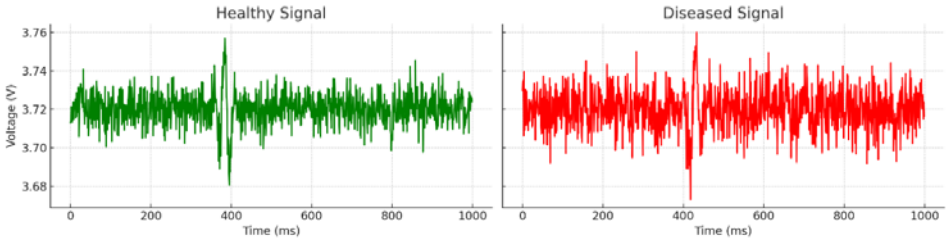

Figure 6. Signal propagation comparison between healthy and unhealthy Harumanis mango

3.2 Machine Learning and Deep Learning Model Performance

Both models were trained using an 80:20 stratified train-test split and evaluated under a three-tier framework: test accuracy on the held-out test set, mean 5-fold cross-validation (CV) accuracy for robustness assessment and prediction accuracy on 10 completely unseen mango samples to simulate genuine field deployment conditions. Features were normalised using StandardScaler prior to training and SelectKBest with an ANOVA F-test was applied for the Random Forest model to retain the most informative features.

Although the 20% held-out test partition offers a statistically reliable assessment of model performance within the experimental dataset, an additional independent validation set comprising 10 separately collected samples was used to evaluate the model's ability to generalise to entirely unseen real-world inputs. These samples were excluded from all stages of model development. This two-level evaluation approach differentiates internal cross-validation performance from external validity, thereby enhancing the credibility and robustness of the reported findings.

3.2.1 Random Forest

The Random Forest classifier was trained on the 13 extracted statistical and frequency-domain features compiled in the ultrasound_features.csv dataset. A total of 54 model configurations were evaluated through GridSearchCV with 5-fold cross-validation, exploring combinations of tree count, maximum tree depth, minimum samples per split and feature selection strategy at each node. The best configuration identified through this search was subsequently used for final model training and evaluation.

The trained Random Forest model achieved a test accuracy of 84 %, which was the highest among all feature-based models evaluated. The confusion matrix in Figure 7 shows a symmetrical and balanced classification outcome, with 42 out of 50 healthy mangoes and 42 out of 50 defective mangoes correctly classified, indicating consistent performance across both classes without bias toward either category. Precision, recall and F1-score were all approximately 0.84, further confirming the model's well-balanced classification performance. The 5-fold cross-validation accuracies were 0.88, 0.86, 0.78, 0.77 and 0.83, resulting in a mean cross-validation accuracy of 82.4 %. The close agreement between the test accuracy (84 %) and the mean cross-validation accuracy (82.4 %) suggests that the model generalised well, with minimal performance variation across different data partitions.

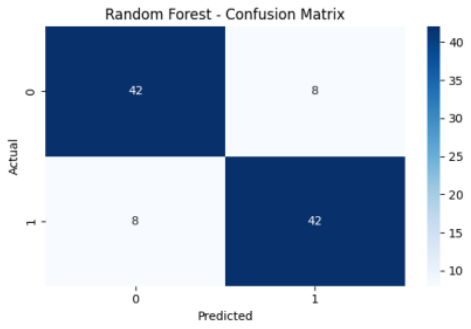

Figure 7. Confusion matrix of Random Forest model

In the independent evaluation using 10 unseen samples, the Random Forest model correctly classified 9 out of 10 samples, representing the strongest real-world generalisation performance observed in this study. The ensemble structure of the model, which combines predictions from multiple independent decision trees trained on random subsets of the data, provided inherent robustness against signal noise and the natural acoustic variability present among individual mango samples.

3.2.2 1 Dimensional Convolution Neural Network (1D CNN)

The 1D CNN was trained directly on raw .npy ultrasound waveform arrays, enabling the model to learn spatial and temporal signal features automatically without requiring manual feature engineering. The architecture consisted of a Conv1D layer with 32 filters and a kernel size of 3, followed by a MaxPooling1D layer for downsampling, a dropout layer for regularisation, a fully connected dense layer and a sigmoid output neuron for binary classification. The model was compiled using the Adam optimiser with binary crossentropy loss, and training was conducted for a maximum of 100 epochs with early stopping monitoring validation loss.

Training progressed steadily from epoch 2 onwards after initial low accuracy in the first epoch. The model reached peak validation accuracy of 87 % at epoch 22, with a training loss that dropped below 0.01 toward the final epochs, indicating effective learning of the signal patterns. The final test accuracy achieved was 87 % the highest test accuracy among all models evaluated

in this study. The confusion matrix (Figure 8) shows 41 out of 50 healthy and 46 out of 50 unhealthy mangoes correctly identified. Notably, the recall for the defective class reached 92 %, making the 1D CNN the most sensitive model for detecting infected fruit of any architecture tested.

