Journal of Scientific & Industrial Research
Vol. 85, February 2026, pp. 140-149
DOI: 10.56042/jsir.v85i2.20761

Fusion-Driven Acoustic Intelligence for Insect Detection in Grain Storage

Rojalin Mishra, Tusar Kanti Dash* & Ganapati Panda
Electronics and Communications Engineering, C. V. Raman Global University, Bhubaneswar, India

Received 17 June 2025; revised 27 August 2025; accepted 13 January 2026

Insect infestations in stored grains continue to pose a serious threat to food quality, safety, and supply, often leading to significant post-harvest losses. Studies show that insects alone contribute to more than 10% of the total post-harvest losses in grains and cereals. Traditional detection methods are not only time-consuming and labor-intensive but also prone to inaccuracies. To address these challenges, this study presents a noninvasive and scalable approach that combines acoustic signal analysis with feature fusion techniques. By analyzing insect-generated sounds from three standard datasets, the system captures movement and feeding activity. A fusion of spectral, cepstral, and statistical features enhances detection performance, while a phase-based speech enhancement method helps reduce background noise for clearer signal interpretation. These features are then used with standard audio classification models to determine insect presence and activity levels. Experimental results show the method achieves an average detection accuracy of 94%. Designed to be both practical and efficient, this solution offers a reliable way to protect stored grains and reduce losses across large-scale storage facilities.

Keywords: Artificial intelligence, Insect detection, Machine learning, Smart farming, Speech recognition

Introduction

Due to their dual roles as pollinators and pests, insects are an integral part of agriculture. Although beneficial insects help produce crops, pest species can result in large yield losses. Reducing crop damage, avoiding the need for chemical pesticides, and enabling prompt intervention all depend on efficient insect detection. New developments in technology have improved the ability to monitor insects accurately and in real time. Insect pests are the primary cause of reduced crop yield and quality globally. Traditional methods of pest control often involve extensive pesticide use, which can have adverse environmental effects. Accurate detection allows for targeted interventions, reducing unnecessary chemical applications and promoting sustainable farming practices.[1] The integration of deep learning techniques, such as Convolutional Neural Networks (CNNs), has revolutionized insect detection. These models can process vast datasets, identifying and classifying insect species with high accuracy, even in complex field conditions.[2] The creation of large-scale datasets, like Insect-1M, has facilitated the training of robust models capable of recognizing a wide variety of insect species. Such datasets encompass diverse insect images, aiding in the development of generalized detection systems applicable across different regions and crops.[3] Implementing real-time monitoring systems using AI and IoT technologies enables continuous surveillance of insect populations. These systems can promptly detect pest outbreaks, allowing for immediate response and minimizing potential crop damage.[4] Hyperspectral imaging offers a non-invasive method to detect and monitor insect pests by capturing detailed spectral information. This technology can identify specific pest species based on their unique spectral signatures, even in challenging field environments.[5] Sound-based detection systems analyze the acoustic signals produced by insects. By employing AI algorithms, these systems can identify pest presence through their characteristic sounds, providing an alternative detection method, especially in dense crop canopies where visual monitoring is challenging.[6] Insect detection is not limited to the field; it is equally critical in post-harvest storage. AI-based detection systems in grain facilities help in early identification of infestations, ensuring the quality and safety of stored products.[7] Acoustic analysis has emerged as a promising non-invasive method for monitoring insect activity, leveraging the

---

*Author for Correspondence
E-mail: tusarkantidash@gmail.com

unique sounds produced by insects during various behaviors such as feeding, movement, and communication. Deep learning techniques have been successfully applied to analyze acoustic signals for the early detection of insect infestations. For the detection of Red Palm Weevil infestations, the acoustic signals have been captured from infested palms. Their approach involved converting sound data into image-like representations for better feature extraction, leading to high accuracy in detecting infestations even at early stages.[8] A study on acoustic analysis highlighted the benefits of non-invasive monitoring, which allows for real-time detection without disturbing the stored products. This method not only reduces labor costs but also minimizes the need for chemical treatments by enabling targeted interventions.[9] Low-cost acoustic systems have been explored for detecting pests in stored rice grains. A recent study demonstrated that specific sound signatures could be reliably associated with insect activity, allowing for efficient monitoring in storage facilities. This low-cost solution is particularly advantageous for resource-limited settings where traditional pest detection methods are impractical.[10] A pest detection system is designed to combine sound analytics with the Internet of Agricultural Things infrastructure. This model employed deep learning for signal classification, achieving high sensitivity and accuracy. Such integration facilitates real-time monitoring, making pest control measures more timely and effective.[11] Noise-robust acoustic systems are being developed to overcome challenges in greenhouse environments, where background noise is typically high. Recent research focused on enhancing the accuracy of insect detection in noisy conditions, allowing for reliable monitoring even in controlled agricultural settings.[12] Feature fusion is a vital technique in audio engineering that significantly enhances the performance and robustness of audio processing systems. By integrating complementary features from different domains, it improves the accuracy, noise resilience, and generalization capabilities of models. In speech emotion recognition, combining pitch, Mel-Frequency Cepstral Coefficients (MFCC), and delta MFCC features captures emotional nuances more effectively than individual features alone.[13] Beyond speech, feature fusion has proven valuable in biomedical signal processing; Bahreini et al. introduced a method that merges handcrafted MFCC features with deep CNN-extracted features, resulting in highly accurate heart sound classification.[14] In marine bioacoustics, dual-feature fusion with MFCC and Delay-Doppler features to enhance the recognition of marine mammal signals amid ocean noise has been used.[15] Insects produce characteristic acoustic signals through wing beats, stridulation, or movement, which can be analyzed for species identification and behavioral studies.[16] Pest insects in particular warrant close attention due to the significant damage they inflict on crops, stored products, and human health.[17] Acoustic monitoring offers a non-invasive, real-time method for early detection of pest activity, enabling timely and sustainable management interventions.[18] This approach holds remarkable promise as a tool for integrated pest surveillance in diverse environments.[9] Recent studies highlight that acoustic-based insect detection has advanced considerably with the rise of deep learning and feature fusion strategies. Large curated datasets such as InsectSound1000 and InsectSet459 have provided strong foundations for training robust models across diverse insect species.[19,21] Research integrating dual-frequency and spectral feature fusion has shown that combining cepstral and spectral cues significantly improves noise resilience and classification accuracy compared to single-feature approaches.[20,21] Practical deployments have also emerged, where low-cost MEMS microphones paired with lightweight CNNs demonstrate the feasibility of real-time detection in grain storage environments, although accuracy under noisy conditions remains a challenge.[22] These findings collectively point to the importance of fusion-driven acoustic intelligence in overcoming the limitations of traditional single-feature models, especially for reliable insect monitoring in post-harvest storage systems.[23]

