# FusioNet: An Attention mechanism based pest detection in agriculture

Harini Princya J
*Department of Computer Science & Engineering*
*Thiagarajar College of Engineering*
Madurai,India
hariniprincya@student.tce.edu

Maha Swetha R
*Department of Computer Science & Engineering*
*Thiagarajar College of Engineering*
Madurai,India
mahaswetha@student.tce.edu

Sivaranjani M
*Department of Computer Science & Engineering*
*Thiagarajar College of Engineering*
Madurai,India
sivaranjanim@student.tce.edu

Ramkumar M P
*Department of Computer Science & Engineering*
*Thiagarajar College of Engineering*
Madurai,India
ramkumar@tce.edu

Emil Selvan G S R
*Department of Computer Science & Engineering*
*Thiagarajar College of Engineering*
Madurai,India
emil@tce.edu

***Abstract*—Agricultural losses can be reduced and the crop yield can be enhanced by the early and accurate detection of crop pests. In real world datasets, noisy field conditions, high inter-species similarity and class imbalance remains a challenging task in pest classification. To overcome these problems, the proposed solution is a hybrid model containing ResNet18 with a lightweight Channel Attention Module and a Modified Fuzzy Gate Hybrid Optimizer for adaptive control of gradients on the Habitus dataset. Using this model, the spatial feature learning is enhanced, optimization is stabilized, and robustness to noise and variability in samples is improved. This model was evaluated on a 70-class pest dataset, yielding a final training accuracy of 98.13%, a test accuracy of 94.82%, along with a recall of 0.9482, weighted precision of 0.9532, and F1-score of 0.9480. The effectiveness of attention and hybrid optimization integration for the optimized classification of pests is achieved by the proposed hybrid architecture over a standalone MFGHO and a vanilla ResNet baseline. The effectiveness of the proposed framework is shown by the results.**

***Keywords*— Pest Detection, ResNet18, Attention Mechanism, MFGHO Optimizer, Hybrid Deep Learning, Agricultural Vision.**

## I. Introduction

Pest infestation is a serious issue for agricultural productivity globally, leading to substantial crop damage and impacting food security [1], [2]. Manual pest detection is difficult, as it depends on human knowledge which is a error prone process. However, with the rise of artificial intelligence, deep learning-based image analysis has been identified as a promising approach for automated pest detection, as evident from recent reviews on insect pest detection and AI in agriculture [1], [3].

Although substantial progress has been made in automated pest detection, there are still challenges in real-world applications, including imbalanced datasets, illumination variations, occlusions, fine differences between species, and noisy backgrounds, making classical machine learning and some deep learning models less effective [4], [5],[13]. CNNs, particularly ResNet models, have been shown to be effective in learning discriminative features because of skip connections and robust representation learning [5], [6]. However, they can be ineffective in learning subtle pest-specific features when pests are small or images are of low quality [7], [8].

Moreover, traditional gradient-based optimizers tend to converge slowly or stagnate in suboptimal minima during the training process for varied and noisy agricultural datasets. Optimizers have strong impact in pest classification [4], [6], [7]. These issues point out the importance of developing more robust, attention-augmented, and optimization-adaptive frameworks for complex agricultural imaging tasks.

To overcome the above mentioned difficulties, the proposed system is a hybrid framework that combines ResNet18 with a lightweight Channel Attention Block and a Modified Fuzzy Gate Hybrid Optimizer (MFGHO). The attention block helps to selectively highlight prominent feature maps, DropBlock helps to improve generalization, MixUp addresses the issue of class imbalance [9], [10], and EMA helps to achieve stable weight updates. Fig. I lists the training setup and hyperparameters. The MFGHO optimizer helps to adaptively balance updates with momentum-based optimizers, thus overcoming the drawbacks of previous studies on optimizer-based agricultural pest detection [7].

