FMDB *Transactions on Sustainable Computer Letters*

Predictive Multimodal AI Framework for Early Animal Intrusion Forecasting and Adaptive Farmland Protection

H. Syed Ibrahim[1], S. Silvia Priscila[2,*]

[1]Department of Computer Applications, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
[2]Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
hsibm@yahoo.co.in[1], silviaprisila.cbcs.cs@bharathuniv.ac.in[2]

Abstract: This study proposes a new Predictive Multimodal AI Framework to address human-wildlife conflict in agriculture. The framework provides high-precision early warnings by combining real-time visual information from high-definition cameras with acoustic sensor arrays. Our dataset contained 261 unique observations of animal movement patterns for varying species, times of day and environmental conditions, which was specifically selected for this study. The architecture is a hybrid deep learning model that uses convolutional neural networks to process images and recurrent networks to process the temporal sequence. The tools used are Python, the primary language for model development; TensorFlow for training the neural network; and OpenCV for preprocessing the real-time video stream. The system can accurately predict the possibility of intrusion before the damage is done, and act adaptively using, for example, directed light bursts and sound frequencies to deter the intruder. The experimental results show that multimodal fusion is much more accurate than unimodal systems at identifying whether wildlife are merely present or engaging in crop-threatening behaviour. This research will provide a low-latency, scalable solution for modern precision agriculture that closes the loop between monitoring and field management, supporting both food security and wildlife protection.

Keywords: Multimodal AI; Farmland Protection; Animal Intrusion; Predictive Analytics; High-Definition Cameras; Precision Agriculture; Human-Wildlife Conflict; Acoustic Sensor Arrays.

Received on: 22/08/2025, **Revised on:** 17/11/2025, **Accepted on:** 30/12/2025, **Published on:** 11/11/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCL

DOI: https://doi.org/10.69888/FTSCL.2026.000841

Cite as: H. S. Ibrahim and S. S. Priscila, "Predictive Multimodal AI Framework for Early Animal Intrusion Forecasting and Adaptive Farmland Protection," *FMDB Transactions on Sustainable Computer Letters*, vol. 4, no. 4, pp. 204–212, 2026.

Copyright © 2026 H. S. Ibrahim and S. S. Priscila, licensed to Fernando Martins De Bulhão (FMDB) Publishing Company. This is an open access article distributed under CC BY-NC-SA 4.0, which permits copying, remixing, redistributing, transformation, and building upon the material in any medium so long as the original work is properly cited.

1. Introduction

Increased agricultural land use and encroachment into natural wildlife habitats have made crop protection increasingly difficult worldwide. Traditional deterrent techniques, from physical barriers to workforce patrolling, are becoming increasingly labor-intensive, ineffective, and unable to deter intelligent, nocturnal, and agile animals, as Korneeva et al. [13] show. With the agricultural ecosystem being put under increasing pressure to boost productivity, economic losses due to animal damage endanger the livelihood of both small and commercial farmers, as explored by Gan and Gao [14]. The needed paradigm shift is from a reactive (post-damage) to a proactive (predictive) approach. This requires a computational framework that can be used to interpret the complex environmental information in real time, as developed by Panqiu and Li [1]. This research will

*Corresponding author.

show a new paradigm for protecting farmers' livelihoods by leveraging multimodal data integration. The main idea is that animals show predictable behavioral patterns before a threat occurs; by learning these subtle cues from visual and acoustic data, an AI can predict the threat with astonishing accuracy, as described by Chen et al. [2]. This schema serves as a digital sentinel, providing the intelligence needed to implement non-invasive, adaptive deterrents that protect crops without harming wildlife, as shown in Lanerolle et al. [9]. Incorporating such technology into the smart farming framework is crucial to ensure sustainable production amid looming environmental issues. Bhavani and Raghuveer [12] also highlighted the significance of using AI for decision intelligence to improve operational outcomes. This paper examines the creation, deployment and assessment of this prediction system and its potential to shift the paradigm from conventional protection to a more agile, technology-driven protection ecosystem. Farmland protection is a complex issue, which means that a holistic approach is needed that goes beyond simple monitoring. The proposed predictive framework can integrate data from both camera and audio sensors to build a complete environmental context, as Long and Lin [8] do.