It has been observed from the brief literature review that feature fusion not only boosts classification accuracy but also strengthens models against noise and domain variability, making it a cornerstone technique in modern audio processing. Despite its proven effectiveness across multiple domains, feature fusion techniques remain largely unexplored in agricultural applications, particularly for insect detection. Given the rising need for automated, real-time pest monitoring to reduce crop damage and improve yield, integrating feature fusion strategies with acoustic insect detection could be transformative. Traditional methods rely mostly on single-feature analysis, which struggles with noise and overlapping sound signatures in field environments. Applying multi-feature fusion, similar to successful approaches in audio engineering, could significantly enhance detection accuracy and robustness in variable outdoor conditions.

Even with many technological advances, current insect detection methods still struggle in grain storage settings. Image-based systems require high-quality pictures with controlled lighting, which is difficult to achieve inside dark storage bins. Hyperspectral imaging can identify pests very accurately, but the equipment is expensive and requires heavy computation. It is also very sensitive to environmental changes and impractical in everyday storage facilities. Acoustic detection is a promising non-invasive alternative, but most approaches rely on just one type of feature, like MFCCs or spectrograms. This makes them vulnerable to background noise and limits their ability to capture the full variety of insect sounds. As a result, models often fail to perform well across different grain types, insect species, or real-world storage conditions. There have been efforts to use deep learning and IoT-based systems for monitoring, but they remain mostly experimental due to the challenges of noise resilience and limited processing power in field settings. Till now, very few works have been done in fusion-driven acoustic analysis for insect detection. In other fields like speech recognition, combining multiple features has already shown big improvements in accuracy and robustness. Bringing this idea into insect detection could make systems far more reliable, even in noisy and complex storage environments. This gap shows the clear need for fusion-based approaches that can provide scalable, practical, and sustainable solutions to protect stored grains.

The agricultural sector is increasingly adopting smart technologies for precision farming, yet insect monitoring still relies heavily on manual observation or simplistic acoustic analysis. Leveraging feature fusion techniques could enable automated, real-time pest detection systems that are more accurate and resilient to noise, much like how it has advanced speech, biomedical, and marine audio processing. Integrating complementary acoustic features such as time-frequency representations, spectral features, and temporal patterns could help differentiate insect species and identify infestations earlier, leading to better pest management and reduced pesticide use. Addressing this gap could revolutionize how pest populations are monitored in smart farming, enhancing both productivity and sustainability. Based on the research gap and motivation, the research objectives of the paper are listed below:

- To quantify the improvement in detection accuracy achieved through multi-feature fusion techniques for acoustic insect pest detection in agricultural environments, measured using metrics such as Accuracy, precision, recall, F1-score, and ROC-AUC.
- To design and implement a noise-resilient insect monitoring system by integrating advanced feature fusion strategies, and to validate its robustness under an ablation study.
- To determine the robustness and generalization ability of the proposed machine learning framework by benchmarking its performance across multiple datasets, using cross-dataset validation and comparative statistical analyses.

Materials and Methods

The details of the database used and the method to develop the proposed insect detection models are presented in this section.

Datasets

In this work, A total of three datasets have been used in this paper, which contain mainly three types of insects. The details of these datasets are listed in Table 1, and the spectrograms of one audio sample are shown in Fig. 1.

