Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 3, June, 2025

IMpc-PyrYOLO: Hybrid YOLO Based Feature Pyramidal Network for Pest Detection in Rice Leaves

Ramesh Shunmugam [1, 2], Sumendra Yogarayan [1, 3*]
Siti Fatimah Abdul Razak [1, 3], Md. Shohel Sayeed [1, 3],

[1] *Center of Intelligent Cloud Computing (CICC), CoE of Advanced Cloud, Multimedia University, Melaka, Malaysia.*

[2] *Department of Mechatronics Engineering, Rajalakshmi Engineering College, Chennai, 602105, India.*

[3] *Faculty of Information Science and Technology, Multimedia University, Jln Ayer Keroh Lama, Melaka, 75450 Malaysia.*

Abstract

Pests pose a significant threat to global food security, making early detection crucial for maintaining crop health. Traditional pest detection models suffered from inefficiencies such as long processing time and low accuracy, which hinder effective disease management. In order to overcome these existing issues, a novel improved efficient channel attention mechanism assisted feature pyramidal network-based YOLO model (IMpc-PyrYOLO) for rice leaf pest detection. The model integrates an efficient channel attention (ECA) mechanism with the feature pyramidal network (FPN) in the YOLO network to improve multi-scale feature extraction and pest classification accuracy. Additionally, an upgraded weighted Gaussian wiener filter (Up-weGaf) is employed for noise reduction, improving image clarity. The model is evaluated on two benchmark datasets such as IP_RicePests and IP102. Experimental results demonstrate that IP_RicePests attains a precision of 95.8%, recall of 96% and F1-score of 95.9%, whereas IP102 attains a higher precision of 97.8%, recall of 96%, mean Average Precision (mAP) of 95.9% and Intersection over Union (IoU) of 97% with processing time of 2.5 seconds. The proposed model significantly outperforms existing methods in accuracy and computational efficiency, which provides a robust and scalable solution for real-time pest detection in agriculture.

Keywords:

Faster R-CNN; Faster-PestNet; Scalable and Efficient Network; Gaussian Filtering; Channel Attention Mechanism and Batch Normalization.

Article History:

Received:	17	February	2025
Revised:	18	May	2025
Accepted:	23	May	2025
Published:	01	June	2025

1- Introduction

Rice is one of the most important food crops in the world. Every year, rice pests and diseases cause significant crop losses, affecting global food security. Precise, real-time monitoring and pest and disease prediction are critical for efficient treatment and preventing crop reduction [1]. Manual surveys in paddy fields are still used on a modest scale to monitor pests and diseases on the rice canopy. Before visually estimating the level of damage, agricultural professionals examine disease and insect lesions on rice leaves and stems in a paddy field [2, 3]. Thus, researchers' experiences determine the inefficiency and labour intensity of the manual survey approach. Maintaining real-time disease and pest monitoring on the rice canopy is challenging [4, 5].

The application of deep learning (DL) techniques to picture categorization has shown promising results. In recent years, tea [6], tomato [7], apple [8], peach, grapevine and pear [9, 10] have been analyzed by using these methods. Five distinct tea plant diseases were identified from tea leaves using a neural network ensemble (NNE). In order to predict rice blast diseases, the neural network used by Bhagawati et al. [11] was trained using meteorological variables like temperature, precipitation, humidity, and wind speed and direction. The above-mentioned research mainly concentrated

* **CONTACT**: sumendra@mmu.edu.my

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-03-029

on precise plant disease identification and classification. The experts utilize a variety of convolutional neural network (CNN) architectures, including LeNet-5, GoogLeNet and AlexNet. Some analyses have used a collection of numerous neural network architectures. These investigations made a substantial contribution to the automated, precise identification and categorization of plant diseases. However, they did not focus on enhancing the training approaches for the model that they had developed and implemented [12, 13].

Among the two-stage techniques, Ren et al. [14] initially proposed the Faster R-CNN. However, it can't provide fast identification because its detection method involves two stages: feature extraction and region suggestion created by a neural network. Redmon et al. [15] proposed a one-stage item identification approach known as you only look once (YOLO). It uses a single CNN feature extraction network to detect object outputs from end to end. This approach varies from Faster R-CNN in that it combines classification with conventional object identification and target localization. This combination greatly enhances the processing speed of the model. The algorithm's enhanced processing speed makes it ideal for detection jobs in complicated agricultural settings.