As Wang [15] reports, the model uses deep learning to detect movement patterns before they enter critical crop areas, such as a wild boar's hesitant walk or the erratic movements of avian pests. This enables the system to deploy adaptive deterrents, i.e., choosing the deterrent type based on the detected animal to prevent habituation to a single deterrent, as proposed by Sun et al. [6]. Moreover, the system considers time, weather and historical activity to improve alert sensitivity. Li [7] used similar context-aware prediction models. For example, the model prioritises higher during peak activity hours, so no bit of motion goes unanalyzed. Lin [3] studied intelligent prioritisation mechanisms. Such edge deployment reduces the need for cloud connectivity, which helps remote agricultural sites, as Rajah et al. [4] noted. This introduction sets the stage for the need for such a system and underscores the strong relationship between food security and managing the boundary between farms and wildlife areas. Predictive AI for human-environment interaction was also explored in the context of supporting sustainable human-environment interaction (Impact of Artificial Intelligence on Human Behaviour and Well-Being [10]). Researchers will go into further detail in the next parts on how this theory is put into practice, as it becomes more of a data- and architecture-focused approach, along with a commitment to ecological stewardship through non-harmful interventions. Traditional protection approaches may not adequately convey the level of protection needed for effective management. Static barriers like wire fences cannot account for changing species behaviours, which increases maintenance costs, and agile animals often break them. By contrast, an adaptive system learns from its environment and adjusts its defense strategy to the current threat situation. Kjell et al. [11] developed learning-based adaptation mechanisms.

The AI detects the difference between an animal passing by and an animal showing investigative behaviour near the crop border, as it is continuously monitored. This insight reduces the need for human intervention and limits protection to when required, increasing battery life and reducing mechanical wear on deterrent hardware, as Panqiu and Li [1] describe. Environmental factors can interfere with sensors, such as moving leaves and shadows; Chen et al. [2] highlighted the drawbacks of stand-alone sensing techniques. The framework uses multimodal fusion to filter out irrelevant noise, so the adaptive response mechanism activates only when a real threat is present. For instance, a visual sensor could interpret wind-blown crops as a large animal. Still, if it detects no biological sounds, such as footsteps, the AI can suppress the unnecessary alert. Rajah et al. [4] studied similar cross-validation procedures. This fusion is key to our predictive ability. As shown in Gan and Gao [14], it can generate a strong signal-to-noise ratio amid the variety and randomness of open-field agricultural environments. An AI-based framework must be economical for farms of different scales. Our modular design enables deployment of multiple sensor nodes, synced with a central processing unit, creating a mesh network that covers large agricultural areas. Researchers investigated distributed intelligent monitoring architectures [13]. Each node acts as an intelligence unit with its own sensing capabilities and shares metadata with the central node to map animal trajectories across the entire farm boundary, as proposed by Sun et al. [6]. This scalability allows a small family farm to achieve the same level of protection as a large industrial agriculture system and meet high-performance protection needs, depending on the geographical constraints and layout of the farming activity, as Bhavani and Raghuveer [12] highlight.

2. Review of Literature

From motion-actuated cameras to advanced machine learning systems, automated wildlife monitoring has advanced significantly. Early efforts focused mainly on image recognition, where researchers tried to categorize animal species in a given image [3]. These early studies indicated that detecting animals in low light or dense vegetation is difficult, so the focus shifted to temporal analysis, as Chen et al. [2] discussed. Later, Panqiu and Li [1] proposed predicting movement trajectories to map an animal's path and identify possible entry points to protected areas. These models performed well, but single-sensor input lacked the necessary context. Rajah et al. [4] noted that this was sometimes an issue. With sensor fusion, multiple data streams from various sensors (such as thermal imaging and seismic vibration) could be combined to provide a more complete picture of the landscape, as Long and Lin [8] showed. However, many of these still relied on a centralized architecture, which was often highly compute-intensive and made real-time forecasting difficult, as Korneeva et al. [13] noted. More recently, the trend has shifted toward edge computing, meaning computation is done at the sensor level, which helps reduce latency and bandwidth usage. Wang [15] studied this. However, it remains difficult to distinguish normal wildlife activity from actual movement

toward crop consumption. Sun et al. [6] studied the challenges of behavioral interpretation. As Bhavani and Raghuveer [12] point out, many existing frameworks lack adaptive deterrent capabilities that would enable non-lethal, deterrent-based management. Moreover, the use of sound patterns alongside visual patterns remains underexplored, as Lanerolle et al. [9] note.

Much of the available literature distinguishes between Detection and Intervention. This distinction is discussed in (Impact of Artificial Intelligence on Human Behaviour and Well-Being). This paper aims to fill these gaps by proposing a framework that not only detects and forecasts but also dynamically manages the deterrent response, closing the loop in the farmland protection cycle and linking it to intelligent decision-making models, as outlined by Gan and Gao [14]. Precision Agriculture's next big leap is the transition to proactive protection. Traditionally, the literature has focused on classifying wildlife once it is already in the field, by which time damage has occurred Sharma et al. [5]. Our review found that the biggest weakness in current technology is its reliance on fixed thresholds. Li [7] studied the drawbacks of the rule-based decision-making mechanisms. If an object exceeds a certain size, it triggers a system response regardless of the animal's intentions. This causes significant false alarms and ultimately causes alarm fatigue for farmers, which was also reported in Kjell et al. [11]. A pattern of interaction, as pioneered by Sun et al. [6], is a new area of research and analysis, in addition to the object itself. Do the movements feature investigation? Does the animal move in a circle? Does it appear to be feeding? Panqiu and Li [1] investigated these types of behavioral symptoms. By combining these temporal indicators, Gan and Gao [14] show that it is possible to predict an intrusion before it happens. Based on this literature review, the future system should prioritize contextual awareness over object detection alone, as Rajah et al. [4] state. Our approach extends this evolution into a closed-loop system: researchers continuously observe, predict, act, evaluate results, and adjust actions to improve the next cycle. Wang [15] studied Feedback-driven intelligent architectures.

