Shiraz University

Iran Agricultural Research

Journal homepage: https://iar.shirazu.ac.ir/?lang=en



Research Article

Field evaluation of an embedded acoustic sensing system for red palm weevil detection in date palms using deep learning

Seyed Ehsan Zahedi , Mohammad Aboonajmi* , Seyed Reza Hassan-Beygi

Department of Agrotechnology, College of Aburaihan, University of Tehran, I. R. Iran

ARTICLE INFO

Keywords:
Acoustic sensing
Date palm
Deep learning
Non-destructive
Red palm weevil

Received: 08 July 2026
Revised: 22 September 2026
Accepted: 26 September 2026

ABSTRACT- The red palm weevil (RPW), *Rhynchophorus ferrugineus*, is a destructive pest of date palm whose concealed larval activity complicates non-destructive detection. This study evaluated an embedded acoustic sensing and deep-learning framework for distinguishing healthy and RPW-infested date palms under orchard conditions in Jahrom, Fars Province, Iran. Acoustic data were collected from 45 palms, comprising 23 healthy and 22 RPW-infested trees. A total of 5,355 raw recordings were acquired using a MAX9814-based microphone front end coupled to an STM32F103C8T6 microcontroller, of which 3,195 recordings passed recording-level quality control (2,255 healthy and 940 infested). The recordings were processed at 16 kHz, transformed into short-time Fourier transform spectrograms, and classified using DenseNet121. Performance was evaluated using stratified, tree-grouped five-fold cross-validation, with all recordings from an individual palm restricted to the same fold. Metrics were calculated independently for each validation fold and then averaged across the five folds. The model achieved a mean overall accuracy of 75.3% and mean balanced accuracy of 77.0%. Mean sensitivity and specificity were 81.2% and 72.8%, respectively, while infested-class precision and F1-score were 55.4% and 65.9%. The corresponding false-positive and false-negative rates were 27.2% and 18.8%. These findings support the potential of the proposed approach for field screening of RPW infestation, while the remaining classification errors indicate that acoustic predictions should complement rather than replace confirmatory inspection.

INTRODUCTION

Date palm (*Phoenix dactylifera* L.) is an economically important fruit crop in arid and semi-arid regions, particularly throughout the Middle East and North Africa. Its production is severely threatened by the red palm weevil (RPW), *Rhynchophorus ferrugineus* (Olivier), an invasive pest capable of causing extensive damage to the palm plantations (Alsanea et al., 2022). The major challenge in RPW management is that the destructive larval stages develop within the trunk and crown, where feeding activity and tissue damage can progress before clear external symptoms become evident. Once visible indicators such as exit holes, sap exudation, frass, and structural deterioration become pronounced, considerable internal damage may already have occurred (Pinhas et al., 2008; Gutiérrez et al., 2010; Hetzroni et al., 2016). Consequently, non-destructive methods capable of identifying RPW activity within the palm have received considerable research attention.

Acoustic sensing is particularly relevant to this problem because the feeding and movement of concealed RPW larvae generate mechanical and acoustic emissions that can propagate through palm tissues. Early bioacoustic studies demonstrated that these signals contain temporal and spectral characteristics that can be used to distinguish larval activity from background conditions (Pinhas et al., 2008; Hussein et al., 2010). Subsequent studies developed dedicated acoustic sensors and signal-processing approaches aimed at improving practical detection in palms (Rach et al., 2013; Hetzroni et al., 2016). More recently, distributed optical-fiber sensing has demonstrated the ability to capture RPW-associated vibrations and has expanded acoustic monitoring toward continuous and larger-scale surveillance applications (Ashry et al., 2020).

The increasing availability of machine learning and deep learning methods has further advanced automated RPW acoustic classification. Wang et al. (2021) combined fiber-optic distributed acoustic sensing with artificial neural networks and convolutional neural networks (CNNs), demonstrating the potential of learned temporal and spectral representations for RPW identification. Ashry et al. (2022) subsequently evaluated a CNN-assisted distributed acoustic sensing system under controlled and outdoor farm

*Corresponding Author: Mohammad Aboonajmi, Associate Professor, Department of Agrotechnology, College of Aburaihan, University of Tehran, I. R. Iran
E-mail address: abonajmi@ut.ac.ir
DOI:10.22099/iar.2026.56582.1767

Copyright© 2026 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publisher.

conditions and showed that environmental factors, particularly wind, can influence false-alarm performance. Deep-learning approaches based on conventional acoustic recordings have also been investigated. Boulila et al. (2023) transformed acoustic information into image-based representations and evaluated several deep neural-network architectures for distinguishing infested and non-infested signals. Torky et al. (2023) applied transfer learning with the VGGish architecture to RPW sound recognition, illustrating the potential of pretrained audio representations for this task. More recently, Kang et al. (2024) reported the large-scale field deployment of machine-learning-aided fiber-optic distributed acoustic sensing in operational date-palm farms, while Saranya and Martin (2025) combined acoustic feature extraction with a bidirectional long short-term memory model and emphasized the continuing need for field validation under diverse environmental conditions.