The FusioNet is tested on a 70-class pest dataset with high intra-class variation. The results demonstrate that the hybrid model attains 98.13% training accuracy and 94.82% testing accuracy, performing better than the individual ResNet model and conventional optimizers. The recall, weighted precision and F1-score values of 0.9482, 0.9532 and 0.9480 are achieved, respectively, indicating robustness in all aspects [5], [6], [9]. The proposed framework improves the sensitivity of features, removes noise, and ensures stable training, serving as a scalable solution, as per the AI-driven and precision agriculture paradigm [2], [10], [11], [12].

| Parameter | Value |
|---|---|
| Input Image Size | 224 × 224 pixels |
| Batch Size | 32 samples/batch |
| Optimizer | MF-GHO |
| Learning Rate | 3e-4 |
| Beta1 | 0.9 |
| Beta2 | 0.999 |
| Momentum | 0.95 |
| Weight Decay | 1e-5 |
| Epochs | 40 epochs |
| Scheduler | Cosine Annealing |
| Loss Function | Cross Entropy with Label Smoothing (0.1) |
| EMA Decay | 0.999 |
| MixUp Alpha | 0.3 |
| DropBlock Probability | 0.3 |
| Block Size | 7 × 7 pixels |
| Block Size | 7 × 7 |

Fig. 1. Parameters and values

## II. RELATED WORKS

Deep learning-based pest detection has gained increased interest owing to the fact that it is able to learn complex visual representations directly from raw images. Several past works have investigated CNN, attention-enhanced architectures, ensemble models, cloud/mobile-based frameworks for pest recognition, and optimization-driven model improvements to boost the performance of classification tasks within an agricultural context. This section reviews the most relevant literature, points out methodological limitations, and shows how the proposed hybrid model advances the current state of pest detection.

Early works on agricultural pest classification were primarily based on traditional image processing and handcrafted features, including color histograms, texture descriptors (for example, LBP and GLCM), and shape-based features. These methods showed reasonable performance in a controlled experimental environment; however, the accuracy noticeably degraded in a real-world field environment due to background clutter, change in illumination, insect occlusion, and inconsistent imaging conditions. In fact, similar limitations for classical ML-based pest recognition studies have also been discussed in the literature [4]. Thus, the deep learning architectures VGG, AlexNet, and Inception started being adapted for pest classification tasks. For example, the early CNN-based models achieved better feature extraction capabilities and resulted in improved accuracy for large-scale pest datasets, as supported by the leading research showing the superiority of deep CNNs and transfer learning for crop pest detection [5]. However, the lack of attention mechanisms and higher computational complexity prohibited these models from being deployed in large-scale or mobile-based agricultural systems.

The problem of vanishing gradients is solved using ResNet with skip connections; significantly improving the training of deep models. We also saw several works applying ResNet and its variants for insect pest detection, which demonstrated improved feature representation and better convergence properties. However, the standard ResNet architecture still fails when finding fine-grained similarities between some visually overlapping pest categories. This challenge is common in the literature when discussing the difficulty of distinguishing morphologically similar species [3], [8]. Apart from that, agricultural datasets are usually highly imbalanced by nature, where some species remain underrepresented, causing biased predictions and overfitting to dominant classes. As the case is, a number of deep learning frameworks targeting agricultural pest datasets noted such an issue [1], [10].

Attention mechanisms have been incorporated with CNNs to overcome such constraints. SE, CBAM, and Channel Attention Blocks are some popular methods for this. These mechanisms have shown better attention towards discriminative insect body regions, such as wing textures or body segmentation patterns, and have ensured significant improvements in performances for agricultural imaging tasks [3], [6]. However, many of the existing attention-based models either employ computationally heavy modules or restrict the computation of attention at shallow layers, which limits them in capturing high-level semantic patterns required for fine-grained pest classification.