Table 1 — Details of the datasets used in the simulation

Name of Dataset	Insect	Categories	Number of samples in each class	Class distribution (%)
Dataset-1	Rice weevil (*Sitophilus oryzae*)	Low	497	30%
		Medium	555	33%
		High	608	37%
Dataset-2	Pulse beetle (*Callosobruchus chinensis*)	Low	336	51%
		High	326	49%
Dataset-3	Mixed insects (Orthoptera and Cicadidae)	Orthoptera	147	44%
		Cicadidae	188	56%

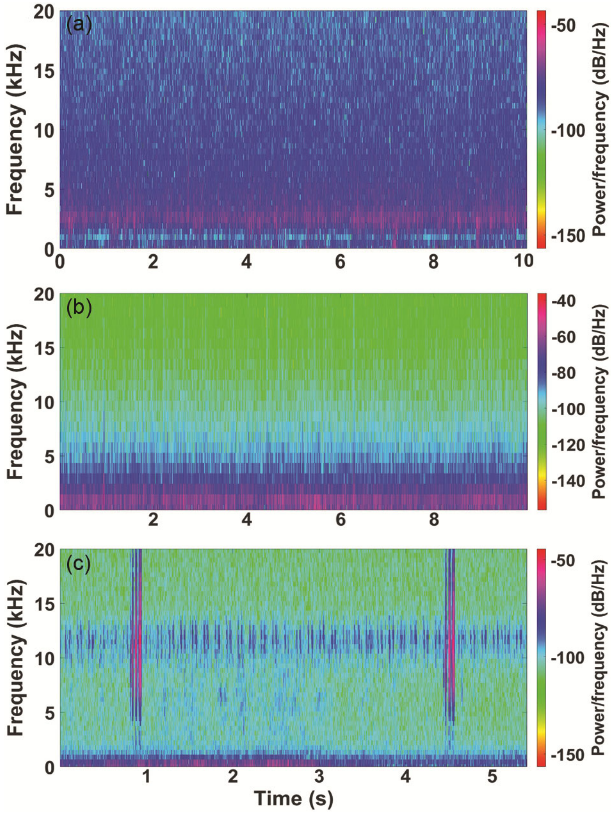

Fig. 1 — Spectrogram of one audio sample of: (a) Dataset-1, (b) Dataset-2, (c) Dataset-3

Dataset-1

Dataset-1 has been downloaded from the stored food grains dataset, which has been created from stored rice grains (*Oryza sativa*) artificially infested with *Sitophilus oryzae* (Rice Weevil).[24] The *Sitophilus oryzae* is one of the most common and destructive pests of cereals. A total of three levels of infestation, including low, medium, and high, are being prepared for data collection. The acoustic signals have been recorded using a standardized audio recording setup with a sampling frequency of 48 kHz. Each recording is of an average of 30 seconds duration. In total, 1,660 audio samples have been collected to form a comprehensive dataset for training and evaluation of detection and classification models. The spectrogram of one signal from the dataset is shown in Fig. 1(a). It has been observed that the energy mainly lies in the lower frequency range, concentrated below 5 kHz. The pattern is steady across time, indicating a continuous buzzing of Rice Weevil.

Dataset-2

Dataset-2 has been collected from stored green gram (*Vigna radiata*) infested with *Callosobruchus chinensis* (Pulse Beetle).[24] It is a major pest that causes significant post-harvest losses in pulses. In this case, two levels of infestation, including low and high, are considered to reflect practical storage conditions. Audio recordings have been obtained under controlled conditions at a sampling frequency of 48 kHz, similar to dataset-1. Each audio sample has been recorded for a duration of 10 seconds. It is focused on capturing short-duration activity signals of the insect. A total of 640 samples are being collected to provide storage pest detection and classification. In Fig. 1(b), the spectrogram displays a broader spread of sound frequencies for a sample from dataset-2. The energy is stronger in the mid and higher frequency ranges, giving the impression of a more intense and sustained sound from Pulses in dataset-2.

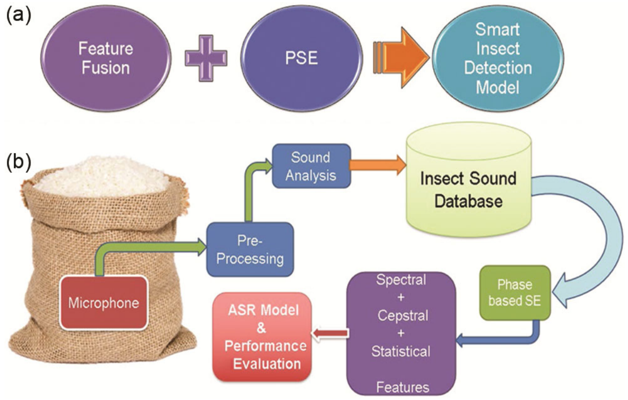

Fig. 2 — Block diagram of the proposed insect detection scheme

Dataset-3

This Dataset contains the insect sounds from the agricultural fields and has mainly two categories of insect families, including nine species belonging to the *Orthoptera* and twenty-three species from *Cicadidae*.[25] A total of three hundred thirty-five sound recordings having a total duration of 57 minutes in 44.1 kHz WAV format have been used in this dataset. The spectrogram of Fig. 1(c) demonstrates one signal from the dataset-3 with distinct vertical bands around 1 second and 4 seconds. These bursts of energy cover a wide frequency range, reflecting sharp pulses or chirps instead of a continuous hum from Orthoptera in dataset-3.

