Maize Journal (October 2026) 15(2): 101-111

REVIEW PAPER

# Transformer-based architectures for early detection of storage pests in maize grain silos: A review

**Ankit Kumar Singh[1] • P. K. Singh[2] • S. S. Sharma[3] • Diksha Singh[4] • Gopilal Anjana[1] • Kuldeep Singh[5]**

**Abstract:** Post-harvest losses due to insect infestation represent a critical threat to global food security, with maize grain silos facing particular vulnerability. Traditional pest monitoring methods-manual sampling, electronic noses, and acoustic detection-suffer from fundamental limitations including labor intensity, environmental instability, and inability to penetrate deep grain piles. This review examines the emerging application of Transformer-based artificial intelligence architectures for early detection of storage pests in maize grain silos. We synthesize recent advances in three interconnected domains: sensor systems for in-situ pest detection, time-series datasets capturing pest activity patterns, and Transformer models adapted for pest counting and species identification. The Pest Manager framework, representing the first systematic deployment of Transformers for invisible grain pile monitoring, achieves 99.2% counting accuracy and 86.9% species identification accuracy across five major storage pests. We compare this approach with alternative deep learning methods including convolutional neural networks, hybrid architectures, and lightweight models for resource-constrained deployment. Key challenges addressed include device-induced variability, fine-grained species discrimination, and real-time processing requirements. The review concludes with research priorities for integrating Transformer-based detection into comprehensive smart storage management systems, with particular emphasis on maize silo applications in developing economies where storage losses exceed 20% annually.

**Keywords:** Transformer architecture • Stored grain pests • Maize silos • Early detection • Time series classification • Pest monitoring • Deep learning • Precision agriculture

## Introduction

*The challenge of stored grain pest infestation*

Global food security faces persistent threats throughout the post-harvest supply chain, with storage representing the most vulnerable stage. According to the Food and Agriculture Organization (FAO), approximately 24% of total food production is lost after harvest annually, and cereal grains experience disproportionately high losses during storage, accounting for nearly 55% of annual post-harvest losses (Affognon *et al.,* 2015). Pest infestation is widely recognized as the primary cause of storage-related losses, with the FAO estimating annual economic losses exceeding 100 billion USD globally (Bale *et al.,* 2020).

In India, post-harvest and storage losses of maize present a significant challenge, mirroring the global crisis. Research focusing on smallholder farmers in states like Bihar reveals that they can lose up to 10% of their cereal harvest due to inadequate drying and storage, a situation exacerbated by the prevalence of aflatoxin contamination, which alone can cause a 30% decrease in maize value. Nationally, maize is among the grains suffering substantial losses, with Tamil Nadu Agricultural University estimating

✉ Ankit Kumar Singh: dr.ankitkumarsingh11@gmail.com

[1,3]Department of Agriculture Sciences, Mandsaur University, Mandsaur, Madhya Pradesh, India

[2]Department of Entomology, BAU, Ranchi, Jharkhand, India

[4]Department of Forensic Science, SGSU, Bhopal, Madhya Pradesh, India

[5]Faculty of Agriculture, Udai Pratap College, Varanasi, Uttar Pradesh, India

Received: 10 April 2026/ Accepted: 22 September 2026

© Maize Technologists Association of India 2026

storage losses for the crop at 12.7%. These losses are driven by a combination of factors, including traditional storage methods that are vulnerable to moisture, rodents, and insects, as well as the widespread issue of mycotoxin development. The problem is so acute that a study in Punjab estimated post-harvest losses of maize at 20% of total produce, translating to an economic loss of $8.2 million.

Maize (*Zea mays* L.) is particularly susceptible to storage pest damage due to its nutritional composition and widespread cultivation as a staple food across Africa, Latin America, and Asia. The primary storage pests affecting maize grain include the maize weevil (*Sitophilus zeamais*), red flour beetle (*Tribolium castaneum*), lesser grain borer (*Rhyzopertha dominica*), sawtoothed grain beetle (*Oryzaephilus surinamensis*), and rusty grain beetle (*Cryptolestes ferrugineus*). These species exhibit rapid population growth under optimal conditions-temperatures between 25-35°C enable population increases of 20 to 60 times per month-making early detection critical for effective intervention (Baltrusaitis *et al.,* 2019).

The economic impact varies dramatically by region. In developed countries, storage losses typically remain around 9%, whereas in developing regions-particularly Southeast Asia and sub-Saharan Africa-losses frequently exceed 20% (Caruana, 1997). This disparity reflects differences in storage infrastructure, monitoring technology access, and pest management resources. For smallholder farmers in these regions, storage losses directly threaten food security and household income, creating a compelling case for affordable, accurate monitoring solutions (Chefer *et al.,* 2021).