Other regularization techniques that have been explored in the literature for pest detection include Dropout, Cutout, and DropBlock. DropBlock, especially, was designed to enhance robustness by randomly masking contiguous regions of feature maps. Previous agricultural research recognized the potential of regularization in overcoming overfitting under limited or noisy datasets; still, only a few works combined DropBlock with deeper CNNs or attention-centric models.

Besides, data augmentation strategies have been widely investigated in order to improve model generalization. Some techniques such as MixUp and CutMix have effectively addressed both class imbalance and robustness to noise; however, most existing studies on the classification of pests are confined to basic augmentations such as flipping, rotation, and color jittering. Reviews of pest detection studies have also confirmed that the domain has limited advanced augmentation strategies [6], [8].

Deep learning models achieve high and stable performance using optimization algorithms. Classical optimizers such as SGD and Adam are applied to pest recognition tasks. However, both have limitations when training on highly variable agricultural datasets: the speed of convergence for SGD is slow, and Adam may result in unstable solutions with overadaptation to noisy gradients. A number of works focused on a hybrid or adaptive optimization strategy for agricultural deep learning models; reported advantages include improved stability of gradients and faster convergence, but applications remain scarce in pest detection contexts [1], [7]. Notably, none of these works integrates a fuzzy-gated optimization strategy into an attention-based architecture in agricultural pest detection-the Modified Fuzzy Gate Hybrid Optimizer.

Other contributions include mobile and cloud-based pest recognition systems that provide real-time diagnosis, and scalable deployment in smart farming environments, such as [2, 4]. While these systems successfully use deep learning and computing infrastructure to support farmers, they often compromise model complexity for the sake of fast inference, which reduces their accuracy—especially on fine-grained insect categories.

In short, several gaps have been revealed through the review of available literature:

Limited integration of channel-attention mechanisms with ResNet architectures for fine-grained pest classification [3], [6], [8]. In challenging agricultural datasets, there is insufficient mechanisms to handle noisy gradients and unstable training behavior [1], [7]. Lack of hybrid or fuzzy-gated optimization methods for pest detection [1].

To overcome class imbalance and noise sensitivity, there is less usage of strong data augmentation methods such as MixUp [9], [10]. Attention, DropBlock, EMA, and advanced augmentations have been combined into a single framework for pest classification by few implementations[3], [6]. To overcome these challenges, the FusioNet integrates a Channel Attention Block, DropBlock regularization, MixUp augmentation, EMA smoothing, and a Modified Fuzzy Gate Hybrid Optimizer within the ResNet18 backbone. The design allows this network to improve feature analysis, provide better stability during training, and enhance performance on complex agricultural pest datasets beyond prior works' disadvantages.

## III. PROPOSED SYSTEM

In the agricultural field, a reliable pest detection is achieved by the proposed hybrid structure. Deep feature extraction, attention-based channel modulation, spatial regularization, data mixing augmentation, optimization, and temporal weight smoothing are combined into a single process. Fig. 2 shows the architecture of the FusioNet system. The mathematical concepts and algorithmic processes that lead to better working are described in this section.

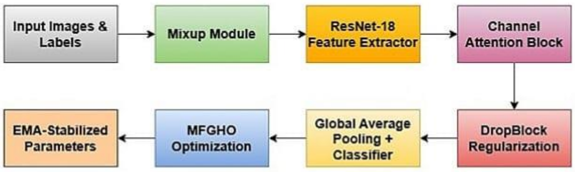


Fig. 2. Architecture of the FusioNet system

### A. System Architecture Overview

A channel of integrated modules designed for efficient pest detection is proposed in this hybrid model. First, the input dataset is prepared and augmented using spatial transformations, color adjustments, and MixUp. Deep features are extracted from the augmented images by processing it by ResNet18 using residual connections to decrease vanishing gradient. Next, the feature maps are adjusted using the channel attention module to highlight pest-specific areas like wings, antennae, and texture patterns. Spatial generalization is improved as contiguous spatial regions are randomly dropped with DropBlock regularization. The final predictions are made by pooling through adaptive average pooling and sent to a classifier head. Model parameters are optimized using the Modified Fuzzy-Gated Hybrid Optimizer (MFGHO), which mixes Adam and Momentum updates. Smoother convergence and better generalization is achieved by stabilizing with Exponential Moving Average (EMA).