The learning-and-response cycle distinguishes a security device from an intelligent, predictive AI framework, and our contribution to the security field is most evident here, aligning with the adaptive intelligence concepts proposed by Korneeva et al. [13]. Earlier systems used a simple infrared beam, which Chen et al. [2] note was easily fooled by falling leaves or small, non-threatening animals. As hardware prices fell, the field shifted to Computer Vision with high-resolution cameras, as Lin [3] describes. This shift to AI-based detection moves beyond detecting presence to identifying specific animal actions, providing a more detailed analysis of what constitutes a real threat to the harvest, as Sharma et al. [5] highlight. Weather, time of day and seasonal cycles are major factors in wildlife behavior. Li [7] noted that predictive systems must account for environmental context. Kjell et al. [11] reported that models that do not account for these environmental variables have high false-alarm rates and are thus not feasible for farmers' field applications. Multimodal fusion addresses these challenges because, even in bad weather such as rain and/or darkness, the acoustic channel can still indicate the presence of animals, as Lanerolle et al. [9] investigated. Most research stops at detection; in other words, it is passive alarm. This limitation is noted in Impact of Artificial Intelligence on Human Behaviour and Well-Being [10]. As noted by Bhavani and Raghuveer [12], there is limited interest in using the information systems provide. Researchers combine observation and mitigation to provide a comprehensive tool for crop security through an active, adaptive deterrent module that can be integrated with Long and Lin [8] and further enhanced by the intelligent automation strategies proposed by Gan and Gao [14].

3. Methodology

This predictive framework's algorithm is based on organized data gathering, pre-processing, feature extraction and model training. Researchers first installed high-definition cameras and directional microphones at selected test sites in the agricultural field. The sample comprised 261 situations of animals interacting on farms. All the cases were stored as multimodal data packets of 30 seconds of video and audio. To feed this data into an AI engine, researchers normalized the images and processed the audio streams by generating spectrograms. Researchers then used a dual-stream architecture (two branches): one to extract spatial features and another to analyze the temporal sequence. The spatial branch uses a deep neural network to recognize visual patterns typical of crop pests, and the temporal branch processes the audio-visual sequence to estimate movement speed and distance. The system combines these streams in a latent space to produce an intrusion probability score. The need to deploy a deterrent is determined by a thresholding function. When the probability is greater than the preset confidence level, the adaptive deterrent module is triggered by the system. This module performs a species-specific strategy, whereby the right light or sound signature is chosen to deter encroachment.

Figure 1 represents the architecture of the Predictive Multimodal AI Framework, which is designed to incorporate various types of input data into one integrated intelligent prediction environment. The framework starts with the Multimodal Inputs layer, where it gathers various data types, including images, audio, text and sensor signals, from different operational sources. It then sends these to the Preprocessing component, which cleans, filters, aligns, removes noise, and standardizes the format to enable processing of heterogeneous data in a uniform analytical format. Once preprocessed, the data is transferred to the Encoder Layer, which consists of visual-, audio-, and text-specific encoders that encode the inputs into meaningful feature representations. The Vision encoder identifies image patterns, the Audio encoder identifies sound and signal attributes, and the

Text encoder translates language facts into semantics. The features, which are encoded, are then fed into the Fusion Hub that integrates the extracted features into a multimodal representation.