The proposed Fusion-Driven smart Insect Detection in Grain Storage is shown in Fig. 2. This method can be divided into two main parts, including smart feature fusion and audio recognition schemes. Accordingly, the block diagram has been divided into two main blocks: Fig. 2(a) & (b) where Fig. 2(a) depicts the development of the improved Phase-

Aware Speech Enhancement framework along with the framing and windowing but Fig. 2(b) depicts the details of the feature fusion.

Preprocessing

Preprocessing of the audio signal is a critical step in building reliable and efficient audio recognition systems. The preprocessing stage typically includes framing the audio signals into segments of 25-ms duration, followed by windowing each frame using a 25-ms Hamming window. This approach helps to minimize spectral leakage and maintain temporal resolution, which is crucial for effective feature extraction. Denoising is crucial in insect sound analysis due to ambient noise from wind, rain, machinery, and other environmental factors that can obscure insect acoustic signals, making detection and classification challenging. Traditional methods like wavelet transforms often struggle with real-world noise, while AI-driven denoising models have shown significant improvements.[26] Effective denoising thus enhances insect monitoring, contributing to sustainable agricultural practices. For the denoising purpose, speech enhancement algorithms are used with the aim of improving the quality and intelligibility of speech signals in noisy environments. In a recent work, an Improved Phase-Aware Speech Enhancement framework has been proposed that combines bio-inspired techniques mimicking auditory processing mechanisms with Artificial Neural Networks (ANNs) to enhance both amplitude and phase components of speech signals.[27] The bioinspired model captures auditory perception characteristics, while the ANN is used for phase estimation. This phase-based speech enhancement technique (PSE) has been used as the first step of the proposed insect detection method.

Features Fusion

For suitable feature extraction, various spectral and cepstral features are extracted at the frame level, and the statistical features are extracted at the sample level.[28] Spectral and cepstral features are essential concepts in speech and audio signal processing, providing insights into the structure and characteristics of sound. Spectral features focus on analyzing the distribution of energy across different frequencies in a signal. Key attributes include the spectral centroid, which represents the perceived brightness of the sound by identifying the "center of mass" of its spectrum; spectral bandwidth, which measures how spread out the frequencies are around that center; spectral flux, which tracks changes in the spectrum over time; and spectral kurtosis, which examines how sharp or flat the spectral peaks are, helping distinguish different sound types. These features are widely used in applications like speech recognition and music analysis. For this purpose, the MATLAB-based audio Feature Extractor function is used.[29] On the other hand, cepstral features transform the spectral information to highlight how these frequency components change over time, which helps separate the source of the sound from its vocal tract characteristics. A fundamental cepstral feature is the cepstrum, obtained by applying an inverse Fourier transform to the logarithm of the spectrum. Another important cepstral feature is the MFCC, which captures essential speech characteristics by mapping frequencies to the mel scale, reflecting how humans perceive sound. Together, these features play a critical role in analyzing, recognizing, and processing speech and audio signals effectively. The details of these features, along with the dimensionality, have been plotted in Table 2, where *n* is the number of frames in each audio sample.

Simulation-based Experiment

The simulation-based experiment for the proposed insect-level detection method is implemented in the following steps, beginning with data collection, where relevant datasets are gathered in the domain of insect sounds. Next, denoising using the PSE scheme, followed by preprocessing and feature extraction, are performed, which contribute to effective diagnosis. Following this, the features undergo fusion and sample-level sorting for consistency and to improve the performance of the model. The processed data is then sent for classification, where machine learning

Table 2 — Details of the feature fusion dimensions

Name of feature	Feature dimension (Frame level)	Feature dimension (Sample level)
Linear spectrum	n*512	1024
Mel spectrum	n*512	1024
Bark spectrum	n*32	64
Equivalent rectangular bandwidth spectrum	n*32	64
MFCC, MFCC delta, MFCC delta delta	n*39	78
Gammatone cepstral coefficients (GTCC), GTCC delta, GTCC delta delta	n*39	78
Total dimension	n*1166	2332

Table 3 — Model parameters

Model	Features	Classifier	Parameters (Best Settings)
1	Cepstral features	Random forest	n_estimators = 50; criterion = gini
2	Cepstral features	SVM	log2c = 1; log2g = 0.7701; kernel = Linear
3	Spectral	KNN	n_neighbors = 5; algorithm = auto; weights = distance
Proposed	Feature fusion	SVM	log2c = 1; log2g = 0.7701; kernel = Linear

algorithms are used for insect-level detection based on extracted features. Subsequently, the model undergoes training and validation to optimize its accuracy and generalization capabilities. Finally, performance measures are evaluated to assess the model's performance, including metrics such as accuracy, precision, recall, and F1-score.