As data size and quality increase, a big data-based technique called DL shows a greater ability to generalize [16, 17]. The researchers expect an extensive and high-quality data set. In addition, the data will be expected to cover a variety of scenarios. It is often difficult to collect enough data to cover every situation. Data augmentation is an excellent method for extending the data set [18-20]. Rice pests and diseases significantly impact global food security by causing substantial crop losses each year. Traditional monitoring approaches, such as manual field surveys, are labour-demanding, time-consuming, and rely on specialist knowledge, making real-time pest and disease identification challenging. While DL models have demonstrated promising outcomes in plant disease identification, current models frequently focus on classification without optimizing training procedures. Two-stage detection models like Faster R-CNN suffered from slow processing speeds, making them unsuitable for real-time agricultural applications. Although one-stage models such as YOLO improve detection efficiency, they require improved feature extraction and noise reduction mechanisms to attain high accuracy in complex agricultural environments. Furthermore, the need for large, high-quality datasets remains a significant barrier, necessitating data augmentation strategies to improve model generalization. To overcome these issues, a novel hybrid DL approach is introduced for rice pest detection. The main objectives are mentioned below:

- To present an "Upgraded weighted Gaussian wiener filter (Up-weGaf)" for removing noise and contrast enhancement. Compared to conventional filters, Up-weGaf's weighted Gaussian feature better retains these small characteristics, enabling the model to distinguish pests from the leaf's background.
- To implement an "Improved efficient channel attention mechanism assisted feature pyramidal network-based YOLO model (IMpc-PyrYOLO)" for detecting pests in rice leaves.
- To minimize the model weights, channel pruning was performed. Scaling pest detection solutions is made simpler and less expensive with channel-pruning as they demand less computing infrastructure.

The following is a discussion of the full research workflow: The most pertinent current studies are mentioned in section 2. Section 3 defines the basic concepts and the suggested methodology. Furthermore, the experimental results are discussed in section 4. The findings of this study are summed up in Section 5.

2- Related Works

Due to its algorithm's intricacy and the scarcity of available data, pest identification methods in the literature are comparatively inaccurate in identifying and classifying pests. For identifying pests, Ullah et al. [21] suggested a new method known as the DeepPestNet framework. The suggested model has eleven learnable layers, including three fully connected (FC) and eight convolutional layers. The recommended model's generalizability was evaluated using image augmentation techniques. Deng's crop dataset was used to assess the performance of the suggested model. Inadequate variance in the dataset or poorly labelled data might result in biases and less generalization.

Li et al. [22] developed a method called the self-attention feature fusion model for rice pest detection (SAFFPest) to detect pests that look similar to rice. An initial step in improving the feature extraction network's capacity to handle pests with varying morphologies was the incorporation of a deformable convolution module. Secondly, in order to adjust the balance feature, the self-attention technique was used to get the balance characteristics of multiple feature maps. This resulted in more successful restoration of semantic data for some pests with similar phrases. The original model's batch normalizing strategy was later updated with the group normalization method to reduce the effect of batch size on the training model. To train and evaluate the model, the IP102 rice pest dataset was evaluated.

Due to its reliance on human abilities, the traditional diagnostic method of visual leaf assessment produces errors that lead to losses. A vision-based diagnostic tool called OryzaNet was suggested by Mendigoria et al. [23] to assess the quality of rice leaves. Using feature-based ML techniques like k nearest neighbour (kNN), naïve byes (NB), linear discrimination analysis (LDA), support vector machine (SVM), and decision tree (DT) for classification, the spectro-textural features of image samples were used to determine the state of health of rice leaves. A total of seventeen characteristics were extracted using the hue, saturation, and value (HSV) colour space thresholding method. Neighborhood-principal component analysis (NCA-PCA) selection was then used to further reduce them, resulting in a four-vector set named (Cr, a*, homogeneity, and contrast). The illness alterations were further classified using deep neural networks (DNN) such as ResNet-101, Inceptionv3, GoogLeNet, ResNet-50, and MobileNetv2. In terms of ML technique classification performance, the testing results demonstrated that the current model achieved better results. However, these ML techniques were highly dependent on data.

To detect and classify crop pests, Ali et al. [24] developed a lightweight DL framework named Faster-PestNet. A unique set of sample characteristics was first extracted by utilizing the MobileNet and subsequently identified by using the 2-step locator of the enhanced Faster-RCNN model. A local crops dataset was collected and tested using the trained Faster-PestNet approach, further demonstrating the recommended model's ability to generalize. The performance of DL models can be adversely affected by a small dataset owing to overfitting, which happens when a model does well on training data but poorly on new, independent test data.

A serious threat to agricultural results, rice hispa disease can significantly reduce crop yield. For classifying rice hispa diseases, Kukreja et al. [25] developed an enhanced model by combining random forest (RF) and convolutional neural networks (CNN). A dataset including 10,000 pictures, each depicting a unique intensity level, was assembled and pre-processed. In the CNN, unique characteristics are retrieved from images, and the RF classifier is responsible for combining these features in the final classification stage. The model's capacity to distinguish among different disease phases was enhanced by confusion matrix analysis. However, the existing model requires more computational power.