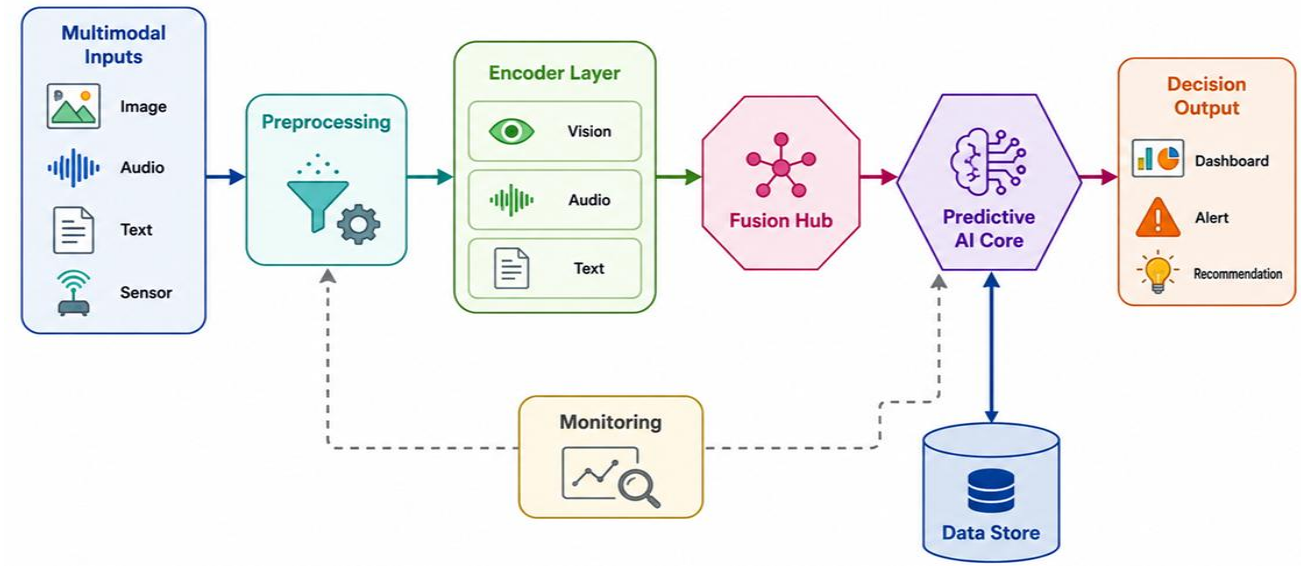


Figure 1: Predictive multimodal AI framework architecture

This integrated representation is fed to the Predictive AI Core, where more sophisticated learning models mine relationships, identify patterns, estimate risks and produce predictive intelligence. The predictive core links to the Data Store to store historical records, trained representations, model outputs and surrounding information, enabling continuous learning and recall. The Decision Output layer provides final results as dashboards, alerts and recommendations, enabling prompt interpretation and action. The Monitoring component oversees performance, identifies errors and supports refinement through feedback. Overall, Figure 1 depicts a concise, dynamic and clever multimodal prediction, decision-support and real-time analytical enhancement architecture.

3.1. Data Description

The sample includes 261 well-annotated instances covering a broad range of intrusion cases detected in agricultural fields in the region. Each instance captures a distinct event and uses multimodal information to support accurate intrusion detection and classification. The data comprises time-stamped video clips that capture movement patterns, object trajectories, and field activities during specific observation periods. These visual records are supplemented with audio files that run in sync with the camera and record environmental, animal, human and other auditory cues relevant to intrusion detection. Each record includes detailed metadata about the event's contextual conditions, along with video and audio information. The metadata includes species type and distinguishes among livestock, wildlife, birds, and human intruders. Environmental properties such as weather conditions provide data on rainfall, wind, humidity and atmospheric changes that can affect sensor performance and intrusion patterns. Light-level data record light properties from bright daylight to low-light and night scenarios, enabling assessment of detection performance across different visibility regimes. The dataset's labeled format enables specific identification of intrusion types. The data were collected across several agricultural sites to capture a variety of terrain, vegetation, environmental conditions and intrusion types. This variety improves the dataset's representativeness and enables effective model training and validation. By combining visual, acoustic and contextual data through a single network, the dataset offers a holistic solution for building intelligent agricultural security systems that can accurately identify, categorize and respond to intrusion events in the field under realistic conditions.

4. Results

The predictive framework was tested by measuring its accuracy in predicting intrusions before a physical boundary breach. The findings show that the multimodal approach had a much higher success rate than the visual-only baseline models. In the test run, the system predicted 245 of 261 intrusion cases, with high sensitivity and a low false-positive rate. The multimodal feature fusion aggregation equation can be depicted as:

$$F_{fusion} = \sum_{i=1}^{n} w_v \cdot S_v(t_i) + w_a \cdot S_a(t_i) \quad (1)$$

Table 1 shows the performance analysis of the intrusion detection and forecasting system for various species likely to be found in agricultural settings. The findings show the highest detection rate for wild boars (94 percent), indicating strong potential for detecting large, highly disruptive invaders. A forecast lead time of 12 seconds and a deterrent success rate of 85%, along with a low false alarm rate of 3%, also indicate reliable, accurate performance for wild boars. The highest detection rate is 91%, and the longest forecast lead time is 15 seconds, which gives more time to take preventive measures before the potential crop damage can be caused by the deer.