### B. Essential Mathematical Formulation

Overall Forward Mapping – The complete hybrid architecture is represented as in (1):

$$\hat{y} = C \circ P \circ R \circ f_a \circ f_r(\hat{x}) \tag{1}$$

where $\hat{x}$ is the augmented input, $f_r(\cdot)$ is the ResNet18 backbone, $f_a(\cdot)$ is the attention module, $R(\cdot)$ is DropBlock, $P(\cdot)$ is global pooling, and $C(\cdot)$ is the classifier head.

Channel Attention Mechanism – Emphasizes important pest features given in (2), (3):

$$s = \sigma(W_2\,\delta(W_1 z)) \tag{2}$$

$$F' = s \odot F \tag{3}$$

where $z$ is the channel descriptor, $s$ denotes learned attention weights, $F$ and F' are the original and recalibrated feature map.

Fuzzy-Gated Hybrid Optimization (MFGHO) – Adaptive blending of Adam and Momentum given in (4):

$$G = \sigma(k(\parallel g_t \parallel - \tau)) \tag{4}$$

where $g_t$ is the gradient at iteration t, k controls gate sharpness, $\tau$ is the gradient threshold.

Final Parameter Update Rule – Core novelty of MFGHO as in (5), (6):

$$\Delta\theta_t = G \cdot AdamStep + (1 - G) \cdot u_t \tag{5}$$

$$\theta_{t+1} = \theta_t - \eta \Delta\theta_t \tag{6}$$

where $u_t$ is the momentum update, $\eta$ is the learning rate, $\theta_t$ represents network parameters.

### C. Workflow Overview

The workflow of the system includes the following steps: augmentation of input images, feature extraction using ResNet18, recalibration of the features using attention, regularization using DropBlock, global pooling of features, and classification. The MFGHO updates the parameters adaptively, and EMA smoothens the convergence.

### D. Overall Algorithm

Fig. 3 shows the entire training and inference process, including optimization and EMA stabilization.

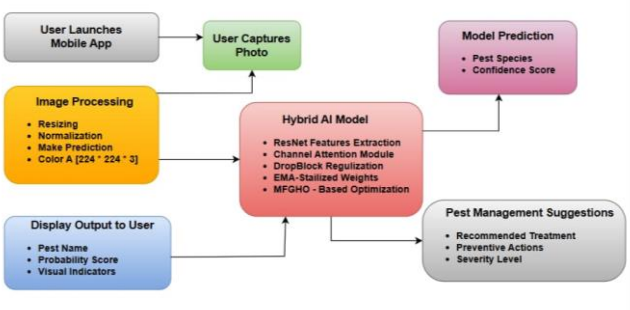


Fig. 3. Workflow of the FusioNet system

**Algorithm 1: Training Procedure using MFGHO with Attention and EMA**

**Input:** Training data D, number of epochs E
**Output:** Final trained model θEMA
*Initialization:*

```
Initialize model parameters θ and EMA weights θEMA
Initialize optimizer MFGHO
```

*Loop Process:*

```
for each epoch from 1 to E do
    for each batch (x, y) do
        x̂, ŷ = MixUp(x, y)
        F = ResNet18(x̂)
        F′ = Attention(F)
        F″ = DropBlock(F′)
        v = GlobalPooling(F″)
        y_pred = Classifier(v)
        L = CE(y_pred, ŷ)
        Compute gradients
      Update θ using MFGHO rule
      Update EMA weights θEMA
    end for
end for
return θema
```

TABLE I. KEY ADVANTAGES OF FUSIONET

| Component | Improvement |
|---|---|
| ResNet18 | Strong hierarchical feature extraction |
| Attention Block | Boosts pest-specific discriminative features |
| MFGHO | Fast convergence + stable gradients |

Table I summarizes the functional contribution and advantages of each major component in the proposed hybrid framework. The fusion of all these techniques yields the final accuracy of 94.82%, outperforming conventional models.