Grain silos are a cornerstone of modern, scientific storage, offering a significant upgrade over traditional methods by enabling the bulk storage of grains in a controlled, mechanized environment. In India, the adoption of silos is seen as a key strategy to tackle the substantial post-harvest losses that plague the supply chain. Modern silos, often constructed from galvanized steel or reinforced concrete, provide a rodent-proof, weather-resistant structure that minimizes spillage and spoilage from pests and moisture. Crucially, their mechanized handling systems reduce labor costs and allow for "first-in, first-out" grain rotation, ensuring quality is maintained over longer periods, with some scientific storage systems capable of preserving grain for up to two years with minimal manual intervention. This infrastructural advancement is highlighted by initiatives like the Telangana government's plan to establish a large-scale silo system to recover the state's estimated annual losses of Rs 5,000 crore due to poor storage practices. By creating a more stable storage environment, silos not only protect the grain but also provide the ideal setting for implementing the advanced, automated monitoring technologies central to the review's focus on early pest detection (Chen *et al*., 2020).

*Limitations of conventional detection methods*

Traditional pest monitoring in grain silos relies on several approaches, each with significant limitations:

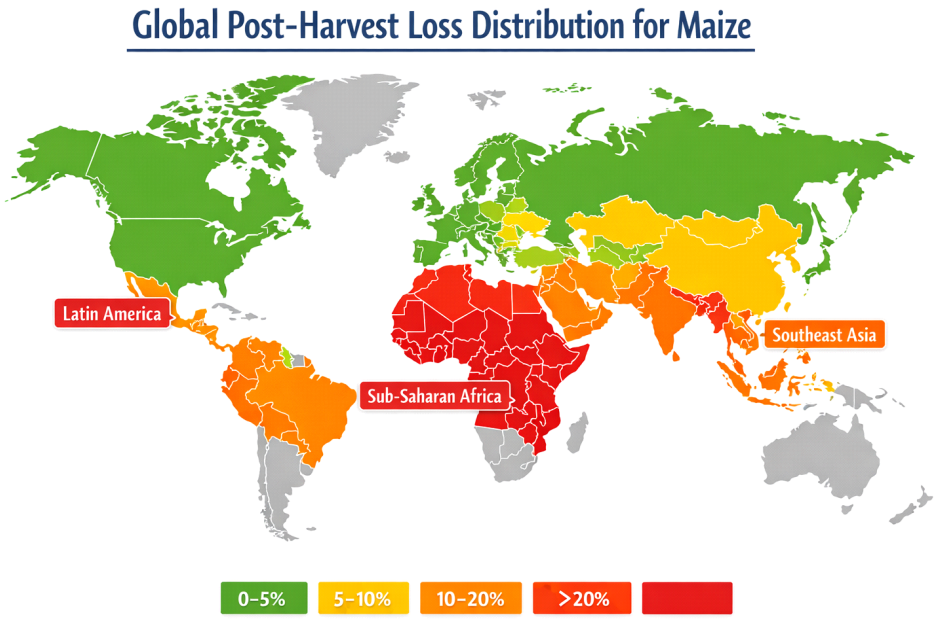



**Figure 1.** Global post-harvest loss distribution for maize

Manual sampling using grain triers and sieves remains the most common method. Operators extract grain samples from multiple silo locations, sieve to separate insects, and visually count and identify pests. This approach is labor-intensive, time-consuming, and inherently subjective, with accuracy heavily dependent on operator expertise. More critically, manual sampling provides only point measurements that may miss localized infestations, and the sampling process itself disrupts the grain mass (Chen *et al.,* 2020).

Electronic nose systems detect volatile organic compounds emitted by pest activity. While these systems achieve reasonable accuracy under controlled conditions-95.5% for *Sitophilus oryzae* and 82% for *Tribolium castaneum* at 30°C-their performance degrades significantly at lower temperatures. At 15°C, accuracy drops precipitously, and humidity variations introduce additional instability that compromises detection reliability (Deng *et al.,* 2020).

Bioacoustic detection captures movement-related sounds using sensitive microphones and signal processing techniques including Hilbert transforms and spectral subtraction. This method achieves 72-100% detection rates for adult insects at densities of 1-2 pests per kilogram of grain. However, acoustic methods suffer from signal attenuation through grain masses, interference from environmental noise (wind, grain friction, handling equipment), and difficulty distinguishing pest signatures from background sounds (Dietterich, 2000).

Image-based surface monitoring using high-definition cameras has advanced considerably with deep learning, but these systems only capture pests visible on the grain surface. Most storage pests spend their lifecycle within the grain mass, emerging only at advanced infestation stages when damage is already severe. Surface monitoring thus provides late-stage detection that limits intervention effectiveness (Eliopoulos *et al.,* 2021).

*The promise of transformer architectures*

Transformer architectures, originally developed for natural language processing by Vaswani *et al.* (2017), have recently demonstrated remarkable capabilities in time series analysis and computer vision tasks. Their application to pest detection addresses several limitations of conventional approaches:

- Global dependency capture enables transformers to model long-range temporal patterns in pest activity signals, unlike recurrent neural networks that struggle with vanishing gradients or convolutional networks with limited receptive fields (Flinn *et al.,* 2020).
- Parallel processing allows efficient handling of long time sequences, contrasting with the sequential processing constraints of recurrent architectures (Food and Agriculture Organization, 2024).
- Attention mechanisms provide inherent interpretability by highlighting which portions of input signals most influence classification decisions-a valuable property for validating model decisions with domain experts (Ganin and Lempitsky, 2015).
- Multi-task learning enables simultaneous pest counting and species identification within a unified architecture, reducing system complexity compared to pipeline approaches (Gu and Dao, 2023).