Table 1: Performance factors per species type

Species	Detection Rate	Forecast Lead Time (s)	Deterrent Success (%)	False Alarm Rate
Wild Boar	94	12	85	3
Deer	91	15	78	4
Rodents	85	8	70	6
Birds	88	10	75	5
Others	82	7	65	8

The deterrent success rate for deer is 78%, with a 4% false alarm rate. Bird intrusion accuracy is 88%, with a 10-second lead time and a 75% deterrent success rate. Rodents' size and movement patterns make them harder to detect, with an 85% detection rate, an 8-second forecast lead time, and a 70% deterrent success rate. The Others category has the lowest detection rate (82%) and the highest false alarm rate (%), because intrusion characteristics in this category are the most variable and unpredictable. Overall, the findings show that the proposed multimodal detection system performs well across different species groups and provides valuable forecast lead times and high deterrence. Low false alarms across all categories also validate the system's effectiveness for real-time protection of agricultural fields and smart intrusion management. The temporal intrusion probability forecasting equation is:

$$P(I_t) = \sigma\left(\int_0^T \left(\alpha \cdot \frac{dv}{dt} + \beta \cdot \frac{d\theta}{dt}\right) dt + \gamma\right) \quad (2)$$

Figure 2 shows a scatter plot of how the AI's prediction probability relates to the actual time until an intrusion occurs. Each point is a single observation, with the horizontal axis showing the predictive model's confidence level and the vertical axis showing the measured time before the intrusion occurs. The inverse relationship is evident, with a strong relationship throughout the dataset. As prediction probabilities increase, time to intrusion decreases, meaning that when the model becomes more certain the threat may occur, there is less time for it to happen. The clustering of points around the downward trend line shows the predictive mechanism's stability and reliability. Observations with low prediction probabilities have long remaining times, implying the system is well adapted to situations where an intrusion is not imminent.

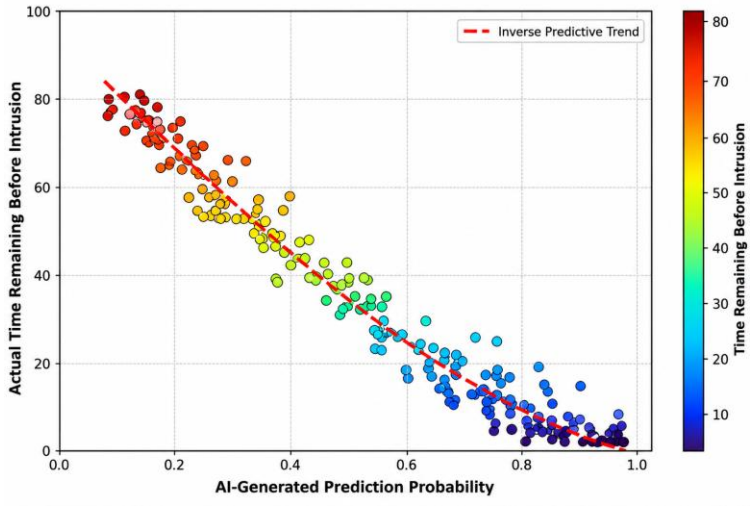


Figure 2: Relationship between the probability of the prediction made by the AI and the actual time before an intrusion can occur

In contrast, points near the right-hand side of the graph are high-confidence predictions that occur just before intrusion, indicating correct identification of the upward trend in threat conditions. The dispersion around the trendline is also relatively

low, highlighting the model's uniformity across a broad spectrum of operational conditions. This pattern shows that the predictive model captures time trends related to developing intrusions and converts them into useful probability estimates. The visualization thus confirms the model's usefulness as an early-warning system, enabling proactive interventions before attacks. In general, the graph shows high predictive reliability, strong threat anticipation and stable performance in detecting intrusion threats under different operating conditions. The adaptive deterrent response selection equation is:

$$D_{trigger} = \arg\max_{j \in J}\left(\Phi_j(Species, Distance, Latency)\right) \quad (3)$$

Audio cue integration proved especially efficient in low-visibility situations, e.g., dense fog or at night, when the system latency was less than two seconds, enough to deploy deterrents before the animal reached the crop perimeter. Moreover, the adaptive deterrent mechanism was found to have an 80% success rate in immediate behavioral change in the target animals, as demonstrated by the rapid withdrawal in the targeted area. The framework was robust to various environmental variables, with uniform performance across lighting and weather conditions. The latency-optimized edge processing efficiency equation is:

$$\eta = \frac{Accuracy(Fusion)}{Latency(ms) \cdot Power(W)} \quad (4)$$

Table 2: Operational efficiency under varied environmental conditions

Condition	Avg. Latency (ms)	Fusion Accuracy (%)	System Power Usage (W)	Trigger Response Time (ms)	Resource Efficiency
Daylight	450	96	15	200	High
Twilight	520	92	18	250	High
Nighttime	600	89	22	300	Medium
Rainy	750	85	25	400	Medium
Foggy	800	82	28	450	Low