Collectively, these studies demonstrate substantial progress in the automated acoustic detection of RPW; however, they also show that the reported performance depends strongly on the sensing technology, dataset, environmental conditions, and evaluation protocol. Recent investigations include controlled experiments, public or benchmark acoustic datasets, transfer-learning approaches, and specialized distributed optical-fiber sensing systems (Wang et al., 2021; Ashry et al., 2022; Boulila et al., 2023; Torky et al., 2023; Kang et al., 2024; Saranya and Martin, 2025). Comparatively less attention has been given to evaluating a compact embedded microphone–microcontroller system using locally acquired recordings under orchard conditions while explicitly separating individual palms between model-development and validation folds. This distinction is important when multiple acoustic recordings originate from the same biological unit because evaluation at the recording level without maintaining tree independence can produce an overly optimistic estimate of performance on previously unseen palms.

Therefore, the objective of the present study was to evaluate an embedded acoustic sensing and deep-learning framework for distinguishing healthy and RPW-infested date palms under real orchard conditions in Jahrom, Fars Province, Iran. Acoustic recordings obtained using a MAX9814-based microphone front end and an STM32F103C8T6 microcontroller were transformed into spectrograms and classified using DenseNet121. The study hypothesis was that acoustic spectrograms acquired with this embedded sensing system would contain sufficient infestation-related information to discriminate healthy from RPW-infested palms when model performance was evaluated using a tree-independent strategy. The distinguishing contribution of the study lies in the combination of field-acquired acoustic data, compact embedded sensing hardware, spectrogram-based deep learning, and stratified tree-grouped five-fold cross-validation, in which recordings from an individual palm remained within the same fold. This evaluation strategy was adopted to assess generalization between palms rather than merely between recordings. The remainder of the paper describes the field acquisition and signal-processing methodology, presents and discusses the cross-validation results in relation to previous RPW detection studies, and concludes with the practical implications, limitations, and directions for future development.

MATERIALS AND METHODS

Study site and dataset

The study was conducted on 45 mature date palms (*Phoenix dactylifera* L. cv. Shahani) in a date-palm orchard in the Hakan area of Jahrom, Fars Province, Iran (WGS 84, UTM Zone 39N, 730,746 m E, 3,167,656 m N; 28.6154° N, 53.3600° E), from September to December 2024. The Jahrom region has a hot semi-arid climate characterized by hot, dry summers and relatively mild winters, with most precipitation occurring during the cooler months. Long-term climatic records indicate an annual mean temperature of approximately 20 °C and annual precipitation of approximately 285 mm (Jokar et al., 2023). The study population comprised 23 healthy palms and 22 RPW-infested palms. The individual palm represented the biological sampling unit, while multiple acoustic recordings were obtained from each palm at different recording times. Initial infestation status was determined at the tree level by expert entomologists using established field indicators, including exit holes, sap exudation, and frass. Following acoustic data collection, palms classified as infested were opened during subsequent treatment, and RPW infestation was directly confirmed. Palms classified as healthy were monitored for approximately six months after recording, during which no RPW-associated symptoms were observed. Accordingly, all recordings obtained from a given palm inherited the tree-level healthy or infested label of that palm.

Acoustic sensing system and recording protocol

Acoustic recordings were acquired using an embedded sensing system comprising a MAX9814-based electret microphone front end and an STM32F103C8T6 microcontroller. The MAX9814 provides microphone amplification with automatic gain control and low-noise microphone bias (Analog Devices, 2022). The analog microphone output was digitized by the microcontroller through its integrated analog-to-digital converter (ADC), while direct memory access (DMA) was used for buffered data transfer during acquisition; these functions are supported by the STM32F103x8/xB architecture (STMicroelectronics, 2025). The recordings were acquired at a sampling rate of 16 kHz and stored on an SD card.

For each measurement, the sensing probe was inserted into the palm trunk through a 10-cm-deep access hole located approximately 1 m above ground level. Because feeding and movement of concealed RPW larvae generate acoustic emissions that can propagate through palm tissues (Pinhas et al., 2008; Hussein et al., 2010), the probe was positioned within the trunk to improve coupling with internally generated signals. Recordings were collected at different times during the day and night to incorporate temporal variation in larval activity and field background conditions. The field deployment and positioning of the embedded acoustic sensing system during data acquisition are illustrated in Fig. 1.