Results and Discussion

To assess the performance and reliability of the proposed insect detection model, a k-fold cross-validation method is used. This approach involves dividing the original dataset into multiple subsets, known as "folds." During each iteration, one fold is set aside for testing, while the remaining folds are used for training the model. This process repeats k times, ensuring that each data point is used for both training and validation exactly once. In the proposed implementation, the k value has been taken as 5 to make it a five-fold cross-validation. By averaging the results across all iterations, a more reliable estimate of the model's performance is obtained, reducing the risk of overfitting and improving generalization. In this study, a 5-fold stratified cross-validation is specifically applied. Unlike basic k-fold cross-validation, the stratified version maintains the original distribution of class labels within each fold. This is crucial, where the number of positive and negative cases can be imbalanced. This strategy prevents any one class from dominating the training or testing process, leading to more balanced and fair evaluations. The evaluation is carried out across five separate datasets as well as a combined dataset that consolidates all the samples. The evaluation has been carried out across three standard datasets to provide insights into the generalization capability across different insect types and data sources. To measure performance, key metrics such as accuracy, precision, recall, F1-score, and the Area Under the ROC Curve (AUC) are calculated during each iteration.[28] By leveraging stratified cross-validation, the study ensures that the model is not only accurate but also robust and reliable across varied datasets, making it a practical tool for real-world scenarios. Several standard machine learning-based models are used as a classifier, including Random Forest (RF), SVM, and KNN.[16,30,31] Grid search is used to find the optimal parameters of the classifiers, and these parameters have been listed in Table 3. The model parameters of all the baseline and the proposed models are shown in Table 3. The Model-1 with Random Forest has been configured with n_estimators = 50, which provides a balance between accuracy and efficiency, and the gini criterion has been used to evaluate split quality. The model-2 with SVM has been set with log2c = 1 as the penalty parameter and log2g = 0.7701 as the kernel coefficient, while the linear kernel has been selected for its effective performance on the feature space. In the model-3, the KNN model has been designed with n_neighbors = 5, the auto algorithm for neighbor search, and distance-based weighting so that nearer neighbors have a stronger influence on predictions. Finally, the proposed feature fusion framework has been assessed using an SVM, a similar setup to model-2, which demonstrates the effectiveness of combining features for improved classification. SVMs have proven to be highly effective in audio signal processing, particularly for classification tasks where robustness and precision are essential. Further, in general audio signal classification, SVMs have outperformed Radial Basis Function Neural Networks when using hybrid MFCC and energy features.[32] A comparative study also highlights that SVMs consistently deliver improved performance over methods like k-Nearest Neighbors and Gaussian Mixture Models, especially when combined with rich feature sets.[33] These findings underline SVM's strong generalization and ability to handle complex, high-dimensional feature spaces in audio processing tasks.

The evaluation results from Table 4 indicate the superior performance of the proposed model over the baseline models, including Model-1(Cepstral Features + RF), Model-2 (Cepstral Features + SVM), and Model-3 (Spectral + KNN) across all datasets. In Dataset-1, binary classification (Low and high), the proposed model provides a consistent value of 0.94 for accuracy, F1, Precision, Recall, and AUC as compared to the other methods, which range between

Table 4 — Performance comparison in multiple datasets

Dataset Name	Evaluation measures	Model-1	Model-2	Model-3	Proposed model
Dataset- 1 (Two class)	Accuracy	0.92	0.90	0.89	0.94
	F1	0.92	0.90	0.89	0.94
	Precision	0.92	0.90	0.89	0.95
	Recall	0.91	0.90	0.88	0.94
	AUC	0.91	0.89	0.88	0.94
Dataset- 2 (Two class)	Accuracy	0.93	0.91	0.91	0.95
	F1	0.93	0.91	0.91	0.95
	Precision	0.93	0.91	0.91	0.95
	Recall	0.92	0.91	0.91	0.95
	AUC	0.92	0.90	0.90	0.94
Dataset- 3 (Two class)	Accuracy	0.94	0.91	0.90	0.94
	F1	0.94	0.92	0.90	0.94
	Precision	0.94	0.91	0.90	0.94
	Recall	0.93	0.91	0.89	0.94
	AUC	0.94	0.91	0.89	0.94
Combined Dataset (Dataset-1 + Dataset-2)	Accuracy	0.93	0.91	0.90	0.96
	F1	0.93	0.91	0.90	0.96
	Precision	0.93	0.91	0.90	0.95
	Recall	0.92	0.91	0.90	0.95
	AUC	0.92	0.90	0.90	0.95
Dataset-1 (Multi class)	Accuracy	0.90	0.90	0.89	0.92
	F1	0.90	0.90	0.89	0.92
	Precision	0.91	0.90	0.89	0.92
	Recall	0.90	0.90	0.88	0.92
	AUC	0.90	0.89	0.88	0.92

0.89 and 0.92. This trend continues with Dataset-2, where the proposed model reaches 0.95 across the board, clearly surpassing the alternatives that peak around 0.91–0.93. The gains are particularly evident in AUC, indicating stronger overall classification performance.

In Dataset-3, the proposed method remains strong, matching the highest accuracy of 0.94 seen with Cepstral + RF, but outperforming in terms of F1, Precision, Recall, and AUC. When the datasets are combined, the Proposed Method scales effectively, reaching an impressive 0.96 in accuracy and 0.95 for all other measures, while competing methods max out at 0.93. This demonstrates not only its adaptability to different data distributions but also its ability to consistently minimize both false positives and false negatives.

This confusion matrix gives a qualitative analysis to understand the picture of how well the model separates low and high insect infestation levels. It has been shown in Fig. 3, and it has been observed that the model gets it right almost all the time, correctly identifying 97% of the low cases and 95% of the high cases. Only a small portion of samples are mixed up, with 3% of low infestations mistaken for high and 5% of high infestations mistaken for low. The high true positive rates and very low false negatives mean the system is both reliable and consistent. The model is very good at indicating the low from high infestations, which is exactly what is needed for practical grain storage monitoring and quick, accurate pest control decisions.