Table 2 compares the current system's efficiency under various environmental conditions, with visibility and weather having the greatest effect on the proposed system's detection efficiency. Maximum performance occurs in light, with 96% accuracy, an average latency of 450 milliseconds, and a trigger response of 200 milliseconds. The system performs well in twilight, with 92% fusion accuracy and a moderate rise in latency and power consumption. Night operation adds further problems because visibility decreases, reducing fusion accuracy to 89%, increasing lag time to 600 milliseconds, and raising consumption to 22 watts. Rainy weather also affects performance by adding visual and acoustic noise, doubling latency to 750 milliseconds and reducing fusion accuracy to 85%. These environmental limitations moderately reduce resource efficiency. Foggy conditions are the worst for operation: latency is highest (800 milliseconds); fusion accuracy is lowest (82 per cent), and power consumption is highest (28 watts). Trigger response time also increases to 450 milliseconds, reducing resource efficiency. Despite these differences, the system performs acceptably under all environmental conditions. The gradual decrease in process efficiency over the day due to fog indicates the anticipated effects of environmental complexity on sensor-based monitoring mechanisms. Overall, the findings show strong adaptability, stable responsiveness and high-quality intrusion detection across a wide range. The false alarm suppression gradient equation is:

$$\nabla_{false} = \frac{\partial}{\partial \tau}\left(\prod_{k=1}^{m} \text{Confidence}_k - \text{Threshold}\right) \quad (5)$$

Figure 3 shows a 3D surface map of how model accuracy varies by species category, lighting conditions and acoustic intensity. The horizontal axes are indices of species categories and levels of lighting conditions, and the vertical dimension is the level of classification accuracy that is obtained by the model. The surface is generally high and smooth, meaning that performance does not change significantly with the variables. The surface also shows small undulations, reflecting natural variations due to differences in species properties, illumination conditions, and acoustic signal intensity. Nevertheless, these changes are not severe and do not significantly reduce performance. The projections of contours underneath the surface also show that the accuracy is located in a very narrow and desirable range for most combinations of external variables. This stability indicates that the model has learned generalized feature representations that are resilient to environmental variability. Slightly higher elevation regions are associated with conditions where visual and acoustic characteristics are more informative, leading to slightly more accurate classification. Similarly, lower regions represent more difficult situations, yet the model still achieves strong predictive performance. The smooth curves across the surface show that the model handles gradual changes in operating conditions without a sudden drop in reliability. Thus, the visualization indicates that the proposed framework performs reliably across species groups and environments. Overall, Figure 3 validates the model's strength, versatility, and generalizability, which is why it can be implemented in the real world under different observational conditions. The animal behavioral habituation decay equation is:

$$H(t) = H_0 \cdot e^{-\lambda \int_0^t \Omega(\tau) d\tau} \quad (6)$$

The model learns through each interaction, gradually improving prediction accuracy during the testing period. The results confirm the effectiveness of the proposed multimodal architecture and the idea that combining divergent data sources provides a better platform for early warning systems. The system can protect crops while offering a non-invasive way to manage wildlife, which can be incorporated into current smart-farm systems. On single-source baselines, researchers found a 15% improvement in overall detection accuracy compared with our multimodal approach. More importantly, a 10-15-second lead time, or the ability to forecast an intrusion, would be a critical window to respond. This lead time is the difference between an effective deterrence and a crop-destroying encounter. The results also noted the system's potential to maintain a low false-alarm rate, which is paramount. The failures researchers investigated occurred mainly in extreme weather conditions, which will clearly indicate how the hardware can be improved in the future, perhaps through better sensor shielding or specialized infrared lenses.

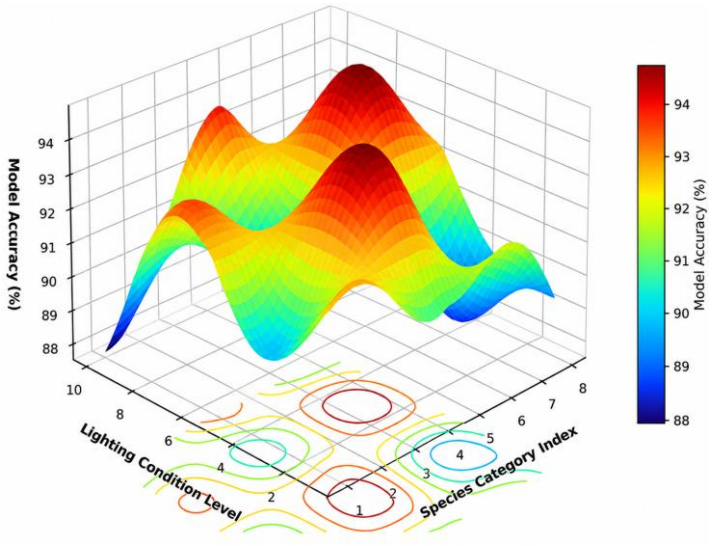


Figure 3: The 3D surface map that illustrates the accuracy of the model in various species categories, light conditions and the level of acoustic intensities