To minimize such losses, it will be helpful to identify rice leaf diseases early by utilizing thermal imaging cameras. The main problems with paddy cultivation are diseases and pests, which cause farmers worldwide to lose approximately 20% of their rice crop. Bharanidharan et al. [26] developed an enhanced Lemurs Optimization Algorithm as a filter-based feature transformation method to improve the precision of diagnosing different paddy diseases using machine learning approaches by examining the thermal pictures of paddy leaves. The five paddy diseases that were investigated were brown leaf spot, hispa, rice blast, bacterial leaf blight, and leaf folder. A total of 636 thermal pictures of six paddy leaves and healthy leaves were analyzed. Four machine learning algorithms, including LDA, RF, kNN, and histogram gradient boosting (HGB) classifiers, were used to evaluate the performance of the proposed model. When it comes to large-scale optimization problems, Lemurs optimization might show slow convergence. This is partly due to its balance between exploration and exploitation, which may not be optimal for speedy problem solving.

It is important to detect insect pest attacks when they are in their early phases so that they do not spread. Conventional techniques like visual detection are labour-intensive, ineffective, and challenging for growing enormous crops. By using DL, Ananyasreya et al. [27] had been developed an automated pest classification technique. Histogram equalization was utilized to pre-process and enhance input data. Common traditional classifiers, including RF, NB, SVM and logistic regression (LR), employ DL to extract features. However, these existing models had over-fitting and data privacy issues.

Tenriola et al. [28] trained various CNN architectures, namely MobileNet and Efficient NetB0, to detect pest attacks on rice. Here, the input images are collected from the open-source dataset, which contains 700 images per class taken directly from the field, where the images depict rice plants that have been peeled or opened to inspect for the presence of pests. A rigorous selection process was applied to enhance the quality and diversity of the dataset, which ensured only high-quality images were used. Data augmentation techniques were used to expand the dataset to 2000 images per class. The suggested model was tested using several metrics, such as accuracy, precision, recall, and f1-score, which were compared to existing models. Chaurasia et al. [29] suggested an extended YOLOV5 model for panicle detection in rice crops named PanicleDet, which was associated with occlusion, object overlap and varying object sizes. The model was trained using a dataset of paddy plants grown under controlled conditions, with images taken from different angles and time stamps before harvesting. In this study, the suggested model was evaluated based on various performances such as IoU, accuracy, and so on.

Tan et al. [30] suggested a DeepBGS and bilayer LSTM model for pest detection in rice leaves. Here, deep convolutional networks were used to extract pertinent information, and a sliding window-based technique was used to transform a one-dimensional spectral band sequence into a more interpretable two-dimensional matrix. Here, local and global correlations were captured by ensemble bilayer LSTM and convolutional layers of the model. A number of performance metrics, including accuracy, precision, recall, and f1-score, were used to assess the suggested model and compare it to other models from recent studies.

Hussain et al. [31] suggested that a novel farmland fertility algorithm with a DL-based automated rice pest detection and classification (FFADL-ARPDC) mechanism be introduced. This approach aids in the identification of rice pests from rice plant images. The images are pre-processed with a bilateral filter to eliminate noise and improve contrast. After the pre-processing technique, image features were retrieved using the NASNetLarge DL architecture. The FFA mechanism was utilized to tune the hyperparameters, which helps to enhance the model performance of the NASNetLarge, hence enhancing classification performance. Based on the Elman recurrent neural network (ERNN), the model accurately categorizes 14 types of pests. Song et al. [32] suggested a novel YOLOv8-SCS architecture with space-to-depth convolution (SPD-Conv) for rice pest detection. In addition to that, the suggested model integrates slide loss and context-guided block (CG block) to improve model performances. The input channel dimensions into spatial dimensions were reorganized by a convolutional module. This approach enables the model to capture fine-grained pest features more efficiently while maintaining a lightweight model architecture. In order to improve generalization across different samples and concentrate the model more on difficult samples during training, the binary cross entropy (BCE) function was employed.