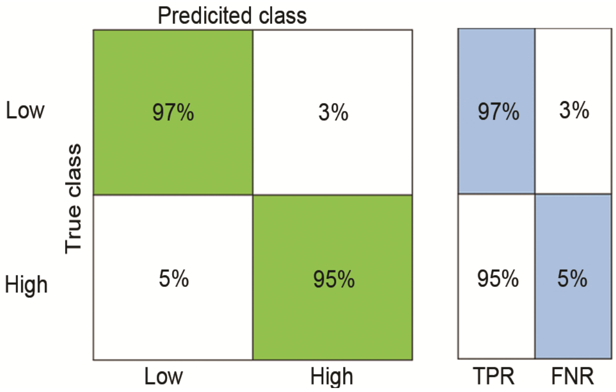

Fig. 3 — Confusion matrix of the proposed model in Dataset-2

Overall, the Proposed Method proves to be more effective and reliable, making it well-suited for real-world scenarios where both accuracy and robustness are crucial. The evaluation results for the multiclass classification demonstrate that the Proposed Model outperforms the baseline models (Model-1, Model-2, and Model-3) in Dataset-3. In Dataset-3, the proposed approach achieves the highest performance across all evaluation metrics, including Accuracy, F1, Precision, Recall, and AUC, each recorded at 0.93. This marks a clear improvement over the baseline models, which achieve values ranging from 0.88 to 0.91. This consistent gain suggests stronger classification capabilities and better class discrimination. A similar trend is observed in Dataset-1, where the Proposed Model records 0.92 for all metrics, outperforming the baseline methods that hover around 0.89 and 0.90. These results indicate that the Proposed Model is more effective at capturing complex patterns in the data, resulting in higher accuracy and better recall across multiple classes. Its ability to consistently outperform traditional models highlights its robustness and suitability for real-world multiclass scenarios. To further evaluate the performance of the proposed model over the baseline models, the t-statistic value between the two classifiers is computed.[30] Most of the t-values in Table 6 are positive, which indicates the superior performance of the proposed model over the standard machine learning-based models.

The table compares the t-statistic values of the Proposed Model against three baseline models (Model-1, Model-2, and Model-3) across different datasets to evaluate the significance of performance differences. For Dataset-1, the proposed approach shows noticeable improvements over Model-1 and Model-3, with t-statistic values of 1.2 and 1.4, respectively, indicating stronger performance. The difference with Model-2 is smaller, reflected by a t-statistic of 0.91, but it still suggests a positive impact. In the case of Dataset-2, the Proposed Model demonstrates its largest margin of improvement over Model-1 with a t-statistic of 1.6, highlighting a clear advantage. The gaps are less significant with Model-2 (0.85) and Model-3 (0.75), although the proposed method still performs better. When evaluating the Combined Dataset (which includes both Dataset-1 and Dataset-2), the t-statistic values drop slightly to 0.82, 0.74, and 0.61 against Model-1, Model-2, and Model-3, respectively, suggesting that while the Proposed Model remains effective, its edge is slightly reduced when data is merged. For Dataset-3, the Proposed Model still shows consistent improvements with t-statistic values of 0.91 compared to Model-1, 1.2 against Model-2, and 0.92 against Model-3, indicating reliable gains. Overall, the results highlight that the Proposed Model generally outperforms the baseline models, especially when datasets are evaluated individually. Its performance remains solid even in combined scenarios, suggesting robust learning capabilities across different data conditions. An ablation study has been carried out, and the results are listed in Table 5. Across all datasets, the combined feature set has been observed to outperform single-feature models consistently. Relative to the best single-feature baseline (cepstral), the fusion model improves accuracy by approximately 6 percentage points. This consistent gain demonstrates that spectral and cepstral representations provide complementary information, and their fusion enhances discriminative power. The ablation, therefore, supports the hypothesis that multi-feature fusion increases robustness and classification accuracy for insect acoustic detection in stored-grain environments.

Previous acoustic-based insect detection studies have primarily relied on single-domain features such as MFCCs or spectrograms with conventional machine learning classifiers or CNN-based approaches.[34] More recent works have begun to explore feature fusion; for example, McLoughlin *et al.* showed that combining spectrogram and cochleogram features improved audio event classification under noisy conditions.[35] However, these studies focus either on limited fusion strategies or controlled environments. In contrast, the proposed model works by integrating spectral, cepstral, and statistical features simultaneously, thereby capturing complementary signal information. Moreover, by incorporating a phase-based speech enhancement model, it explicitly addresses noise resilience, which is often overlooked in earlier insect detection studies.

Table 5 — Ablation study of the proposed model

	Spectrum features	Cepstral	Combined
Dataset-1	0.85	0.88	0.94
Dataset-2	0.86	0.88	0.95
Dataset-3	0.84	0.87	0.94

Table 6 — Comparison of t-statistic values

Dataset	Proposed vs Model-1	Proposed vs Model-2	Proposed vs Model-3
Dataset-1	1.2	0.91	1.4
Dataset-2	1.6	0.85	0.75
Combined Dataset	0.82	0.74	0.61
Dataset-3	0.91	1.2	0.92

This combination of multi-domain feature fusion with PBSE-based preprocessing differentiates our approach and is designed specifically to enhance robustness and accuracy in post-harvest pest management scenarios.