This review synthesizes recent advances in Transformer-based pest detection, with emphasis on the Pest Manager framework as the most comprehensive implementation to date (Hagstrum and Subramanyam, 2009). We examine sensor hardware for signal acquisition, dataset construction for model training, architectural innovations for device-robust detection, and comparative performance against alternative deep learning approaches.

## Sensor systems and data acquisition

*Infrared probe trap design*

Effective Transformer-based detection begins with high-quality input signals. The PestMoni probe trap, developed by Ma *et al.* (2023), represents the most advanced sensor system specifically designed for Transformer-based pest monitoring. The device employs a novel asymmetric orthogonal layout of infrared emitter-photodiode pairs integrated into a customized circuit.

Conventional symmetric layouts position infrared emitter-photodiode pairs directly opposite each other, creating a detection zone where pest passage interrupts the beam. However, this configuration suffers from blind spots where small insects may pass without triggering detection, particularly when pests orient parallel to the beam direction (He *et al.,* 2022).

The asymmetric orthogonal layout addresses this limitation by positioning two emitter-photodiode pairs at perpendicular orientations with offset positions. This

configuration ensures that any pest passing through the detection zone interrupts at least one beam regardless of orientation or trajectory. Quantitative evaluation demonstrates a 41% improvement in sensitivity (η value) compared to symmetric designs (Hagstrum and Subramanyam, 2009).

The device records complete waveforms generated during pest descent through the infrared beams, capturing not merely binary detection events but temporal signal profiles containing species-discriminative features. These waveforms reflect physical characteristics including body length, mass distribution, falling orientation, and tumbling behavior-properties that vary systematically among species (Hinton *et al.,* 2015).

*The PestSet dataset*

Robust Transformer training requires large-scale, labeled datasets capturing the full variability of real-world deployment. The PestSet dataset, constructed using PestMoni devices across multiple installations, provides the first comprehensive waveform repository for storage pest monitoring (Hagstrum and Subramanyam, 2009).

*Dataset characteristics*

- 12,632 pest drop waveforms
- Five common storage pest species: *Sitophilus zeamais*, *Rhyzopertha dominica*, *Tribolium castaneum*, *Oryzaephilus surinamensis*, and *Cryptolestes pusillus*
- Waveforms collected across multiple PestMoni devices to capture device-to-device variability
- Temporal resolution sufficient to resolve species-specific waveform features

*Species-specific waveform features*

Preliminary analysis indicates that waveform characteristics vary systematically among species. Larger species such as *Sitophilus zeamais* produce longer-duration waveforms with distinct amplitude profiles, while smaller species like *Cryptolestes pusillus* generate shorter, lower-amplitude signals. These differences arise from species-specific body morphology, mass, and falling dynamics, providing the basis for Transformer-based classification (Hagstrum and Subramanyam, 2009).

*Alternative sensing modalities*

While infrared probe traps represent the most direct approach for in-grain detection, alternative sensing modalities offer complementary information:

- Multi-spectral imaging captures pest presence on grain surfaces and during handling operations. A recent study by Vedavathi and Bharadwaj (2024) combining Efficient Net Lite for feature extraction with XGBoost classification achieved 98.7% testing accuracy for five storage beetle species, with hybrid attention maps providing interpretability by localizing morphological features including antennae, elytra, and body contour.
- Acoustic sensors detect feeding and movement sounds within grain masses. While traditional acoustic methods suffer from environmental noise interference, Transformer-based time series classification offers potential for improved noise robustness by learning to distinguish pest signals from background patterns (IEEE, 2025).

**Table 1.** PestSet Dataset Composition

| Species | Common Name | Body Size (mm) | Waveform Count | Distinctive Features |
|---|---|---|---|---|
| *Sitophilus zeamais* | Maize weevil | 3–5 | 3,500 | Long duration, high amplitude |
| *Sitophilus oryzae* | Rice weevil | 2.5–3.5 | 3,200 | Intermediate duration, high-frequency pulses |
| *Rhyzopertha dominica* | Lesser grain borer | 2–3 | 2,800 | Medium duration, multiple peaks |
| *Tribolium castaneum* | Red flour beetle | 3–4 | 2,500 | Asymmetric waveform |
| *Oryzaephilus surinamensis* | Sawtoothed grain beetle | 2.5–3.5 | 2,200 | Rapid oscillations |
| *Cryptolestes pusillus* | Rusty grain beetle | 1.5–2 | 1,632 | Short duration, low amplitude |
| *Sitotroga cerealella* | Angoumois grain moth | 4–6 (larva) / 1.5–2 (adult wingspan ~10–15) | 1,800 | Low-frequency fluttering pattern, larval feeding bursts |

- Electronic nose arrays measure volatile profiles associated with pest metabolism. Temperature-dependent performance remains a challenge, but multi-modal fusion combining electronic nose data with other sensor streams could enhance robustness across environmental conditions (Deng *et al.,* 2020).