From this overall analysis, the existing DL-based models such as DeepPestNet, SAFFPest and faster-PestNet struggle with generalization due to small and imbalanced datasets, which leads to biases and hinders accurate pest identification. Although deformable convolution and self-attention mechanisms have been introduced to enhance feature extraction, they failed to achieve robust classification when pests exhibit similar morphological characteristics. In addition to that, some traditional methods relying on ML-based classification often require extensive handcrafted feature extraction, making it more computationally expensive and dependent on dataset quality. Additionally, two-stage models, such as faster R-CNN, suffered from high latency, limiting their real-time applicability in large-scale paddy fields. Conventional vision-based techniques, including RF and CNN hybrid models, are computationally intensive and unsuitable for real-time deployment in resource-constrained environments. In contrast, optimization algorithms like Lemurs optimize and enhance disease diagnosis accuracy but result in slow convergence, which degrades timely detection and intervention. Furthermore, existing models often lack efficient noise reduction techniques, which makes them more vulnerable to variations in lighting, background clutter and occlusions in real-world agricultural settings. In addition to that, the minimum accuracy, efficiency, and time were consumed in these existing models. In order to overcome these gaps, a novel IMpc-PyrYOLO model is introduced for precise and real-time rice pest detection. Unlike typical CNN-based models, the proposed framework uses an improved FPN to improve multi-scale feature extraction, allowing for successful pest detection with a wide range of sizes and morphological variations. The attention mechanism (ECA) is included in the YOLO architecture to improve feature representation by stressing important spatial and channel-specific information. Compared to conventional two-stage models, the proposed method achieved end-to-end real-time detection with much less complexity while keeping excellent precision. The model's lightweight architecture enables effective deployment on edge devices, making it ideal for large-scale paddy field surveillance. By incorporating these novel improvements, the proposed method exceeds existing pest detection frameworks in terms of accuracy, speed, and adaptability to real-world agricultural conditions, contributing to long-term rice production and food security.

3- Proposed Methodology

Detecting and managing pests is essential in modern agriculture for crop health and yield, especially for staple cereals like rice and maize, which play a critical role in the world's food supply. In the face of obstacles such as complicated backgrounds and a multitude of parameters, timely and accurate pest identification is crucial for agricultural output. In this research, a novel advanced YOLO model with a feature pyramidal network for pest detection in rice leaf images. Gathered from an open-source dataset, the input photos undergo a pre-processing step to improve their quality by eliminating noise. Upgraded weighted Gaussian wiener filters (Up-weGaf) are used in the pre-processing stage. After pre-processing, improved efficient channel attention mechanism assisted feature pyramidal network-based YOLO model (IMpc-PyrYOLO) for pest detection in leaf images. Channel pruning is also performed to reduce the model weights, which helps to minimize model complexity. The proposed methodology is shown in Figure 1.

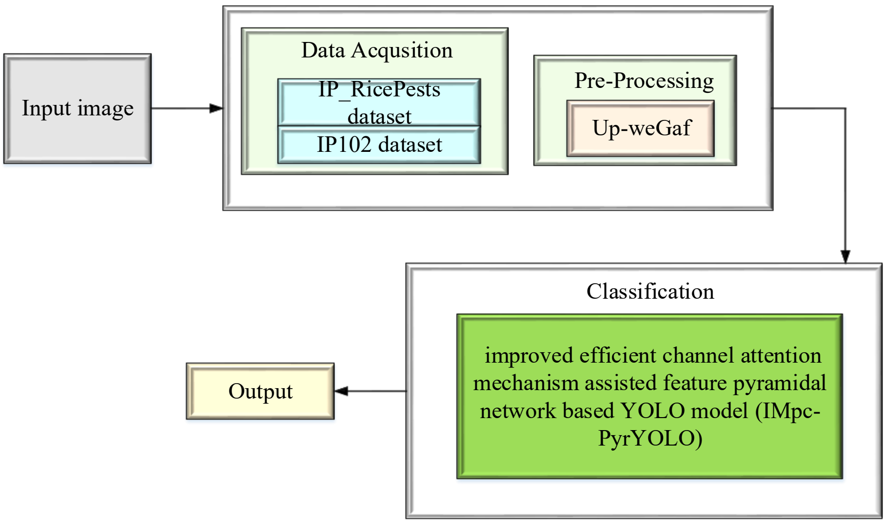

Figure 1. Block Diagram of Proposed Methodology

By addressing significant shortcomings in current techniques, the suggested framework provides an innovative approach to improve insect identification in rice crops. In the pre-processing stage, Up-weGaf were introduced to remove the noise present in the image. In contrast to conventional filtering techniques, Up-weGaf achieves a compromise between edge preservation and noise reduction, guaranteeing that the image retains most of the important pest-related features. In order to maximize feature extraction, this model presents a distinctive pyramid architecture coupled with a channel attention mechanism. The model generates higher detection accuracy for small, complicated and overlapping pest patterns by combining multi-scale feature representations and prioritizing critical pest-specific characteristics. In terms of accuracy and computing efficiency, these two approaches exceed current approaches to provide a reliable, effective, and scalable pest detection system.