Fig. 1. Field deployment of the embedded acoustic sensing system during data acquisition in the date-palm orchard**.**

A total of 5,355 raw acoustic recordings were collected, most with a duration of approximately 20 s. Recording-level quality control was performed before signal preprocessing and consisted of numerical signal-quality screening together with visual and aural inspection. Recordings affected by unacceptable wind or environmental noise, handling or mechanical artifacts, corrupted or incomplete acquisition, incomplete metadata, very short duration, or very weak acoustic content were excluded. After quality control, 3,195 recordings were retained for analysis, comprising 2,255 recordings from healthy palms (70.6%) and 940 recordings from RPW-infested palms (29.4%). Thus, the 45 palms represented the biological sampling units, whereas the 3,195 recordings represented repeated acoustic observations obtained from those palms. The difference between the nearly balanced numbers of healthy and infested palms and the recording-level class distribution resulted from differences in the number of recordings available from individual palms and a higher exclusion rate among recordings from infested palms during quality control. The tree-level and recording-level composition of the retained acoustic dataset is summarized in Table 1.

Table 1. Tree-level and recording-level composition of the acoustic dataset after quality control

Class	Number of palms	Retained recordings	Recording proportion
Healthy	23	2,255	70.6%
RPW-infested	22	940	29.4%
Total	45	3,195	100%

Note: RPW, red palm weevil.

Audio preprocessing and spectrogram generation

Signal processing and spectrogram generation were performed using MATLAB. The recordings were processed at their native sampling rate of 16 kHz. Each signal was baseline-corrected by linear detrending and band-pass filtered between 100 and 7,000 Hz to suppress low-frequency drift and out-of-band interference. Recordings shorter than 20 s were extended by circular repetition, whereas recordings longer than 20 s were truncated to 20 s. During model training, random circular temporal shifting was applied only to the training recordings as a data-augmentation procedure. No temporal-shift augmentation was applied to recordings in the held-out validation fold.

Time–frequency representations were generated using the short-time Fourier transform (STFT), a standard approach for representing the time-varying spectral content of acoustic signals (Allen and Rabiner, 1977). The STFT was calculated using a fast Fourier transform (FFT) size of 256 samples and a hop length of 256 samples, corresponding to non-overlapping analysis frames and 129 frequency bins. The resulting spectrograms were adjusted to a fixed length of 1,251 time frames by cropping or zero-padding where required. Spectrogram magnitude was converted to the decibel (dB) scale and represented as a single-channel input tensor of 129 × 1,251 × 1 (frequency bins × time frames × channel). The principal preprocessing parameters are summarized in Table 2.

Table 2. Signal preprocessing and spectrogram-generation parameters

Parameter	Value
Sampling rate	16 kHz
Baseline correction	Linear detrending
Band-pass range	100–7,000 Hz
Recording duration	20 s
Temporal augmentation	Random circular shift (training data only)
Fast Fourier transform (FFT) size	256 samples
Hop length	256 samples
Frequency bins	129
Fixed time frames	1,251
Spectrogram scale	Decibel (dB)
Input tensor size	129 × 1,251 × 1 (frequency bins × time frames × channel)

Deep learning model and training

The spectrograms were classified using a CNN. CNNs learn hierarchical spatial patterns directly from image-like inputs and are therefore well suited to identifying localized time–frequency patterns in spectrogram representations. DenseNet121 was selected as the classification architecture because it is a well-established CNN in which each layer receives feature maps from preceding layers, promoting feature reuse and effective gradient propagation (Huang et al., 2017). Although dense connectivity can increase computational and memory requirements, it provides an efficient mechanism for retaining both lower- and higher-level spectro-temporal information.

The final classification layer consisted of two output neurons corresponding to the healthy and RPW-infested classes, followed by a softmax activation. Model optimization was performed using Adam (Kingma and Ba, 2015) with categorical cross-entropy loss (Goodfellow et al., 2016). A batch size of 32 and a maximum training duration of 10 epochs were used in

each fold. The 10-epoch duration was selected following preliminary development trials with different training durations, which indicated that extending training beyond approximately 10 epochs did not provide a consistent improvement sufficient to justify a longer training schedule. The same maximum training duration was subsequently applied across all five cross-validation folds. For each fold, the model checkpoint associated with the lowest validation loss was retained. The model architecture and principal training parameters used in each cross-validation fold are summarized in Table 3.

Table 3. Model training configuration used in each cross-validation fold

Training parameter	Value
Architecture	DenseNet121
Output classes	Healthy/RPW-infested
Output activation	Softmax
Optimizer	Adam
Loss function	Categorical cross-entropy
Batch size	32
Maximum epochs	10
Checkpoint criterion	Lowest validation loss

Note: DenseNet121, Densely Connected Convolutional Network with 121 layers.