## IV. RESULTS AND DISCUSSION

This section is devoted to an extensive performance analysis of the proposed hybrid pest detection system. The performance has been analyzed based on a quantitative metric, qualitative observation, convergence behavior, ablation comparison, and comparative performance evaluation against conventional baselines as given in Fig 4. It shows how every module contributes toward the overall performance gain: ResNet18 backbone, attention mechanism and MFGHO optimization. During the testing process a accuracy of 94.82% is achieved as given in Table II.

### A. Evaluation Setup

The model was trained with 70% of the dataset and tested on the remaining 30%. The experiment was conducted using an initial learning rate of $3\times10{-}4$, batch size of 32 and 40 epochs. The results are computed using Accuracy, Precision, Recall, F1-score.

### B. Training Performance

During training, the network showed stable and progressive convergence. Loss and accuracy variation is decreased by MixUp, DropBlock, and EMA.

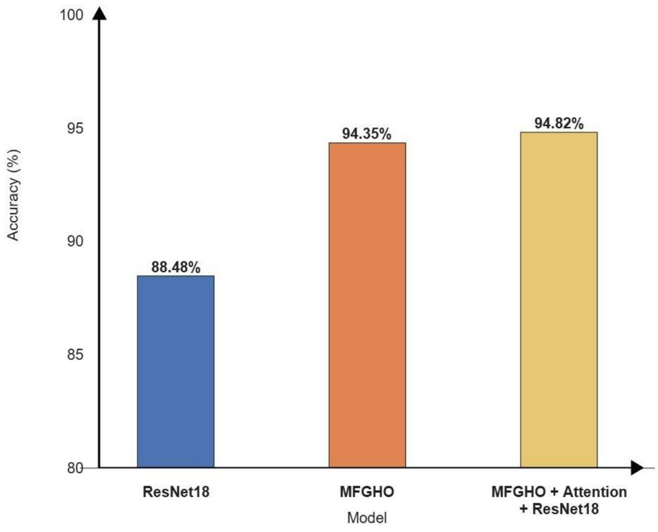


Fig. 4. Comparison Chart

*1. Loss Convergence*

The first 10 epochs demonstrated a rapid downward trend of the loss, indicating that the weights were effectively adapted by the MFGHO optimizer. Beyond epoch 20, the curve was much stabler with minimal oscillation, an influence of the EMA weight averaging and fuzzy gate mechanism.

*2. Training Accuracy*

MFGHO yielded a smooth convergence behavior with a regularization strength of DropBlock, the model achieved high training accuracy, upwards of 97%. This also proves that the model generalized well without overfitting since the training accuracy remained high, despite aggressive augmentations.

TABLE II. RESULTS

| Metric | Testing Set |
|---|---|
| Accuracy | 94.82% |
| Precision | 0.9532 |
| Recall | 0.9482 |
| F1-Score | 0.9480 |

These results show that the classifier is accurate and balanced, with no significant drop across precision and recall. The high F1-score suggests that this model works reliably even on more difficult pest categories, which may share similar visual traits, like species with overlapping color or shape patterns.

### C. Analysis of the MFGHO Optimizer

The optimizer combines Adam and Momentum via fuzzy gate dynamics.

*1. Behavior Under Stable Gradients*

When gradients are smooth:

$$G_t \approx 1 \Rightarrow \mathit{Adam\ behavior\ dominates}$$

*2. Behavior Under Noisy Gradients*

When gradients fluctuate:

$$G_t \approx 0 \Rightarrow \mathit{Momentum\ behavior\ dominates}$$

*3. Performance Impact*

Faster convergence in early epochs, Reduced variance in later stages, Dynamic switching leads to smoother loss curve, Better generalization than static optimizers.