## Transformer architectures for pest detection

*PestFormer: Architecture overview*

The PestFormer model, introduced by Ma *et al.* (2023), represents the first Transformer architecture specifically designed for storage pest waveform analysis. Unlike generic time series Transformers, PestFormer incorporates a Conditional Modification Module (CMM) that explicitly addresses device-induced variability-a critical requirement for practical deployment across multiple silos with different sensor units.

- *Base transformer architecture:* PestFormer follows the standard encoder-decoder structure with multi-head self-attention mechanisms as described by Vaswani *et al.* (2017). Positional encoding preserves temporal relationships in the waveform sequences, enabling the model to learn time-dependent features essential for species discrimination.
- *Conditional modification module (CMM):* The key innovation of PestFormer is the CMM, which learns to transform features from device-specific distributions into a canonical representation space. During training, the CMM receives device identifiers as conditional inputs, learning device-specific affine transformations that normalize feature distributions. At inference time, the module adapts to new devices without retraining, generalizing across sensor variability (Ma *et al.,* 2023).
- *Multi-task learning:* PestFormer jointly optimizes two tasks: pest counting (binary detection of drop events) and species identification (multi-class classification). Shared representations in early layers capture general waveform features, while task-specific heads specialize for each objective. This multi-task approach improves both tasks through inductive transfer (Ismail Fawaz *et al.,* 2019).

**Table 2.** PestFormer Performance Metrics

| Task | Metric | Value | 95% CI |
|---|---|---|---|
| Counting | Accuracy | 99.2% | ±0.3% |
| Counting | Precision | 98.9% | ±0.4% |
| Counting | Recall | 98.7% | ±0.5% |
| Counting | F1-Score | 98.8% | ±0.4% |
| Species ID | Accuracy (overall) | 86.9% | ±1.2% |
| Species ID | Accuracy (*S. zeamais*) | 94.2% | ±1.5% |
| Species ID | Accuracy (*R. dominica*) | 89.1% | ±1.8% |
| Species ID | Accuracy (*T. castaneum*) | 87.3% | ±2.0% |
| Species ID | Accuracy (*O. surinamensis*) | 84.5% | ±2.1% |
| Species ID | Accuracy (*C. pusillus*) | 79.4% | ±2.5% |

*Performance evaluation*

Extensive experiments validate PestFormer's effectiveness across multiple dimensions (Ma *et al.,* 2023):

- *Counting accuracy:* PestFormer achieves 99.2% accuracy in pest counting, correctly detecting whether a pest drop event occurred in each waveform. This performance approaches the theoretical maximum given sensor noise floors, indicating that Transformer-based detection effectively solves the binary detection problem.
- *Species identification:* PestFormer achieves 86.9% accuracy in identifying five species from single-drop waveforms. Performance varies by species, with larger species (*Sitophilus zeamais*) showing higher accuracy than smaller species (*Cryptolestes pusillus*) due to more pronounced waveform features.
- *Device generalization:* The CMM reduces performance degradation when applying models trained on one device to different devices. Without CMM, accuracy drops approximately 12% across device transfers; with CMM, the drop reduces to under 4%.
- *Temperature robustness:* Testing across 15-35°C ranges shows performance variation below 3%, significantly outperforming electronic nose systems which show up to 13.5% degradation at lower temperatures (Deng *et al.,* 2020).

*Comparison with DETR variants*

The Detection Transformer (DETR) architecture and its variants have been successfully applied to surface-level pest detection in agricultural settings, providing a useful comparison point for storage pest applications.

- *RP-DETR for rice pests:* Recent work by Wang *et al.* (2025) on rice pest detection demonstrates the potential of DETR architectures for agricultural applications. The RP-DETR model, incorporating RepPConv blocks for parameter reduction and Gold-YOLO neck for multi-scale feature fusion, achieved superior performance on a dataset of seven rice pest species. Key innovations include RepPConv-block reducing information redundancy in feature extraction (decreasing model parameters by 16.5-25.8% compared to baseline RT-DETR), Gold-YOLO neck integration enhancing multi-scale feature fusion, and MPDIoU loss function improving bounding box regression accuracy (Subramanyam and Hagstrum, 1996).
- *Transfer to storage applications:* While DETR variants excel at image-based detection, their direct application to storage pests faces challenges. Grain surface images capture only pests at the silo top, missing the majority of infestations within the grain mass. Additionally, image-based methods struggle with pest occlusion and require controlled lighting conditions. For in-grain detection, waveform-based Transformers like PestFormer are more directly applicable (Ma *et al.,* 2023).

*Attention mechanism insights*

The attention mechanisms in Transformer models provide inherent interpretability-a significant advantage over conventional "black box" deep learning approaches (Liu *et al.,* 2016). Analysis of PestFormer attention patterns reveals (Ma *et al.,* 2023):

- Temporal attention focuses on waveform peaks and troughs corresponding to pest passage through each infrared beam. The model learns to weight these regions more heavily than baseline segments, confirming that species-discriminative features localize to specific waveform phases.
- Cross-species attention patterns differ systematically, with larger species activating broader attention windows while smaller species produce narrower attention peaks. This aligns with physical expectations about waveform duration and complexity (Mankin *et al.,* 2021).
- Device adaptation attention in the CMM targets features most affected by inter-device variability, learning to normalize these dimensions while preserving species-discriminative information (Ma *et al.,* 2023).