Stratified tree-grouped five-fold cross-validation

Model performance was evaluated using stratified, tree-grouped five-fold cross-validation. The individual palm was used as the grouping variable because multiple acoustic recordings originated from each tree. The 45 palms were partitioned into five mutually exclusive folds while approximately preserving the distribution of healthy and infested trees across folds. All recordings sharing the same tree identifier (Tree ID) were assigned to the same fold and therefore could not occur simultaneously in the training and validation subsets. Data augmentation was restricted to the training subset within each fold and was not applied to the corresponding held-out validation recordings. This group-based strategy is important for datasets containing repeated observations from the same biological unit because record-wise splitting can produce overly optimistic performance estimates when observations from the same subject or source are represented in both subsets (Tougui et al., 2021).

In each of the five iterations, approximately four-fifths of the palms (about 36 trees) were used for model training, and the remaining one-fifth (about 9 trees) served as the held-out validation fold. Each fold was used once for validation, while the remaining folds were used for training. Performance metrics were calculated independently for each validation fold. After completion of the five iterations, the final reported value for each metric was obtained as the arithmetic mean of its five fold-specific values. Therefore, this procedure assessed tree-disjoint cross-validation performance rather than the performance obtained from a random recording-level split.

Model evaluation

Model performance was evaluated using complementary recording-level classification metrics. Because the retained dataset was imbalanced at the recording level, both overall and balanced accuracy were reported together with class-specific sensitivity, specificity, precision, and F1-score. Accuracy, sensitivity, specificity, precision, and F1-score were calculated using standard binary-classification definitions (Sokolova and Lapalme, 2009), while balanced accuracy was defined as the arithmetic mean of sensitivity and specificity (Brodersen et al., 2010). For metric calculation, RPW-infested recordings were treated as the positive class and healthy recordings as the negative class. Overall accuracy was calculated as:

$$\text{Accuracy}(\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \quad \text{Eq. (1)}$$

Sensitivity (recall for the RPW-infested class) was calculated as:

$$\text{Sensitivity}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \quad \text{Eq. (2)}$$

Specificity (recall for the healthy class) was calculated as:

$$\text{Specificity}(\%) = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \quad \text{Eq. (3)}$$

Precision for the RPW-infested class was calculated as:

$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \quad \text{Eq. (4)}$$

The F1-score was calculated as the harmonic mean of precision and sensitivity:

$$\text{F1}(\%) = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad \text{Eq. (5)}$$

Balanced accuracy was calculated as:

$$\text{BalancedAccuracy}(\%) = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad \text{Eq. (6)}$$

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. The false-negative rate was calculated as 100-sensitivity, while the false-positive rate was calculated as 100-specificity. All performance metrics were computed separately for each held-out validation fold, and the final reported value for each metric was obtained as the arithmetic mean of the corresponding values across the five folds. For visual interpretation of class-specific performance, a summary row-normalized performance matrix was constructed using the mean recall values for the healthy and RPW-infested classes across the five folds.

An overview of the complete experimental and computational workflow is presented in Fig. 2.

Acoustic recordings were collected from 45 date palms, screened for recording quality, converted into STFT spectrograms, and classified using DenseNet121. Model performance was evaluated using stratified, tree-grouped five-fold cross-validation, with all recordings from an individual palm restricted to a single fold.

RESULTS AND DISCUSSION

Mean cross-validation performance

The classification results are summarized in Table 4. Performance metrics were calculated independently for each of the five stratified, tree-grouped validation folds, and the values reported here represent the arithmetic mean of the corresponding fold-specific metrics. The DenseNet121 model achieved a mean overall accuracy of 75.3% and a mean balanced accuracy of 77.0%. The

mean sensitivity for RPW-infested recordings was 81.2%, whereas the mean specificity for healthy recordings was 72.8%. The precision and F1-score for the infested class were 55.4% and 65.9%, respectively. The corresponding mean false-negative and false-positive rates were 18.8% and 27.2%, respectively.