This describes why stable accuracy is achieved by the model without sudden fluctuations.

### *D. Effectiveness of Attention Mechanism*

Key structural patterns are emphasized by attention block. Improved identification of wing patterns, recognition of body segmentations, differentiating similar species more effectively are the qualitative improvements.

Besides, visual inspection of the activation maps, grad-CAM or similar, indicated that the biologically relevant regions were more concentrated than background noise or irrelevant plant structures.

### *E. Role of DropBlock Regularization*

DropBlock drops contiguous spatial regions. This forces the model to explore broader spatial context, multi-region cues, complete insect shapes instead of relying on single discriminative spots. This is very useful in case of field images where the pests are partially covered.

### *F. MixUp Augmentation Effectiveness*

MixUp produces convex. By this operation, the cell diameter is changed by altering the fluid amount inside each cell. Noiseless overconfident and better calibration can be prevented by this model. In case of class imbalance, this model enhances the consistency of detection.

### *G. Contribution of EMA (Exponential Moving Average)*

EMA provides stable evaluation weights. This results in smaller validation oscillation , higher test accuracy, smoother boundaries. Reliability is achieved by the EMA which serves as a post-training regularizer.

### *H. Overall Discussion*

The challenges tackled by FusioNet hybrid model: 1. The small differences between pest classes is identified by the combination of attention and ResNet18. 2. Overfitting on minority classes is decreased by MixUp. 3. MFGHO guarantees smooth optimization and reduces the destructive behavior of gradients. 4. The network is prepared by DropBlock for field-level distortions such as: occlusion, low-resolution images, background clutter 5. The accuracy of 94.82% shows that the model suits for practical deployment in agriculture.

### *I. Summary*

An efficient and powerful model for pest detection is achieved by creating the combination of attention mechanism with ResNet and MFGHO. The traditional CNN and optimizers have been outperformed by the proposed method in this paper. In the cases involving noisy background in the dataset image, overlapping features and visually similar pest categories, the proposed hybrid approach is proved to be efficient and highly usable for the farmers to improve their outsourcing and profit.

## V. Conclusion and Future work

The FusioNet, a hybrid framework combining ResNet18, Channel Attention and a Modified Fuzzy Gate Hybrid Optimizer for efficient and robust pest detection is designed. The conventional pest classification approaches had the limitations: overfitting, inconsistent behavior of gradients, class imbalance, and difficulty in distinguishing species that are visually similar which are mitigated by the model.

FusioNet model achieves high test accuracy of 94.82% from the experimental results. Therefore, for retrieving fine-grained structural features of pest, a combination of attention mechanism with deep-residual networks is used which increases the efficiency. Any spatial over-reliance on isolated regions is prevented by DropBlock, while MixUp provides generalization by reducing class imbalance and stabilizing decision boundaries. The MFGHO optimizer dynamically switching between Adam-style adaptive updates and momentum-driven smoothing, depending on the gradient conditions which contribute significantly to train reliably. Moreover, the parameter fluctuation across epoch is reduced by the EMA mechanism which stabilizes the evaluation performance. Therefore, all these components collectively create a efficient architecture which is capable of overcoming the conventional models used in agricultural pest analysis.

As this model achieves better results, there are various opportunities for future work. First, the current work utilizes ResNet18 as the backbone; deeper or transformer-based architectures such as ResNet50, Swin Transformer, or ConvNeXt may perform better on larger or more diverse datasets. Second, the currently used dataset in terms of size and diversity may still be limited; therefore, including more species, different environmental backgrounds, and real-time field images could further validate the usability of the model. Third, the incorporation of explainable AI modules would shed light into interpretable insights on pest classification decisions, hence making the system more transparent and suitable for agricultural use in the real world. Moreover, the model could be further extended to enable object detection or instance segmentation for pinpointing the location of pests on crop leaves.