Conclusions

In the current study, a non-invasive method for the detection of Insects in stored grains is proposed using acoustic analysis of the insect sounds and a feature fusion technique. This model has been tested in multiple datasets, both with binary classification and multiclass classification scenarios. It has been observed that the proposed model consistently outperforms baseline models in both binary and multiclass scenarios in Accuracy, F1, Precision, Recall, and AUC, indicating its strong classification capabilities and enhanced reliability. The t-statistic analysis further supports the superiority of the proposed model. The overall research contributions are: enhanced classification, robustness across multiple datasets, statistical validation, and a scalable solution for real-world applications. These contributions highlight the potential of the proposed Model as a reliable and scalable solution for complex classification tasks, advancing the state of research in this domain.

Acknowledgements

- Availability of Data - The insect dataset can be accessed from the Data Hub IIT Kharagpur (https://www.datahubkgp.org)
- This work is partially supported by the IIT Kharagpur AI4ICPS I Hub Foundation under the core Project Grant No. PID AR-051-224. For accessing the rice insect dataset, please refer to the Data Hub, IIT Kharagpur.
- Ethics Declarations - The authors declare no competing interests.

References

1 Lima M C F, de Almeida Leandro M E D, Valero C, Coronel L C P & Bazzo C O G, Automatic detection and monitoring of insect pests—A review, *Agric*, **10(5)** (2020) 1–24, doi: 10.3390/agriculture10050161

2 Ung H T, Ung Q H, Nguyen T T & Nguyen B T, An efficient insect pest classification using multiple convolutional neural network based models, *SoMeT*, (2022) 584–595, doi: 10.48550/arXiv.2107.12189

3 Nguyen H Q, Truong T D, Nguyen X B, Dowling A, Li X & Luu K, Insect-foundation: a foundation model and large-scale 1M dataset for visual insect understanding, *Proc IEEE CVF Conf Comput Vis Pattern Recognit*, (2024) 21945–21955. https://openaccess.thecvf.com/content/CVPR2024/html/Nguyen_Insect-Foundation_A_Foundation_Model_and_Large-scale_1M_Dataset_for_Visual_CVPR_2024_paper.html

4 Kapetas D, Christakakis P, Faliagka S, Katsoulas N & Pechlivani E M, AI-driven insect detection, real-time monitoring, and population forecasting in greenhouses, *Agri Eng*, **7(2)** (2025) 29–29, doi: 10.3390/agriengineering7020029.

5 Ferrari V, Calvini R, Boom B, Menozzi C, Rangarajan A K, Maistrello L, Offermans P & Ulrici A, Evaluation of the potential of near infrared hyperspectral imaging for monitoring the invasive brown marmorated stink bug, *Chemometr Intell Lab Syst*, **234** (2023) 104751–104751, doi: 10.1016/j.chemolab.2023.104751.

6 Ali M A, Dhanaraj R K & Nayyar A, A high performance-oriented AI-enabled IoT based pest detection system using sound analytics in large agricultural field, *Microprocess Microsyst*, **103** (2023) 104946–104946, doi: 10.1016/j.micpro.2023.104946.

7 Mendoza Q A, Pordesimo L & Neilsen M, Enhancing grain facility management with AI-based insect detection and identification system, *Proc SPIE Sens Agric Food Qual Saf XV*, **12545** (2023) 149–156, doi: 10.1117/12.2672253.

8 Boulila W, Alzahem A, Koubaa A, Benjdira B & Ammar A, Early detection of red palm weevil infestations using deep learning classification of acoustic signals, *Comput Electron Agric*, **212** (2023) 108154–108154, doi: 10.1016/j.compag.2023.108154.

9 Mankin R, Hagstrum D, Guo M, Eliopoulos P & Njoroge A, Automated applications of acoustics for stored product insect detection, monitoring, and management, *Insects*, **12(3)** (2021) 259–259, doi: 10.3390/insects12030259.

10 Balingbing C, Kirchner S, Siebald H, Hung N V & Hensel O, Determining the sound signatures of insect pests in stored rice grain using an inexpensive acoustic system, *Food Secur*, **16(6)** (2024) 1529–1538, doi: 10.1007/s12571-024-01493-6.

11 Dhanaraj R K & Ali M A, Deep learning-enabled pest detection system using sound analytics in the internet of agricultural things, *Eng Proc*, **58(1)** (2023) 123, doi: 10.3390/ecsa-10-16205.

12 Branding J, von Hörsten D, Wegener J K, Böckmann E & Hartung E, Towards noise robust acoustic insect detection: from the lab to the greenhouse, *Künstl Intell*, **37(2)** (2023) 157–173, doi: 10.1007/s13218-023-00812-x.

13 Sekkate S, Khalil M, Adib A & Jebara S B, An investigation of a feature-level fusion for noisy speech emotion recognition, *Computers*, **8(4)** (2019) 91, doi: 10.3390/computers8040091.