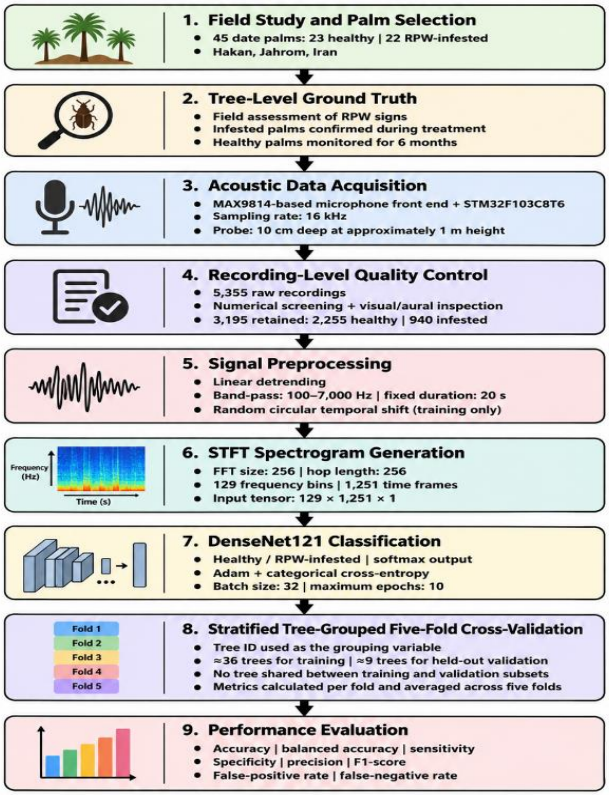


Fig. 2. Overview of the experimental and computational workflow used for red palm weevil (RPW) acoustic classification.

The difference between overall and balanced accuracy is important because the dataset was nearly balanced at the biological unit level (23 healthy and 22 infested palms) but imbalanced at the recording level, with 70.6% of all retained recordings originating from healthy palms. For context, a classifier assigning every recording in the complete retained dataset to the majority healthy class would achieve 70.6% accuracy but only 50% balanced accuracy. This illustrates why overall accuracy can be misleading under class imbalance. The observed mean balanced accuracy of 77.0%, together with the class-specific sensitivity and specificity, therefore provides stronger evidence of discrimination between the two classes than overall accuracy alone (Brodersen et al., 2010; Sokolova and Lapalme, 2009).

Class-wise performance and screening interpretation

The model showed greater recall for infested recordings than for healthy recordings. On average across the five folds, the sensitivity of 81.2% indicates that approximately four out of five RPW-infested recordings were correctly classified, whereas the corresponding mean false-negative rate of 18.8% indicates that approximately one in five infested recordings was missed. For the healthy class, the mean specificity of 72.8% corresponded to a false-positive rate of 27.2%. Thus, slightly more than one-quarter of healthy recordings generated an infested classification. The 8.4-percentage-point difference between sensitivity and specificity indicates that the classifier favored detection of infestation to some extent, while retaining meaningful discrimination of the healthy class.

Table 4. Mean recording-level classification performance across the five stratified, tree-grouped cross-validation folds

Metric	Mean across five folds (%)
Overall accuracy	75.3
Balanced accuracy	77.0
Precision – RPW-infested	55.4
Sensitivity – RPW-infested	81.2
Specificity – healthy	72.8
F1-score – RPW-infested	65.9
False-positive rate	27.2
False-negative rate	18.8

Note: RPW, red palm weevil.

The mean infested-class precision of 55.4% also requires particular consideration. This corresponds to a mean false-discovery proportion of 44.6% across the five folds, indicating that a substantial fraction of positive acoustic predictions would require subsequent verification. Precision depends on both classifier performance and class prevalence; therefore, the lower proportion of infested recordings contributed to the moderate positive predictive performance despite the comparatively higher sensitivity (Sokolova and Lapalme, 2009). For an orchard-screening application, this profile may be acceptable when an acoustic alert is treated as an indication for further inspection rather than as a definitive diagnosis. However, unnecessary follow-up caused by false-positive detections and the potential consequences of missed infestation both remain operationally important.

Fig. 3 summarizes the mean class-wise behavior across the five validation folds. The diagonal values correspond to the mean healthy-class recall (72.8%) and mean infested-class recall (81.2%), while the off-diagonal values represent the corresponding false-positive (27.2%) and false-negative (18.8%) rates. These values are recording-level cross-validation measures and should not be interpreted as the percentages of individual palms correctly diagnosed. Although Tree ID grouping prevented recordings from the same palm from occurring in both training and validation subsets, multiple recordings were still obtained from each biological unit. Therefore, the tree-level diagnostic performance remains to be established in a larger independent palm cohort.

Comparison with previous RPW acoustic studies

The performance observed in the present study is lower than that reported in some recent RPW acoustic-classification studies; however, direct numerical comparison requires caution because sensing technologies, datasets, infestation conditions, feature representations, and validation strategies differ substantially among studies. For example, Ashry et al. (2022) combined optical-fiber distributed acoustic sensing with CNN classification and reported an approximately 97% classification accuracy, including experiments conducted in an outdoor farm. Their field results also

demonstrated that environmental wind noise can substantially increase false alarms, emphasizing the challenge of transferring acoustic classification from controlled to outdoor conditions. Torky et al. (2023) reported 94.5% accuracy using VGGish-based transfer learning, while Saranya and Martin (2025) reported 98.02% accuracy using acoustic features derived from linear predictive coding (LPC) and a bidirectional long short-term memory (Bi-LSTM) classifier; the latter study explicitly identified controlled-dataset evaluation and the need for large-scale field validation as a limitation.