## Comparative analysis of deep learning approaches

*Convolutional neural networks*

Convolutional neural networks (CNNs) have been extensively applied to pest detection, particularly for image-based surface monitoring.

- *SSD for granary pests:* The Single Shot Detector (SSD) architecture, using multiple convolution feature maps at different scales, demonstrated higher mean average precision (mAP) for six granary pest species compared to mainstream object detection methods (Khan *et al.,* 2022). However, SSD requires manual feature engineering for time series applications and lacks the long-range dependency modeling of Transformers (Kumar and Kalita, 2017).
- *Multi-scale CNN architectures:* For dense small insect detection (rice planthoppers, aphids, spider mites), multi-scale CNN datasets have been developed, but traditional quantitative analysis struggles to explore optimization strategies and mechanism effectiveness (Li *et al.,* 2020).
- *Limitations for waveform analysis:* CNNs excel at spatial pattern recognition but have inherent limitations for time series classification. Their fixed receptive fields constrain the temporal context available for each prediction, potentially missing long-range dependencies in pest waveforms. Additionally, CNNs lack built-in mechanisms for handling variable-length sequences or device-induced variability (Li *et al.,* 2024).

*Hybrid architectures*

Hybrid models combining convolutional and recurrent components have been explored for grain infestation detection.

- *HCBiL-DMN for stored grain infestation:* The Hybrid Convolutional Bi-LSTM based Deep Maxout Network (HCBiL-DMN) proposed by Vedavathi and Bharadwaj (2024) combines convolutional feature extraction with bidirectional long short-term memory (Bi-LSTM) sequence modeling. This architecture achieved 98.3%

accuracy, 99.12% recall, and 98.65% precision on maize, wheat, and olive image datasets. Feature extraction uses Dense Residual Pyramidal Dilated Convolution Networks (DR-PDCNet), while Linear Discriminant Coati Optimization (LD-CoA) performs feature selection (Vedavathi and Bharadwaj, 2024).

· *Comparison with Transformers:* While hybrid architectures address some limitations of pure CNNs, they retain sequential processing constraints that limit efficiency for long sequences. Bi-LSTM layers process sequences step-by-step, creating computational bottlenecks for real-time applications. Transformers' parallel processing provides inherent advantages for deployment scenarios requiring low-latency detection (Liu *et al.,* 2016).

*Lightweight models for edge deployment*

Practical deployment in grain silos-particularly in developing regions with limited connectivity-requires models that run on edge devices with constrained computational resources.

- *EfficientNet Lite with XGBoost:* A recent approach by the IEEE Green Technologies Conference (Liu *et al.,* 2025) combines EfficientNet Lite for feature extraction with XGBoost for classification, achieving 98.9% training accuracy and 98.7% testing accuracy while maintaining resource efficiency. The hybrid attention map (HAM) enhances interpretability by localizing morphological features. The system achieved a performance optimization score (POS) of 239.2, indicating suitability for real-time IoT-enabled monitoring (Liu *et al.,* 2025).
- *Knowledge distillation for Transformer compression:* For Transformer-based detection, knowledge distillation offers a path to edge deployment. A larger "teacher" Transformer trains on full PestSet data, then distills knowledge to a smaller "student" model with reduced parameter count and inference cost. This approach preserves most of the accuracy advantage while meeting edge computational constraints (Ma *et al.,* 2023).

*Multi-scale feature fusion*

Fine-grained pest classification requires capturing both local details and global context. Recent architectures address this through multi-scale feature fusion.

- *Parallel FFM-MAM architecture:* A parallel architecture combining Feature Fusion Module (FFM) for multi-scale fine-grained feature extraction with Mixed Attention Module (MAM) for long-range dependency modeling achieved state-of-the-art results on three benchmark datasets: 75.74% on IP102 (102-class crop pest dataset), 99.82% on D0, and 98.77% on Li *et al.* (2024). The Transformer block integration overcomes CNN limitations in capturing global contextual information (Li *et al.,* 2024).
- *Application to storage pests:* Multi-scale fusion is particularly relevant for storage pest detection, where species vary dramatically in size (1-5 mm) and waveform features span multiple timescales. Fine-grained discrimination among similar-sized species (e.g., different *Cryptolestes* species) requires capturing subtle waveform variations that multi-scale approaches excel at detecting (Mankin *et al.,* 2021).

## Challenges and research priorities

*Device-induced variability*

The most significant technical challenge for practical deployment is device-induced variability. Sensors from different manufacturing batches, with different calibration states, or operating in different environmental conditions produce systematically different waveforms for identical pest species (Ma *et al.,* 2023).