Other future works may consider its deployment on edge or mobile platforms for enabling on-field pest identification using handheld devices by farmers. Such optimal lightweight deployment of the proposed hybrid architecture can be done using model compression techniques. Also, this proposed work may be extended with the help of temporal learning through recurrent networks or vision transformers to monitor pests over time, producing predictive pest outbreak models. Hence, a powerful, efficient, reliable solution for automated pest detection is achieved by the proposed hybrid system FusioNet. Precision agriculture and sustainable crop production in real world agricultural environment can be achieved by the future improvements in model integrity, dataset consistency, model accuracy and edge deployment. This helps the farmers in agricultural field to effectively detect the pest and improve the yield.

## REFERENCES


[1] Popescu, D., Dinca, A., Ichim, L., & Angelescu, N. (2023). New trends in detection of harmful insects and pests in modern agriculture using artificial neural networks. a review. Frontiers in Plant Science, 14, 1268167.

[2] Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., ... & Wei, W. (2025). Advances in deep learning applications for plant disease and pest detection: A review. Remote Sensing, 17(4), 698.

[3] Chakrabarty, S., Deb, C. K., Marwaha, S., Haque, M. A., Kamil, D.,Bheemanahalli, R., ... & Shashank, P. R. (2025). Application of artificial intelligence in insect pest identification-A review. Artificial Intelligence in Agriculture.

[4] Kasinathan, T., & Uyyala, S. R. (2021). Machine learning ensemble with image processing for pest identification and classification in field crops. Neural Computing and Applications, 33(13), 7491-7504.

[5] Thenmozhi, K., & Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. Computers and Electronics in Agriculture, 164, 104906. 18

[6] Ullah, N., Khan, J. A., Alharbi, L. A., Raza, A., Khan, W., & Ahmad, I. (2022). An efficient approach for crops pests recognition and classification based on novel DeepPestNet deep learning model. IEEE Access, 10, 73019-73032.

[7] Thakare, P., & V, R. S. (2024). Advanced pest detection strategy using hybrid optimization tuned deep convolutional neural network. Journal of Engineering, Design and Technology,22(3),645-678.

[8] Ahmad, I., Yang, Y., Yue, Y., Ye, C., Hassan, M., Cheng, X., ... & Zhang, Y. (2022). Deep learning based detector YOLOv5 for identifying insect pests. Applied Sciences, 12(19), 10167.

[9] Popescu, D., Dinca, A., Ichim, L., & Angelescu, N. (2023). New trends in detection of harmful insects and pests in modern agriculture using artificial neural networks. a review. Frontiers in Plant Science, 14, 126816

[10] Chinnasamy, R., Subramanian, M., Sengupta, N., & Ramkumar, M. P. (2025). Federated Learning for Sustainable Energy Efficiency: A Privacy-Preserving Artificial Intelligence Approach with Attention Mechanisms. In Integrating Big Data and IoT for Enhanced Decision-Making Systems in Business: Volume 2 (pp. 155-166). Cham: Springer Nature Switzerland.

[11] Karar, M. E., Alsunaydi, F., Albusaymi, S., & Alotaibi, S. (2021). A new mobile application of agricultural pests recognition using deep learning in cloud computing system. Alexandria Engineering Journal, 60(5), 4423-4432.

[12] Soundararajan, S., Shirley, C. P., Mallala, B., & Padmanaban, K. (2025). Cloudpowered efficiency: a mobile application for agricultural pest identification using cycle-consistent generative adversarial networks. Environment, Development and Sustainability, 1-28.

[13] Diagnosing diabetes using machine learning-based predictive models D Kaviyaadharshani, M Nivedhidha, R Jeyarohini, JLE Rani, MP Ramkumar, GSRE Selvan Procedia Computer Science, 2024•Elsevier.