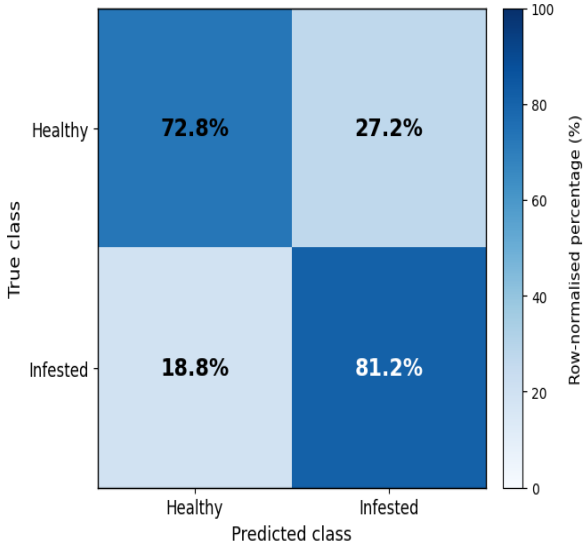


Fig. 3. Row-normalized classification performance averaged across five cross-validation folds. Diagonal cells indicate correct classification of healthy (72.8%) and RPW-infested (81.2%) recordings, while off-diagonal cells indicate false-positive (27.2%) and false-negative (18.8%) rates.

The present study used a direct STFT spectrogram representation acquired with a compact embedded sensing system and evaluated the classifier using tree-grouped cross-validation on locally collected orchard recordings. Therefore, the lower numerical performance observed here should not be interpreted simply as inferiority of the classifier architecture; it reflects a different sensing configuration, data representation, field environment, and validation design.

A methodological strength of the present evaluation is that recordings from the same palm were never distributed between the training and validation folds. This reduces the source-overlap bias that can occur when repeated measurements from the same biological unit are randomly divided between the model development and evaluation subsets (Tougui et al., 2021). Nevertheless, the present design represents internal tree-disjoint cross-validation rather than external validation, and the held-out fold was also used for validation and checkpoint selection. Accordingly, the reported results should be interpreted as internal cross-validation performance and should be confirmed on a completely independent set of palms in future studies. A comparison of the present study with selected recent machine-learning-based RPW acoustic detection studies is provided in Table 5.

Table 5. Comparison of the present study with selected recent machine-learning-based red palm weevil (RPW) acoustic detection studies

Study	Sensing / model	Evaluation context	Main result/ relevance
Ashry et al. (2022)	Optical-fiber distributed acoustic sensing (DAS) + convolutional neural network (CNN)	Controlled + outdoor farm	~97% accuracy; wind-related false alarms
Torky et al. (2023)	Acoustic embeddings + VGG-derived audio classification architecture (VGGish) transfer learning	Acoustic dataset	94.5% accuracy
Saranya and Martin (2025)	Linear predictive coding (LPC)-based features + bidirectional long short-term memory (Bi-LSTM)	Controlled dataset	98.02% accuracy; field validationh identified as a need
Present study	Embedded sensing + short-time Fourier transform (STFT) + Densely Connected Convolutional Network with 121 layers (DenseNet121)	Orchard data; tree-grouped five-fold cross-validation (CV)	75.3% mean accuracy; 77.0% balanced accuracy; tree-disjoint evaluation

Practical implications, limitations, and future work

The results support the use of the proposed system primarily as an orchard screening tool rather than as a stand-alone diagnostic method. The sensitivity of 81.2% is favorable for identifying acoustic recordings associated with infestation, but the 27.2% false-positive rate would generate a non-negligible number of unnecessary alerts. A practical deployment strategy would therefore use the acoustic classifier to prioritize palms for subsequent visual, entomological, or other confirmatory inspections rather than trigger treatment solely from an individual acoustic prediction. Similar challenges with environmental false alarms have been reported for outdoor RPW acoustic sensing, particularly under wind-induced noise (Ashry et al., 2022).

The field nature of the dataset also resulted in substantial recording rejection during quality control. Of the 5,355 raw recordings, 2,160 (40.3%) were excluded before analysis, leaving 3,195 recordings. Exclusion was associated with inadequate signal quality, environmental or handling noise, incomplete acquisition or metadata, short recordings, and very weak acoustic content. This proportion illustrates the practical difficulty of acquiring consistently usable acoustic signals in orchard environments and indicates that improved mechanical coupling, environmental-noise suppression, and more robust signal-quality assessment should be prioritized in subsequent system development.