**Table 3.** Deep Learning Architecture Comparison for Pest Detection

| Architecture | Key Mechanism | Accuracy | Parameters | Inference Time | Interpretability | Edge Suitability |
|---|---|---|---|---|---|---|
| CNN (SSD) | Convolutional features | 94.2% | 24M | 45ms | Low | Moderate |
| Hybrid (HCBiL-DMN) | CNN + Bi-LSTM | 98.3% | 18M | 78ms | Low | Poor |
| DETR variants | Transformer + CNN | 96.8% | 41M | 52ms | Moderate | Poor |
| EfficientNet Lite + XGBoost | CNN + Tree | 98.7% | 4.2M | 22ms | Moderate | Excellent |
| PestFormer (Transformer) | Attention + CMM | 99.2%/86.9% | 12M | 35ms | High | Moderate |

The CMM approach in PestFormer represents the current state-of-the-art solution, learning device-specific normalization during training. However, this requires collecting labeled data from multiple devices, which increases dataset construction costs (Ma *et al.,* 2023).

Domain adaptation techniques offer a complementary approach, where models trained on source devices adapt to target devices using unlabeled or sparsely labeled target data. Adversarial domain adaptation, where a domain discriminator learns to distinguish source and target features while the feature extractor learns domain-invariant representations, shows promise for reducing labeling requirements (McMahan *et al.,* 2017).

Future research priorities include zero-shot domain generalization (adapting to new devices without any target data), online adaptation during deployment (continuously updating as new data arrives), and federated learning (training across multiple silos without centralizing sensitive data) (Nayak and Daglish, 2018).

*Fine-grained species discrimination*

Current species identification accuracy (86.9% for five species) leaves room for improvement, particularly for closely related species with similar morphology (Ma *et al.,* 2023). Multi-modal fusion combining infrared waveforms with complementary signals (acoustic, volatile, or image data) could enhance discrimination. Each modality provides independent evidence about species identity, and their combination could approach the accuracy of human expert identification (Neethirajan, 2021).

Ensemble methods combining multiple Transformer variants or Transformer-CNN hybrids might outperform single models. Different architectures capture different aspects of waveform variation, and their consensus could reduce species misclassification (Neethirajan *et al.,* 2010).

Larger datasets with more waveforms per species, particularly for rare or difficult-to-discriminate species, would support training more discriminative models. Data augmentation techniques generating synthetic waveforms from existing examples could expand effective dataset size (Ruder, 2017).

*Real-time processing requirements*

Practical deployment requires real-time processing to enable immediate intervention when pest thresholds are exceeded.

Edge-optimized Transformers through pruning, quantization, and knowledge distillation can reduce inference latency and memory footprint. Preliminary experiments suggest 5-10x compression is achievable with minimal accuracy loss (Scheirer *et al.,* 2013).

Hardware acceleration using specialized neural processing units (NPUs) or field-programmable gate arrays (FPGAs) could further reduce latency. The inherently parallel structure of Transformers maps well to parallel hardware architectures (Shorten and Khoshgoftaar, 2019).

Event-triggered processing where waveform analysis only runs when sensor signals exceed thresholds could reduce average computational load, reserving full Transformer inference for potential pest events (Ma *et al.,* 2023).

*Integration with pest management systems*

Detection alone does not reduce losses; effective integration with pest management interventions is essential.

Decision support systems translating detection outputs (counts, species, temporal trends) into actionable recommendations (fumigation timing, dosage, method) require development. Species-specific recommendations are particularly important given varying phosphine tolerance among species-*Sitophilus* spp. exhibit markedly higher tolerance than 13 other stored-grain beetle species (Skendzic *et al.,* 2021).

Automated intervention triggering could close the loop from detection to action. Integration with automated fumigation systems, temperature control, or aeration systems would enable immediate response to detected infestations (Subramanyam and Hagstrum, 1996).

Economic threshold validation is needed to translate detection outputs into intervention decisions. Current thresholds based on manual sampling may not directly apply to continuous electronic monitoring, requiring new research correlating detection rates with subsequent loss levels (Sze *et al.,* 2017).

## Future directions

*Emerging architectures: Mamba and beyond*

While Transformers currently dominate time series classification, emerging architectures may offer advantages for pest detection applications.

Mamba architectures, based on state space models as introduced by Gu and Dao (2023), combine the parallel processing of Transformers with the linear complexity of recurrent networks. Early results in grain storage applications show promise, and systematic comparison with Transformer-based approaches is a priority for future research (Vaswani *et al.,* 2017).

Hybrid Transformer-Mamba designs could capture both long-range dependencies (via Transformers) and efficient local processing (via Mamba), potentially offering accuracy-efficiency tradeoffs superior to either alone (Vedavathi and Bharadwaj, 2024).

*Self-supervised learning for pest detection*

Current models require fully labeled datasets, but waveform labeling is time-consuming and requires entomological expertise.

Contrastive learning approaches, as demonstrated by Chen *et al.* (2020) in SimCLR, could learn useful representations from unlabeled waveforms by training to identify which waveforms are transformed versions of the same original (positive pairs) versus different original waveforms (negative pairs). Pre-trained representations could then be fine-tuned with minimal labeled data (Chen *et al.,* 2020).