3-1- Pre-Processing

3-1-1- Upgraded weighted Gaussian Wiener Filter (Up-weGaf)

For image processing applications like pest detection, a number of different pre-processing methods have been employed. Some of them are bilateral filter, anisotropic diffusion, adaptive histogram equalization (AHE) and non-local means (NLM) filtering. The bilateral filters are computationally costly and less efficient for complex textures or high noise levels. Anisotropic diffusion needs precise parameter adjustment and, if done incorrectly, might result in over-smoothing. NLM filtering is computationally intensive and very sensitive to parameter selection. In order to overcome these disadvantages, Up-weGaf were introduced for the pre-processing stage. One linear smoothing filter that can be used to eliminate Gaussian noise is the Gaussian filter. It is frequently employed in image processing. Gaussian filtering [33] is an overall image-weighted average method. The weighted average of every pixel's value and the outcomes of other pixels in close proximity determines its value. Every pixel's grayscale value was substituted with the weighted average of its nearby pixels' grayscale values using a Gaussian filtering template. The mathematical representation for the Gaussian filter is illustrated in Equation 1.

$$G_\sigma = \frac{1}{2\pi\sigma} f^{\frac{-(a^2+b^2)}{2\sigma^2}} \tag{1}$$

Here, the Gaussian function was denoted as G_σ. The standard deviation of Gaussian distribution was represented asσ. The overall distance within other pixels in the neighbourhood was indicated as a^2andb^2. The normalization factor of Gaussian distribution was denoted as $\frac{1}{2\pi\sigma}$. The exponential component of the Gaussian function was represented as $f^{\frac{-(a^2+b^2)}{2\sigma^2}}$

The goal of the Wiener filter, a linear filter, is to decrease the mean square error (MSE) in both the original and filtered pictures. Images that have been impacted by extra noise are expected to be improved [34]. The wiener filter achieves a balance between reducing noise and maintaining significant visual features by modifying the filter's characteristics. The mathematical representation for the wiener filter was expressed in Equations 2 to 4.

$$\hat{P}(a,b) = U(a,b)Q(a,b) \tag{2}$$

$$U(a,b) = \frac{J*(a,b)A_p(a,b)W}{|J(a,b)|^2 A_u(a,b) + WA_n(a,b)} \tag{3}$$

$$U(a,b) = \frac{J*(a,b)}{|J(a,b)|^2 + \frac{A_n(a,b)}{A_u(a,b)}} \tag{4}$$

where, the power spectrum of image and noise technique were denoted as $A_u(a,b)$ and $A_n(a,b)$. The estimated restored image in the frequency domain and observed degraded image in the frequency domain is denoted as $\hat{P}(a,b)$ and $U(a,b)$. The Wiener filter function is applied to restore the image complex conjugate of the original image spectrum, and the regularization parameter is denoted as $Q(a,b)$, $J*(a,b)$ and W.

A single filter type cannot properly manage some real-world noise patterns. This hybrid technique, which combines weighted Wiener and Gaussian filters, can successfully optimize the quality of images.

3-2- Rice Crop Detection by Improved Efficient Channel Attention Mechanism Assisted Feature Pyramidal Network based YOLO Model (IMpc-PyrYOLO)

Conventional CNNs, Random Forest (RF), and Support Vector Machines (SVMs) are examples of typical machine learning algorithms that are commonly used in current methods for detecting pests in rice leaves. Small item sizes or overlapping infestations are difficult for traditional approaches to handle. Conventional networks fail to capture multi-scale properties. YOLO-based feature pyramidal network has been chosen due to its real-time efficiency, superior localization, accuracy and ability to detect small and overlapping objects, which makes it well-suited for pest detection in complex agricultural environments. Unlike traditional CNN-based models like VGG16, ResNet and DenseNet, YOLO processes entire images in a single pass, which significantly reduces inference time. FPN architecture has been integrated into the YOLO model to enhance multi-scale feature representation and help to improve the detection of pests with varying sizes. Additionally, an ECA mechanism has been incorporated to refine feature selection and improve precision. Here, IMpc-PyrYOLO was implemented to address the aforementioned drawbacks. The channel attention method prioritizes key traits, increasing the detection accuracy. Tiny and overlapping elements are more correctly spotted because of the feature of the pyramidal network. The new YOLO-based technique allows for real-time pest identification with reduced computing overhead.

One of the most effective single-stage object detection methods is the YOLO. After dividing a picture into many grids, it uses non-maximum suppressions (NMS) to get rid of boxes that overlap. Then, the bounding boxes and item categories were predicted to correspond to every grid. YOLOv1 is well-known for its fast detection speed; however, it performs less well on small or closely spaced objects. A scalable and efficient network (E-ELAN) was introduced by YOLOv7 [35]. This network includes new transition modules. Additionally, YOLOv7 introduced parameterization techniques with E-ELAN. These improvements resulted in better detection performance. Furthermore, E-ELAN enhances feature extraction and semantic expression. YOLO adjusts effectively to visual changes because it uses all the image information during training. This model retains its excellent identification accuracy even in challenging situations with shifting illumination or varying object scales. Figure 2 illustrates the structure of IMpc-PyrYOLO.