Several limitations should be considered when interpreting the findings. First, although 3,195 recordings were retained, they originated from only 45 independent

palms. Therefore, the large number of recordings does not represent an equivalent number of independent biological samples. The use of Tree ID grouping mitigated overlap between training and validation data; however, a larger number of independent palms is required to establish population-level generalizability. Second, the reported metrics are recording-level measures rather than tree-level diagnostic outcomes. A future operational system should aggregate repeated measurements from each palm and evaluate sensitivity and specificity at the tree level. Third, the infested palms were initially identified using field symptoms and subsequently confirmed during treatment, whereas healthy palms were monitored for approximately six months without developing RPW-associated symptoms. This strengthened the reliability of the reference labels; however, because the infested cohort was initially identified from symptomatic trees, the current study does not establish performance for pre-symptomatic or very early infestation.

A further limitation is that the five-fold procedure provides internal tree-disjoint validation rather than evaluation on an entirely independent external orchard cohort. Future work should therefore include larger multi-orchard datasets, independent tree-level test cohorts, and measurements collected across different seasons and environmental conditions. Additional work should investigate improved noise-robust preprocessing, calibration of classification thresholds, alternative or ensemble deep-learning architectures, and decision-level aggregation of multiple recordings from the same palm. These developments should focus particularly on improving specificity and precision while maintaining sufficient sensitivity for screening applications.

CONCLUSION

This study demonstrates the feasibility of combining compact embedded acoustic sensing with spectrogram-based deep learning for RPW screening under real orchard conditions. An important methodological feature was the use of tree-grouped cross-validation, which ensured that recordings from the same palm were not shared between training and validation subsets and therefore provided a more appropriate assessment of generalization between palms than a random recording-level split. The results indicate that the system can distinguish acoustic recordings from healthy and RPW-infested palms with useful screening capability, although the remaining false-positive and false-negative errors preclude its use as a stand-alone diagnostic method.

The findings should be interpreted in the context of the relatively small number of independent palms and the recording-level nature of the reported performance. In addition, although infestation status was strengthened through direct confirmation during treatment and six-month follow-up of healthy palms, the present dataset was not designed to establish pre-symptomatic detection. Future development should therefore prioritize validation on larger and independent multi-orchard cohorts, tree-level decision aggregation, improved robustness to field noise, and methods for improving specificity and precision without compromising sensitivity. With such validation and refinement, the proposed framework could support rapid, non-destructive prioritization of palms for subsequent expert inspection and targeted pest-management intervention.

FUNDING

none.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization: Seyed Ehsan Zahedi and Mohammad Aboonajmi; Methodology: Seyed Ehsan Zahedi, Mohammad Aboonajmi, and Seyed Reza Hassanbeygi; Validation: Seyed Ehsan Zahedi and Mohammad Aboonajmi; Writing—original draft preparation: Seyed Ehsan Zahedi, Mohammad Aboonajmi, and Seyed Reza Hassanbeygi; Writing—review and editing: Seyed Ehsan Zahedi, Mohammad Aboonajmi, and Seyed Reza Hassanbeygi; Project administration: Mohammad Aboonajmi; Funding acquisition: Mohammad Aboonajmi.

DECLARATION OF COMPETING INTEREST

The authors have no relevant financial or non-financial interests to disclose.

ETHICAL STATEMENT

This study did not involve human participants or vertebrate animals; therefore, it did not require approval from a research ethics committee. The work was conducted on privately owned date palm trees, and permission to access the orchard and to carry out the measurements on the trees was obtained from the owner. The manuscript does not contain any data, images, or information relating to identifiable individuals; therefore, no consent to publish was required.

DATA AVAILABILITY

The datasets generated and analyzed in the current study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

The authors sincerely thank the entomologists who confirmed the infestation status of the trees, the Jahrom County Agricultural Jihad Management for their assistance with the fieldwork and data collection, and the orchard owner for granting access to the trees and permission to conduct the measurements.

REFERENCES

Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE*, *65*(11), 1558–1564. https://doi.org/10.1109/PROC.1977.10770

Alsanea, M., Habib, S., Khan, N. F., Alsharekh, M. F., Islam, M., & Khan, S. (2022). A deep-learning model for real-time red palm weevil detection and localization. *Journal of Imaging*, *8*, 170. https://doi.org/10.3390/jimaging8060170

Analog Devices. (2022). *MAX9814: Microphone amplifier with AGC and low-noise microphone bias (Rev. 4)*. Analog Devices, Inc.