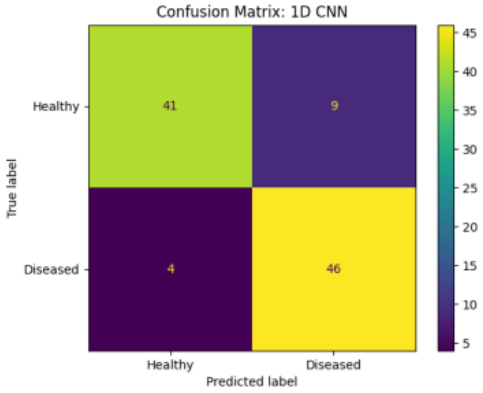

Figure 8. Confusion matrix of 1D CNN model

On the independent 10-sample unseen evaluation, the 1D CNN correctly classified 8 out of 10 samples. One false positive was recorded in which a healthy signal was classified as 'Diseased' with a high confidence score of 0.95, likely attributed to overlapping acoustic waveform characteristics between certain healthy and diseased samples, or minor signal acquisition artefacts introduced during the data collection process. Despite this isolated case, the 1D CNN demonstrated that raw signal learning was a highly effective approach for ultrasound-based mango disease classification.

Although the 1D CNN achieved encouraging performance, with a test accuracy of 87 % and a defective-class recall of 92 %, the relatively limited dataset of 500 samples remains a major constraint on the model's generalisability. In the independent 10-sample evaluation, the misclassification of a healthy signal as defective with a high confidence score of 0.95 indicates that the learned decision boundary may not yet be sufficiently robust. This suggests potential sensitivity to edge cases or subtle signal variations that were underrepresented in the training data. While the current dataset size is acceptable for preliminary research, it is insufficient to fully utilise the representational capacity of a 1D CNN architecture. Therefore, future work should prioritise dataset expansion through additional data collection campaigns and data augmentation techniques to enhance model reliability and reduce classification uncertainty.

3.3 Comparative Analysis and Model Selection

Table 5 presents a direct comparison of both models across all three evaluation criteria applied in this study. Both models demonstrated strong and reliable classification performance, although each showed strengths that align with different deployment priorities. The 1D CNN achieved the higher test accuracy at 87 %, compared with 84 % for the Random Forest, and also recorded the highest recall for the diseased class at 92 %, indicating that it is the more sensitive model for detecting infected fruit. In addition, its ability to learn directly from raw waveform arrays removes the need for manual feature engineering, thereby simplifying the data pipeline and reducing the risk of information loss during feature extraction.

In contrast, the Random Forest showed stronger real-world generalisation, correctly classifying 9 out of 10 unseen samples, compared with 8 out of 10 for the 1D CNN. It also demonstrated stable cross-validation performance across all five folds, with a mean cross-validation accuracy of 82.4 %, which supports its robustness to data variability.

Table 5 Performance comparison between Random Forest and 1D CNN models

Parameters	Random Forest	1D CNN
Test accuracy	84 %	87 %
Precision (avg)	0.84	0.87
Recall healthy class	84 %	82 %
Recall diseased class	84 %	92 %
F1-Score (Avg)	0.84	0.87
CV accuracy (5-fold)	82.4 %	-
Unseen sample accuracy	9/10	8/10

Based on the overall three-tier evaluation, Random Forest was selected as the recommended model for deployment in this system. Its combination of balanced test accuracy, strong cross-validation robustness and the highest unseen generalisation performance of 9 out of 10 made it the most consistently reliable model under conditions that best reflect real-world field deployment. The 1D CNN is recommended as the preferred alternative for future system iterations, particularly as dataset size grows, since deep learning models such as 1D CNN are known to improve substantially with larger training sets and are expected to close the generalisation gap as more field samples become available.

4. CONCLUSION

This study successfully developed and validated a non-destructive ultrasound-based disorder and disease detection system for Harumanis mango, integrating a custom-built portable hardware device with machine learning and deep learning models to distinguish healthy from diseased or defective fruit without causing any physical damage to the sample. The research was motivated by the limitations of conventional visual inspection and laboratory-based testing methods, which are unable to detect internal disease and disorder such as IFR, Anthracnose and *Bactrocera Dorsalis* that produce no visible external symptoms in the early stage yet cause severe internal tissue deterioration.

4.1 Key Findings

Ultrasound at 40 kHz was confirmed as an effective non-destructive sensing modality for detecting internal quality disorder or disease in Harumanis mango. The measurable differences in signal attenuation, entropy, peak definition and amplitude variability between healthy and diseased samples demonstrated that ultrasound waveforms carry sufficient discriminatory information to support reliable automated classification, including for unhealthy that are entirely undetectable through visual inspection.