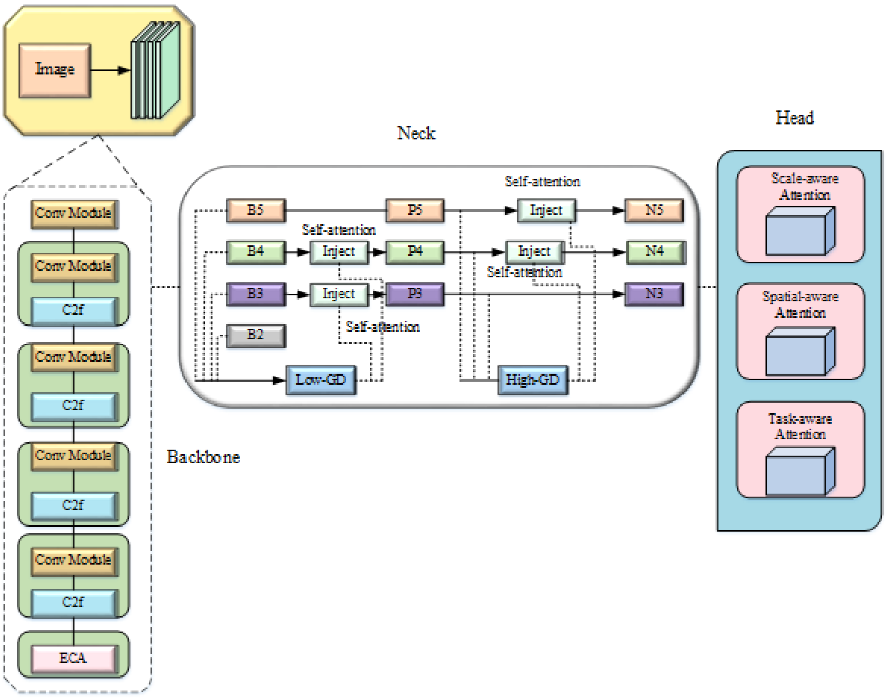

Figure 2. Structure of IMpc-PyrYOLO

In order to capture internal relationships inside the sequence, the self-attention mechanism assigns different weights to various parts and computes attention within the sequence directly. Attention is generated by several "heads" in a multi-head self-attention mechanism, each of which concentrates on a distinct region of the feature space. By allowing it to collect data from many representational spaces at various locations, this method improves the model's robustness and adaptability. To concentrate on globally significant information, the self-attention mechanism calculates relationships between a query$P$, value$R$and key$N$matrix. The mathematical representation for scaled dot-product attention was expressed in Equation 5.

$$Attention(PNR) = soft\,max\left(\frac{PN^T}{\sqrt{d_n}}\right)R \tag{5}$$

where, the dot products were normalized by the scaling factor $\sqrt{d_n}$.

This study uses the GD mechanism to enhance YOLOv8's feature fusion method. This improvement generates YOLOv8, which is more suitable for small object recognition by reducing information loss in conventional feature pyramid network (FPN) structures. The original recursive approach was replaced with a new "Gather-and-Distribute"

(GD) mechanism. By utilizing a single module, this method collects and combines data from all levels before transferring it to further layers. It has a Low GD branch and a High GD branch to improve the model's object detection capabilities across a range of sizes. Each of these branches uses different-sized feature maps to extract and combine attributes. By ensuring that each layer receives input that is both relevant and appropriate for its context, the information distribution module (Inject) enhances the model's overall detection ability.

Channel attention mechanisms are essential for highlighting or diminishing specific channels according to their importance in representing the input data [36]. With the efficient channel attention (ECA) mechanism in particular, a computationally effective approach can be used to quantify its importance without a large computer cost. The input feature map Y with dimensions $H \times W \times C$. The channel descriptor is initially obtained through Global Average Pooling (GAP). It is expressed in Equation 6.

$$g_c = \frac{1}{H \times W} \sum_{a=1}^{H} \sum_{b=1}^{W} Y_{abc} \tag{6}$$

where, the width, height and total number of channels were represented as W, H, and C.

For the one-dimensional (1D) convolution, the adaptive kernel size was marked as $j(C)$, is then calculated using a basic linear relationship. It is demonstrated in Equation 7.

$$j(C) = \chi C + \alpha \tag{7}$$

where, the parameters of adaptive kernel size were denoted as χ and α.

After the 1D convolution using this kernel is complete, the channel weights are normalized using the softmax function. The mathematical representation for normalizing channel weights was explained in Equations 8 to 10.

$$w = Conv1D\big(g, j(E)\big) \tag{8}$$

$$\widehat{w}_e = \frac{exp(w_e)}{\sum_{e'=1}^{E} exp\left(w_{c'}\right)} \tag{9}$$

$$X_{abe} = \widehat{w}_e Y_{abe} \tag{10}$$

where, the output of the 1D convolution operation was represented as w. The convolution function was indicated as $Conv1D\big(g, j(E)\big)$. The normalized weights were denoted as $\widehat{w}_e$. The input characteristics or value connected within channel e and spatial positions (a, b) were indicated as X_{abe}.