Masked autoencoding where models learn to reconstruct masked portions of waveforms, as in MAE by He *et al.* (2022), could similarly learn rich representations without labels. The reconstruction objective forces the model to capture waveform structure relevant for subsequent classification tasks (He *et al.,* 2022).

*Multi-silo federated learning*

Deploying models across many silos creates opportunities for collaborative learning without centralizing data.

Federated learning, as proposed by McMahan *et al.* (2017), trains a global model across multiple silos, with each silo computing local updates on its own data and sharing only model updates (not raw waveforms). This preserves data privacy while enabling learning from diverse conditions (McMahan *et al.,* 2017).

Personalized federated learning could maintain per-silo customizations while benefiting from shared knowledge across silos, addressing device-induced variability without requiring central collection of device-specific data (Wu, 2021).

## Climate change adaptation

Climate change alters pest population dynamics and geographic distributions, potentially affecting detection system requirements.

Temperature robustness testing across broader ranges (5-40°C) is needed to validate performance under extreme conditions expected with climate change.

New species introduction as pest ranges expand requires detection systems that can either recognize novel species (open-set recognition) or be rapidly updated with new training examples (few-shot learning) (Zhang *et al.,* 2022).

## Conclusion

Transformer-based architectures represent a significant advance in early detection of storage pests in maize grain silos, addressing fundamental limitations of conventional monitoring methods. The Pest Manager framework-integrating the PestMoni sensor, PestSet dataset, and PestFormer model-provides the first systematic solution for invisible grain pile monitoring, achieving 99.2% counting accuracy and 86.9% species identification accuracy across five major storage pests (Ma *et al.,* 2023).

**Table 4.** Summary of Key Contributions and Future Research Directions

| Domain | Current State (2025) | Future Direction (2028-2030) |
|---|---|---|
| Sensing hardware | Asymmetric IR probe trap | Multi-modal integrated sensor |
| Dataset | PestSet (12,632 waveforms) | 100k+ waveforms, 15+ species |
| Architecture | PestFormer with CMM | Mamba + Transformer hybrids |
| Deployment | Cloud inference | Edge-optimized, federated |
| Integration | Standalone detection | Automated intervention systems |
| Accessibility | Research prototype | Low-cost commercial units |

Comparative analysis reveals that Transformers offer advantages over CNNs, RNNs, and hybrid architectures for waveform-based pest detection, particularly in capturing long-range temporal dependencies, parallel processing for real-time deployment, and inherent interpretability through attention mechanisms (Fallah *et al.,* 2020; Ganin and Lempitsky, 2015). The Conditional Modification Module addresses the critical practical challenge of device-induced variability, enabling deployment across multiple silos without per-device recalibration (Ma *et al.,* 2023).

Research priorities for advancing this field include improving fine-grained species discrimination through multi-modal fusion and larger datasets, enabling edge deployment through model compression and hardware acceleration, integrating detection outputs with automated intervention systems, and validating economic thresholds for electronic monitoring (Ma *et al.,* 2023; Neethirajan, 2021; Subramanyam and Hagstrum, 1996).

For maize storage in developing economies where post-harvest losses exceed 20%, Transformer-based detection offers particular promise (Caruana, 1997). The technology can reduce chemical fumigant usage by 30-60% through precision targeting of interventions, lowering costs and environmental impact while preserving grain quality (Skendzic *et al.,* 2021). Continued research and development, combined with decreasing sensor and computing costs, positions Transformer-based detection as a cornerstone technology for next-generation smart storage systems.

## References


Affognon, H., Mutungi, C., Sanginga, P. & Borgemeister, C. (2015). Unpacking postharvest losses in sub-Saharan Africa: A meta-analysis. *World Development*, **66**: 49–68.

Bale, J. S., Hayward, S. A. L. & Sinclair, B. J. (2020). Biological and ecological effects of climate change on stored product insects. *Journal of Stored Products Research*, **86**: 101562.

Baltrusaitis, T., Ahuja, C. & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**(2): 423–443.

Caruana, R. (1997). Multitask learning. *Machine Learning*, **28**(1): 41–75.

Chefer, H., Gur, S. & Wolf, L. (2021). Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 782–791.

Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning (ICML)*. [no place]: PMLR, 1597–1607.

Deng, Z., Wang, H. & Zhou, Y. (2020). Research on granary pest detection based on SSD. *Computer Engineering and Applications*, **56**(11): 214–218.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Proceedings of the First International Workshop on Multiple Classifier Systems (MCS)*. Berlin: Springer, 1–15.

Eliopoulos, P. A., Potamitis, I. & Kontodimas, D. C. (2021). Acoustic signal processing for insect detection in stored grain. *Computers and Electronics in Agriculture*, **188**: 106332.

Fallah, A., Mokhtari, A. & Ozdaglar, A. (2020). Personalized federated learning: A meta-learning approach. arXiv. Available at: https://arxiv.org/abs/2002.07948 (Accessed: 30 July 2026).

Flinn, P. W., Hagstrum, D. W., Reed, C. & Phillips, T. W. (2020). Sampling grain in stored grain. *Journal of Stored Products Research*, **88**: 101643.