Ashry, I., Mao, Y., Al-Fehaid, Y., Al-Shawaf, A., Al-Bagshi, M., Al-Brahim, S., Ng, T. K., & Ooi, B. S. (2020). Early detection of red palm weevil using distributed optical sensor. Scientific Reports, 10, 3155. https://doi.org/10.1038/s41598-020-60171-7

Ashry, I., Wang, B., Mao, Y., Sait, M., Guo, Y., Al-Fehaid, Y., Al-Shawaf, A., Ng, T. K., & Ooi, B. S. (2022). CNN-aided optical fiber distributed acoustic sensing for early detection of red palm weevil: A field experiment. *Sensors, 22*, 6491. https://doi.org/10.3390/s22176491

Boulila, W., Alzahem, A., Koubaa, A., Benjdira, B., & Ammar, A. (2023). Early detection of red palm weevil infestations using deep learning classification of acoustic signals. *Computers and Electronics in Agriculture, 212*, 108154. https://doi.org/10.1016/j.compag.2023.108154

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In 2010 20th International Conference on Pattern Recognition (pp. 3121–3124). IEEE. https://doi.org/10.1109/ICPR.2010.764

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Gutiérrez, A., Ruiz, V., Moltó, E., Tapia, G., & Téllez, M. del M. (2010). Development of a bioacoustic sensor for the early detection of red palm weevil (*Rhynchophorus ferrugineus* Olivier). *Crop Protection, 29*, 671–676. https://doi.org/10.1016/j.cropro.2010.02.001

Hetzroni, A., Soroker, V., & Cohen, Y. (2016). Toward practical acoustic red palm weevil detection. *Computers and Electronics in Agriculture, 124*, 100-106. https://doi.org/10.1016/j.compag.2016.03.018

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4700–4708). https://doi.org/10.1109/CVPR.2017.243

Hussein, W. B., Hussein, M. A., & Becker, T. (2010). Detection of the red palm weevil *Rhynchophorus ferrugineus* using its bioacoustics features. *Bioacoustics, 19*, 177-194. https://doi.org/10.1080/09524622.2010.9753623

Jokar, P., Masoudi, M., & Pradhan, B. (2023). A new spatial model for ecological suitability assessment of irrigated farming in Jahrom County, Iran. *Revista Caatinga, 36*(1), 177-188. https://doi.org/10.1590/1983-21252023v36n119rc

Kang, C. H., Ashry, I., Marin, J. M., Ariff, T., Rjeb, A., Gunawan, W. H., Ng, T. K., & Ooi, B. S. (2024). On the large-scale field experiment of machine learning-aided fiber-optic distributed acoustic sensing for early detection of red palm weevils. In CLEO 2024, Technical Digest Series, Paper SF3A.5. Optica Publishing Group. https://doi.org/10.1364/CLEO_SI.2024.SF3A.5

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA. https://doi.org/10.48550/arXiv.1412.6980

Pinhas, J., Soroker, V., Hetzroni, A., Mizrach, A., Teicher, M., & Goldberger, J. (2008). Automatic acoustic detection of the red palm weevil. *Computers and Electronics in Agriculture, 63*, 131-139. https://doi.org/10.1016/j.compag.2008.02.004

Rach, M. M., Gomis, H. M., Granado, O. L., Malumbres, M. P., Campoy, A. M., & Martín, J. J. S. (2013). On the design of a bioacoustic sensor for the early detection of the red palm weevil. *Sensors, 13*(2), 1706-1729. https://doi.org/10.3390/s130201706

Saranya, S., & Martin, B. (2025). Machine learning-enabled acoustic sensing for RPW infestation detection. *Scientific Reports, 15*, 38391. https://doi.org/10.1038/s41598-025-22306-6

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

STMicroelectronics. (2025). STM32F103x8, STM32F103xB: Medium-density performance line Arm®-based 32-bit MCU with 64 or 128 KB Flash, USB, CAN, 7 timers, 2 ADCs, 9 communication interfaces (DS5319 Rev. 20). Retrived from: https://www.st.com/resource/en/datasheet/stm32f103rb.pdf.

Torky, M., Dahy, G., & Hassanien, A. E. (2023). Recognizing sounds of red palm weevils (RPW) based on the VGGish model: Transfer learning methodology. *Computers and Electronics in Agriculture, 212*, 108079. https://doi.org/10.1016/j.compag.2023.108079

Tougui, I., Jilbab, A., & El Mhamdi, J. (2021). Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. *Healthcare Informatics Research, 27*(3), 189-199. https://doi.org/10.4258/hir.2021.27.3.189

Wang, B., Mao, Y., Ashry, I., Al-Fehaid, Y., Al-Shawaf, A., Ng, T. K., Yu, C., & Ooi, B. S. (2021). Towards detecting red palm weevil using machine learning and fiber optic distributed acoustic sensing. *Sensors, 21*(5), 1592. https://doi.org/10.3390/s21051592