Random Forest was identified as the most suitable model for real-world deployment based on its consistent performance across all three evaluation tiers. Its robustness to signal noise, resistance to overfitting on a moderate-sized dataset of 500 samples and superior generalisation to unseen field samples made it the most dependable classifier for practical agricultural application. The 1D CNN, while achieving higher test accuracy and diseased-class recall, is better positioned as the preferred model for future iterations of the system as the dataset expands, given the well-established tendency of deep learning architectures to improve substantially with larger training sets.

4.2 Contribution of the Research

A portable, low-cost, field-deployable ultrasound signal acquisition device was designed, built, and validated for Harumanis mango disorder and disease detection. The use of commercially available components kept the total system cost accessible to smallholder farmers. A balanced, orchard-acquired ultrasound signal dataset of 500 labelled Harumanis mango samples was constructed, stored in three complementary formats raw waveform arrays (.npy), and extracted feature files (.csv) and made available to support future research in non-destructive mango quality assessment.

A rigorous three-tier model evaluation framework was applied and validated, demonstrating its superior utility over single-metric evaluation for assessing the true deployment readiness of machine learning models in agricultural internal quality detection. The comparative analysis of a classical ensemble model against a deep learning architecture further contributed practical guidance on model selection for ultrasound-based signal.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Universiti Malaysia Perlis for providing the facilities and support necessary to conduct this research. Appreciation is also extended to all staff of Universiti Malaysia Perlis for their assistance during the research process.

REFERENCES

[1] AirAsia Grocer, 2023. Harumanis Mangoes: From Perlis to the World [Online]. Available: https://www.airasiagrocer.com/customers/pages/news-details/10

[2] IndexBox, Inc., 2025. Malaysia: Mango and Mangosteen Market 2025 [Online]. Available: https://www.indexbox.io/store/malaysia-mangoes-mangosteens-and-guavas-market-report-analysis-and-forecast-to-2020/

[3] A. K. Dofuor, N. K.-A. Quartey, A. F. Osabutey, A. K. Antwi-Agyakwa, K. Asante, B. O. Boateng, F. K. Ablormeti, H. Lutuf, J. Osei-Owusu, J. H. N. Osei, W. Ekloh, S. K. Loh, J. O. Honger, O. F. Aidoo, K. D. Ninsin, 2023. Mango anthracnose disease: the current situation and direction for future research, Front. Microbiol., vol. 14, p. 1168203, 2023, doi: 10.3389/fmicb.2023.1168203.

[4] S. Tarmizi, T. T. Malik, M. Pauziah, T. Zahrah, A. T. Sapii, T. T. Maamun, P. Muda, Z. Talib, 1993. Incidence of insidious fruit rot as related to mineral nutrients in Harumanis mangoes, MARDI Res. J., vol. 21, no. 1, pp. 43–49. http://jtafs.mardi.gov.my/jtafs/21-1/Fruit-rot.pdf

[5] UF/IFAS Extension St. Lucie County, 2020. Mango tree sooty mold [Online]. Available: https://blogs.ifas.ufl.edu/stlucieco/2020/09/21/mango-tree-sooty-mold/

[6] Business Queensland, 2024. Pink disease [Online]. Available: https://www.business.qld.gov.au/industries/farms-fishing-forestry/forests-wood/pests-diseases/trees-timber/pink-disease

[7] Plantix Plant Disease Library, Stem end rot of mango [Online]. Available: https://plantix.net/en/library/plant-diseases/100340/stem-end-rot-of-mango/

[8] N. Salleh, A. Abdullah, S. Sudin, N. Zakaria, 2023. Development of Harumanis mango insidious fruit rot (IFR) detection by utilising vibration-based sensors and PCA with random forest, J. Phys.: Conf. Ser., vol. 2641, p. 012013, doi: 10.1088/1742-6596/2641/1/012013.

[9] N. Zakaria, A. Abdullah, F. Saad, M. Hassan, M. Bakar, A. Saad, S. Sudin, M. Ibrahim, 2021. Potential of near-infrared (NIR) spectroscopy technique for early detection of insidious fruit rot (IFR) disease in Harumanis mango, J. Phys.: Conf. Ser., vol. 2107, p. 012014, doi: 10.1088/1742-6596/2107/1/012014.

[10] V. Ashok, R. Bharathi, P. Shivakumara, 2023. Building a medium scale dataset for nondestructive disease classification in mango fruits using machine learning and deep learning models, Int. J. Image, Graph. Signal Process., vol. 15, no. 4, pp. 83–95, doi: 10.5815/ijigsp.2023.04.07.
[11] W. B. Demilie, 2024. Plant disease detection and classification techniques: a comparative study of the performances, J. Big Data, doi: 10.1186/s40537-023-00863-9.
[12] A. Sharma, R. K. Bijral, J. Manhas, V. Sharma, 2022. Mango leaf diseases detection using deep learning, International Journal of Knowledge Based Computer Systems, vol. 10, no. 1, pp. 40–44.