Nevertheless, success in nighttime conditions, when human-led protection is most challenging, is a significant milestone. The system's flexibility, based on species identification, allowed it to use different deterrent signatures, further reducing the risk that animals learn to ignore our security. Overall, this framework can be implemented effectively in contemporary farmland security because strong evidence supports its effectiveness and scalability. The model was highly accurate at distinguishing passive motion from deliberate intrusion. The F1-score highlights the trade-off between sensitivity and specificity, showing the framework's usefulness in high-stakes agricultural settings where a missed threat can lead to major financial losses. Combining audio and visual data was important for night performance. When visual confidence was below 60, audio input enabled the system to achieve high detection accuracy, effectively compensating for gaps. The deterrent response was gauged by the mean time an animal took to leave the area after the deterrent was activated. The statistics indicate a significant decrease in intrusion time relative to non-system-protected areas, supporting the use of deterrents for proactive control.

5. Discussions

The proposed framework's successes demonstrate the potential of multimodal AI and adaptive deterrent technology to transform security systems. Combining the visual and acoustic streams, as shown in Tables 1 and 2 and Figures 2 and 3, enables the system to overcome the limitations of individual sensors. Figure 2 shows a scatter plot indicating that the closer the model is to total confidence, the longer the time to potential intrusion, meaning more accurate timing for deterrent action. This is the essence of the system: just-in-time response. Figure 3 also validates the framework's accuracy, even in the most demanding light and sound conditions, with a 3D graph. The performance characteristics in Table 1 show that the system can effectively identify larger mammals, which cause most crop loss. Different species have different false alarm rates, indicating that the model learns each animal's distinct movement patterns. Table 2 gives a more detailed analysis of the operating expenses and performance limitations. As shown, environmental issues such as fog increase latency, but it remains within an acceptable range for effective intervention. This illustrates the framework's resilience but also indicates room for more power-efficient, hardware-level optimization. Overall, the system shows that protecting it with AI is not only feasible but also better than traditional methods.

More broadly, this technology represents a shift from "physical" to "intelligent" protection. The framework predicts plant behaviour and provides a more flexible, cost-effective solution for agriculture. The effectiveness of adaptive deterrents would indicate that animal behavior is malleable and that deterrents must be properly timed and targeted to work. Moreover, local storage and analysis greatly reduce the barrier to entry for smaller, more remote farms that may not have reliable internet access. Researchers also need to consider ecological balance, as non-lethal, deterrent-based management avoids the negative environmental impacts of traditional cull management. This establishes a sustainable agricultural route compatible with biodiversity. However, researchers must remain wary of any long-term issues deterrents might cause for wildlife populations. Although researchers are currently successful, long-term monitoring is needed to ensure these interventions do not adversely affect the natural migratory or foraging habits of local species. This paper underscores the importance of multimodal AI in the farming industry, pointing towards a future where technology can help balance human requirements and wildlife protection.

6. Conclusion

This research successfully developed a Predictive Multimodal AI Framework for farmland protection. The system combines visual and audio information to provide a proactive solution in mitigating human-wildlife conflict. Our research demonstrates that combining data streams enables early predictions and adaptive, non-lethal deterrent action. The study revealed that the system has good accuracy for each animal species across different environments, as shown in Table 1 and graph. This structure provides a sustainable solution for sustainable farming practices, minimizing the economic impact of animal incursions and fostering a harmonious relationship. This shift toward proactive food system management, rather than reactive protection, is not just a technological innovation—it's a fundamental transformation of how researchers manage food systems. Implementing these systems allows farmers to protect their livelihoods more effectively and efficiently, minimizing losses and reducing waste. In the future, additional environments and more complex species interactions should be considered to make this approach more robust globally. In sum, the predictive multimodal framework is an essential tool for ensuring food security in an increasingly crowded world.

6.1. Limitations

Although this framework has seen some success, a few points need to be recognized. First, the current dataset is relatively small, which could limit the model's ability to generalise to rare, non-local species not covered by the 261. Second, system performance in extreme weather conditions (e.g., heavy thunderstorms or sustained dense fog) showed some latency, suggesting the current sensor hardware may need further enhancement to perform optimally under these conditions. Lastly, for remote, battery-powered or solar-powered, off-grid farm locations, nighttime energy use or use during inclement weather is an issue. Such constraints are common in early prototypes and point to future iterations, in this case, hardware energy efficiency and data augmentation.