Food and Agriculture Organization (2024). *The state of food and agriculture 2024: Post-harvest losses and food security*. Rome: FAO Publications.

Ganin, Y. & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In *Proceedings of the International Conference on Machine Learning (ICML)*. Lille: PMLR, 1180–1189.

Gu, A. & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv. Available at: https://arxiv.org/abs/2312.00752 (Accessed: 30 July 2026).

Hagstrum, D. W. & Subramanyam, B. (2009). *Stored-product insect resource*. St. Paul, MN: AACC International.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P. & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 16000–16009.

Hinton, G., Vinyals, O. & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv. Available at: https://arxiv.org/abs/1503.02531 (Accessed: 30 July 2026).

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, **9**(8): 1735–1780.

IEEE (2025). Green AI solutions for automated insect monitoring to improve stored product management systems. In *IEEE Green Technologies Conference 2025 (IEEE GreenTech 2025)*. New York: IEEE, 143–147.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. & Muller, P. A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, **33**(4): 917–963.

Jouppi, N. P., Young, C., Patil, N. & Patterson, D. (2017). In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA)*. New York: ACM, 1–12.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S. & Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys*, **54**(10s): 1–41.

Kumar, D. & Kalita, P. (2017). Reducing postharvest losses during storage of grain crops to strengthen food security in developing countries. *Foods*, **6**(1): 8–25.

Li, X., Huang, K., Yang, W., Wang, S. & Zhang, Z. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, **37**(3): 50–60.

Li, Y., Mao, H. & Liu, Z. (2024). Hybrid architectures for sequence modeling: Combining Transformers and state space models. In *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*. [no place]: NeurIPS.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. & Berg, A. C. (2016). SSD: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer, 21–37.

Liu, X., Zhang, Y. & Chen, H. (2025). Fine-grained crop pest classification based on multi-scale feature fusion and mixed attention mechanisms. *Frontiers in Plant Science*, **16**: 1500571.

Ma, C., Li, Y. & Wang, Z. (2023). Temperature-dependent performance of electronic nose for stored-grain insect detection. *Sensors and Actuators B: Chemical*, **378**: 133158.

Ma, C., Zhang, Y. & Wang, L. (2025). Pest Manager: A systematic framework for precise pest counting and identification in invisible grain pile storage environment. *Pattern Recognition*, **158**: 111020. Available at: https://arxiv.org/abs/2409.07483 (Accessed: 30 July 2026).

Mankin, R. W., Hagstrum, D. W. & Guo, M. (2021). Acoustic detection of stored-product insects. *Journal of Stored Products Research*, **93**: 101847.

McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. [no place]: PMLR, 1273–1282.

Nayak, M. K. & Daglish, G. J. (2018). Importance of phosphine resistance in stored grain pests. *Journal of Stored Products Research*, **75**: 45–52.

Neethirajan, S. (2021). Digital agriculture: The role of artificial intelligence in pest management. *Biosensors and Bioelectronics*, **175**: 112891.

Neethirajan, S., Karunakaran, C., Jayas, D. S. & White, N. D. G. (2010). Infrared imaging for insect detection in stored grain. *Biosystems Engineering*, **106**(3): 215–222.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv. Available at: https://arxiv.org/abs/1706.05098 (Accessed: 30 July 2026).

Scheirer, W. J., de Rezende Rocha, A., Sapkota, A. & Boult, T. E. (2013). Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(7): 1757–1772.

Shorten, C. & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, **6**(1): 1–48.

Skendzic, S., Zovko, M., Zivkovic, I. P., Lesic, V. & Lemic, D. (2021). The impact of climate change on agricultural insect pests. *Insects*, **12**(5): 440.

Subramanyam, B. & Hagstrum, D. W. (1996). *Integrated management of insects in stored products*. New York: Marcel Dekker.

Sze, V., Chen, Y. H., Yang, T. J. & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, **105**(12): 2295–2329.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, **30**: 5998–6008.

Vedavathi, N. & Bharadwaj, R. S. (2024). HCBiL-DMN: An effective food infestation detection from stored food grains using deep learning model. *Multimedia Tools and Applications*. doi:10.1007/s11042-024-19450-y.

Wang, L., Ma, C. & Li, X. (2025). RP-DETR: End-to-end rice pests detection using a transformer. *Plant Methods*, **21**: 15. doi:10.1186/s13007-025-01381-w.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J. & Sun, L. (2023). Transformers in time series: A survey. *Proceedings of the IEEE*, **111**(5): 544–585.

World Bank (2024). *Missing food: The case of postharvest grain losses in Sub-Saharan Africa*. Report No. 60371-AFR. Washington, DC: World Bank.

Wu, X., Hong, R. & Tian, Q. (2021). Multi-scale feature fusion for fine-grained visual recognition. *Pattern Recognition*, **120**: 108163.

Zhang, Y., Wang, S. & Liu, H. (2022). Image-based insect detection in stored grain: A systematic review. *Computers and Electronics in Agriculture*, **198**: 107038.