Through this process, channels that are considered important are amplified while less important ones are suppressed. The structure of the ECA module is demonstrated in Figure 3.

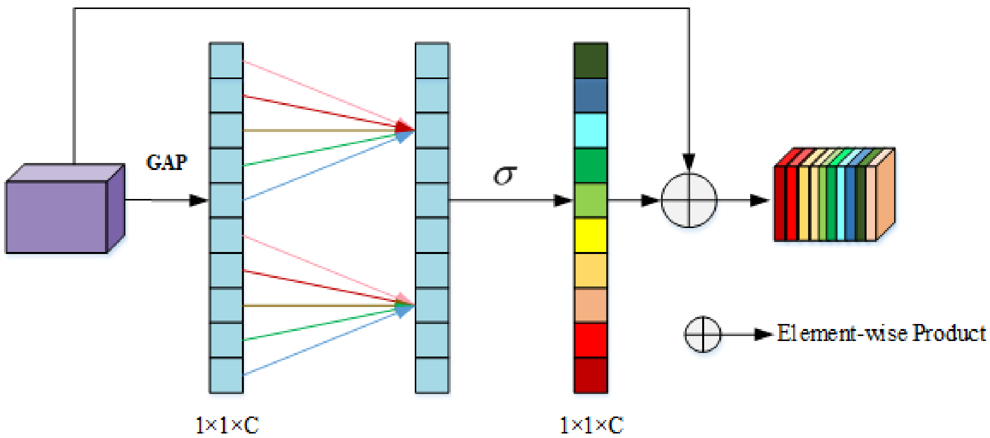

Figure 3. Structure of ECA Module

Global average pooling is applied to the input feature map, resulting in a matrix shift from $[h, w, c]$ to, as seen in Figure 4. Finally, an important element of ECA is the adaptive re-calibration of feature responses on a channel-by-channel basis. Because it focuses on inter-channel connections rather than the computational burden of traditional attention approaches, ECA is able to successfully highlight significant features while suppressing less important ones. The mid-layer architecture of the YOLOv8 series utilizes the conventional FPN structure, which has many branches for multi-scale feature fusion. The structural architecture of traditional FPN is demonstrated in Figure 4.