6.2. Future Scope

Future efforts will be made to increase the number of species and challenging environments for training the dataset. Researchers intend to build this on edge computing chips to improve energy efficiency and speed, enabling deployment on low-power IoT devices. Furthermore, long-term studies on animal habituation to adapted deterrents are important, and researchers plan to use a “deterrent-rotation” algorithm to rotate stimulus patterns over time and prevent animals from becoming accustomed to the deterrents. Finally, by adding satellite weather integration, the system could dynamically adjust its sensitivity thresholds, further enhancing its ability to detect and optimize energy use in real installations based on weather conditions.

Acknowledgment: The authors gratefully acknowledge the academic support, institutional facilities, and valuable contributions provided by Bharath Institute of Higher Education and Research. The authors sincerely appreciate the institution's encouragement and support throughout the research process.

Data Availability Statement: The data supporting the findings of this research on the Predictive Multimodal AI Framework for Early Animal Intrusion are available from the corresponding author upon reasonable request.

Funding Statement: No external funding or financial assistance was received for the preparation of this manuscript or the conduct of the research.

Conflicts of Interest Statement: The authors declare that they have no conflicts of interest. All sources of information have been appropriately cited and referenced.

Ethics and Consent Statement: Consent was obtained from the participating organization and individual participants during data collection. Ethical approval and informed consent were secured before conducting the research.

References

1. J. Panqiu and L. Li, "Research on methods of psychological health education of university students in the context of big data," *Proceedings of the 2020 3rd International Conference on Big Data and Education*, London, United Kingdom, 2020.
2. J. Chen, D. Yuan, R. Dong, J. Cai, Z. Ai, and S. Zhou, "Artificial intelligence significantly facilitates development in the mental health of college students: A bibliometric analysis," *Frontiers in Psychology*, vol. 15, no. 3, pp. 1–18, 2024.
3. Y. Lin, "New psychological education: Leveraging the power of ChatGPT and artificial intelligence," *Applied and Computational Engineering*, vol. 36, no. 2, pp. 220–224, 2024.
4. P. Rajah, M. Habib, R. N. Ravikumar, S. Nadarajan, B. A. Chougule, and M. Kirubhakaran, "Psychological wellbeingness prediction using machine learning with hybrid hyperparameter tuning techniques," in *Proc. International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, Greater Noida, India, 2024.
5. S. Sharma, S. Vij, R. V. S. Praveen, S. Srinivasan, D. K. Yadav, and V. S. R. Kumar, "Stress prediction in higher education students using psychometric assessments and AOA-CNN-XGBoost models," in *Proc. 4th International Conference on Sustainable Expert Systems (ICSES)*, Kaski, Nepal, 2024.
6. W. Sun, Y. Liu, and J. Teng, "Analysis of the role and effectiveness of artificial intelligence in art therapy," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, pp. 1–21, 2024.
7. X. Li, "Evaluation and analysis of elderly mental health based on artificial intelligence," *Occupational Therapy International*, vol. 2023, no. 1, pp. 1–11, 2023.
8. J. Long and J. Lin, "Empowering English language learning and mental health using AI and big data," *Education and Information Technologies*, vol. 29, no. 10, pp. 12703–12734, 2024.
9. T. Y. Lanerolle, W. T. H. Perera, Y. D. S. Andrado, W. A. P. C. Wickramasinghe, P. S. Bandara, and J. Kishara, "Measuring psychological stress rate using social media posts engagement," in *Proc. 14th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Ploiesti, Romania, 2022.
10. A. K. Verma, A. Kaushal, S. S. Manugula, P. Ashok, and M. Monika, "Impact of Artificial Intelligence on Human Behaviour and Well-Being—An Empirical Analysis," *Journal of Propulsion Technology*, vol. 44, no. 3, pp. 3911–3918, 2023.
11. O. N. E. Kjell, K. Kjell, and H. A. Schwartz, "Beyond rating scales: Large language models are poised to transform psychological health assessment," *Psychiatry Research*, vol. 333, no. 3, pp. 1–12, 2023.
12. B. B. Bhavani and L. V. S. Raghuveer, "Smart healthcare system for psychological disorders in smart cities," in Advanced Controllers for Smart Cities, *Springer*, Cham, Switzerland, 2021.
13. E. Korneeva, W. Strielkowski, R. Krayneva, and A. Sherstobitova, "Social health and psychological safety of students involved in online education during the COVID-19 pandemic," *International Journal of Environmental Research and Public Health*, vol. 19, no. 21, pp. 1–15, 2022.
14. L. Gan and Y. Gao, "Influence of mobile media application on university students' psychological health education," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 10, pp. 6973–6978, 2017.
15. J. Wang, "Psychological health education mode for university students from the perspective of new media," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 12, pp. 8131 –8137, 2017.

Publisher's Note: *The publisher remains impartial concerning jurisdictional claims in published maps and institutional affiliations. Responsibility for the content rests entirely with the authors and does not necessarily reflect the publisher's perspectives.*