14 Bahreini M, Barati R & Kamali A, Cardiac sound classification using a hybrid approach: MFCC-based feature fusion and CNN deep features, *EURASIP J Adv Signal Process*, **2025(2)** (2025) 2, doi: 10.1186/s13634-025-01203-0.

15 Lü Z, Shi Y, Lü L, Han D, Wang Z & Yu F, Dual-feature fusion learning: an acoustic signal recognition method for marine mammals, *Remote Sens*, **16(20)** (2024) 3823, doi: 10.3390/rs16203823.

16 Sueur J, Farina A, Gasc A, Pieretti N & Pavoine S, Acoustic indices for biodiversity assessment and landscape investigation, *Acta Acust United Ac*, **100(4)** (2014) 772–781, doi: 10.3813/AAA.918757.

17 Mankin R W, Hagstrum D W, Smith M T, Roda A L & Kairo M T, Perspective and promise: a century of insect

acoustic detection and monitoring, *Am Entomol*, **57(1)** (2011) 30–44, doi: 10.1093/ae/57.1.30.

18 Mankin R W & Moore A, Acoustic indicators for targeted detection of stored product and urban insect pests by inexpensive infrared, acoustic, and vibrational detection of movement, *J Econ Entomol*, **103(5)** (2010) 1636–1646, doi: 10.1603/EC10126.

19 Branding J, Hörsten D, Böckmann E, Wegener J K & Hartung E, InsectSound1000: a dataset for deep learning based acoustic insect recognition, *Sci Data*, **11** (2024) 475–475, doi: 10.1038/s41597-024-03301-4.

20 He H, Chen J, Chen H, Zeng B, Huang Y, Zhaopeng Y & Chen X, Enhancing insect sound classification using dual-tower network: a fusion of temporal and spectral feature perception, *Appl Sci*, **14** (2024) 3116, doi: 10.3390/app14073116.

21 Faiß M & Stowell D, InsectSet459: an open dataset of insect sounds for bioacoustic machine learning, *arXiv*, (2025) 2503.15074–2503.15074, doi: 10.48550/arXiv.2503.15074.

22 Faiß M & Stowell D, Adaptive representations of sound for automatic insect recognition, *PLoS Comput Biol*, **19(10)** (2023) e1011541, doi: 10.1371/journal.pcbi.1011541.

23 Ebrahimi M A, Khoshtaghaza M H, Minaei S & Jamshidi B, Vision-based pest detection based on SVM classification method, *Comput Electron Agric*, **137** (2017) 52–58, doi: 10.1016/j.compag.2017.03.016.

24 Dash T K, Insect detection in stored food grains using speech recognition and ML-based models, (2025). https://www.ai4icps.in/projects?section=ai_env

25 Faiß M, InsectSet32: dataset for automatic acoustic identification of insects (Orthoptera and Cicadidae), (2022). https://zenodo.org/records/7072196

26 Liu X, Zhang H, Jiang Q, Ren L, Chen Z, Luo Y & Li J, Acoustic denoising using artificial intelligence for wood-boring pests semanotus bifasciatus larvae early monitoring, *Sensors*, **22(10)** (2022) 3861, doi: 10.3390/s22103861.

27 Dash T K, Solanki S S & Panda G, Improved phase aware speech enhancement using bioinspired and ANN techniques, *Analog Integr Circuits Signal Process*, **102(3)** (2020) 465–477, doi: 10.1007/s10470-019-01566-z.

28 Dash T K, Chakraborty C, Mahapatra S & Panda G, Gradient boosting machine and efficient combination of features for speech-based detection of COVID-19, *IEEE J Biomed Health Inform*, **26(11)** (2022) 5364–5371, doi: 10.1109/JBHI.2022.3197910.

29 Matlab, Audio feature extractor function, (2020). https://it.mathworks.com/help/audio/ref/audiofeatureextractor.html/(2020).

30 Wong T T, Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets, *Pattern Recognit*, **65** (2017) 97–107, doi: 10.1016/j.patcog.2016.12.018.

31 Pao T L, Chen Y T, Yeh J H, Cheng Y M & Lin Y Y, A comparative study of different weighting schemes on KNN-based emotion recognition in Mandarin speech, *Int Conf Intell Comput*, (2007) 997–1005, doi: 10.1007/978-3-540-74171-8_101.

32 Lu L, Ge F, Zhao Q & Yan Y, A SVM-based audio event detection system, *Int Conf Electr Control Eng*, (2010) 292–295, doi: 10.1109/iCECE.2010.78.

33 Rong F, Audio classification method based on machine learning, *Int Conf Intell Transp Big Data Smart City*, (2016) 81–84, doi: 10.1109/ICITBS.2016.98.

34 Noda J J, Travieso C M, Sánchez-Rodríguez D, Dutta M K & Singh A, Using bioacoustic signals and support vector machine for automatic classification of insects, *Int Conf Signal Process Integr Netw*, (2016) 656–659, doi: 10.1109/SPIN.2016.7566778.

35 McLoughlin I, Zhang H, Xie Z, Song Y & Xiao W, Robust sound event classification using deep neural networks, *IEEE ACM Trans Audio Speech Lang Process*, **23(3)** (2015) 540–552, doi: 10.1109/TASLP.2015.2389618.