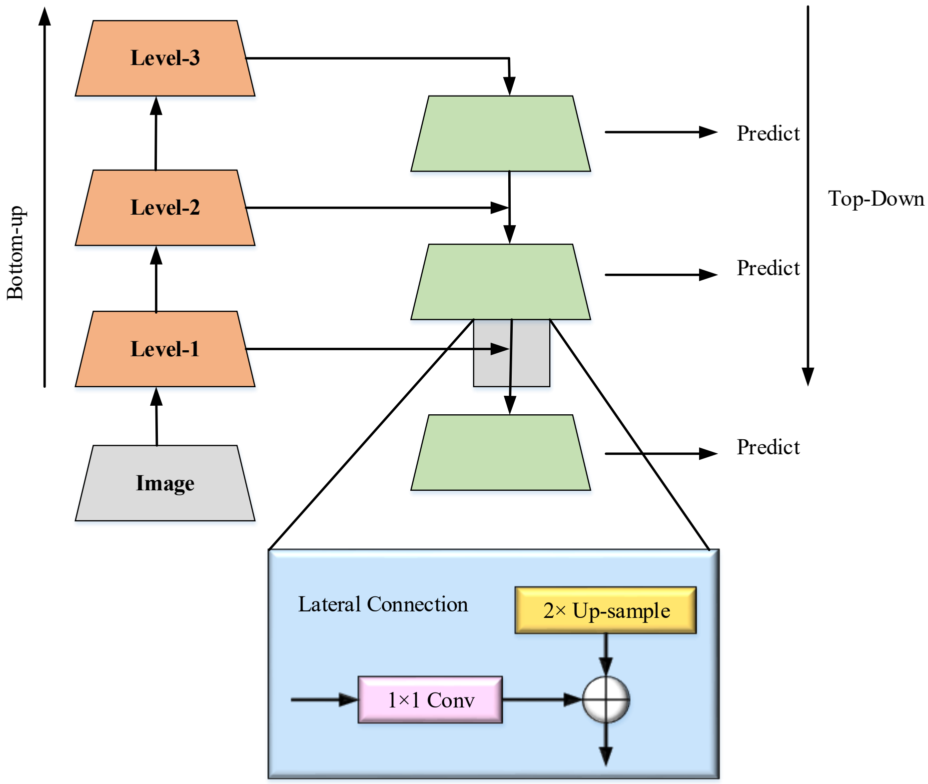

Figure 4. Structural Architecture of FPN

Through a top-down approach and lateral connections, the FPN is capable of combining low-resolution and semantically strong characteristics with high resolutions. It is especially useful in situations where target sizes vary greatly because each level of the pyramid may be utilized to identify objects at various scales. In a conventional FPN architecture, information from nearby levels is directly combined, allowing features to be integrated progressively across layers. This technique indirectly enables the network to access data from distant layers through repeated connections within adjacent levels, creating a flow of information similar to recursive. The fusion within current levels 1, 2, and 3 was generated more easily by FPN, which is ordered from top to bottom. Within the conventional FPN framework, this transfer mechanism can lead to a substantial loss of information.

By essentially reducing the model's size and parameter count, channel pruning lowers the computational workload. The β co-efficient of batch normalization (BN) was utilized in channel pruning procedures to evaluate every channel's contribution score. The computing process in the BN layer is illustrated in Equations 11 to 14.

$$\pi = \frac{1}{n}\sum_{a=1}^{n} x_a \tag{11}$$

$$\sigma^2 = \frac{1}{n}\sum_{a=1}^{n} x_a - \pi^2 \tag{12}$$

$$\hat{x}_a = \frac{x_a - \pi}{\sqrt{\sigma^2 + \chi}} \tag{13}$$

$$y_a = \beta\hat{x}_a + \alpha \equiv BN_\beta x_a \tag{14}$$

where, the mean and variance of input values$x_a$were represented as π and σ^2. The normalized value of input was denoted as $\hat{x}_a$. The constant value added to variance was noted as χ. The learnable parameters were denoted as β and α. After BN, the output was represented as y_a.

The YOLO model's capacity to identify pests of varied sizes and shapes that may be seen at different scales in images of rice leaves is improved by the FPN structure, which enables the model to efficiently integrate data from many scales. In order to maintain high processing speed, effective attention mechanisms like ECA are added to YOLO models, which are intended for real-time pest recognition. By concentrating on the most pertinent channels, the channel attention method aids the model in differentiating pests from these backgrounds, lowering false positives and increasing accuracy. In practical applications, an FPN-based YOLO model with ECA support improves detection accuracy, preserves real-time performance, lowers false detections, and is appropriate for managing agricultural pests on rice fields.

4- Results and Discussion

In this study, an improved DL model was proposed to automatically identify and detect pests on rice leaves. The Python programming language is used to implement the task. IP_RicePests and IP102 Datasets are used in the implementation of the suggested approach. The system configuration is demonstrated in Table 1.

Table 1. System Configuration

Installed RAM	16.0 GB
Pen and Touch	No pen or Touch Input is available.
Type of System	x64-based process, 64-bit operating system,

4-1- Dataset Description

IP_RicePests Dataset: The IP_RicePests dataset is designed for machine learning model training and evaluation, specifically for rice leaf pest detection and classification. For timely pest control in rice cultivation, the dataset helps create automated methods for recognizing pests. It serves as a valuable resource for automating pest identification, aiding in timely pest control strategies crucial for rice evaluation. The dataset comprises multiple pest species that commonly infest rice crops, covering a broad spectrum of pest types that affect plant health. The dataset was gathered under different lighting conditions, such as natural vs. artificial light, shadow effects, and diverse backgrounds.

IP102 Dataset: Over 75,000 images from 102 groupings are comprised in the IP102 dataset. It displays a long-tailed distribution by nature. Furthermore, bounding boxes were added to 19,000 images in order to detect objects. According to IP102's hierarchical taxonomy, insect pests that mainly affect a single agricultural commodity are grouped together in a single upper-level category. Figures 5 and 6 shows the visualization results of two datasets. This dataset exhibits a long-tailed distribution, meaning that certain pest classes contain significantly more samples than others, which reflects natural occurrences of pests in agricultural settings.

4-2- Visual Comparison of Results of IP102 and IP_RicePests Datasets

Original **Pre-processed** **Detected Images**

Figure 5. Visualization of Results for IP102 Dataset

Original **Pre-Processed** **Detected Images**

Figure 6. Visualization of Results for IP_RicePests Dataset

4-3- Performance Evaluation of IP_RicePests Dataset

After reviewing Figure 7, it is clear that the adaptive optimizer used in this study performs better than the others, attaining the maximum mean average precision (mAP), Precision, and Recall. This implies that the proposed approach is more suited for this particular object identification problem than classic optimizers. Adam and AdamW rely on momentum and adaptive learning rate adjustments, whereas traditional SGD employs a fixed learning rate. Local minimum, saddle points, and gradient vanishing are only a few of the difficulties that any optimizer may run across in practical implementations.

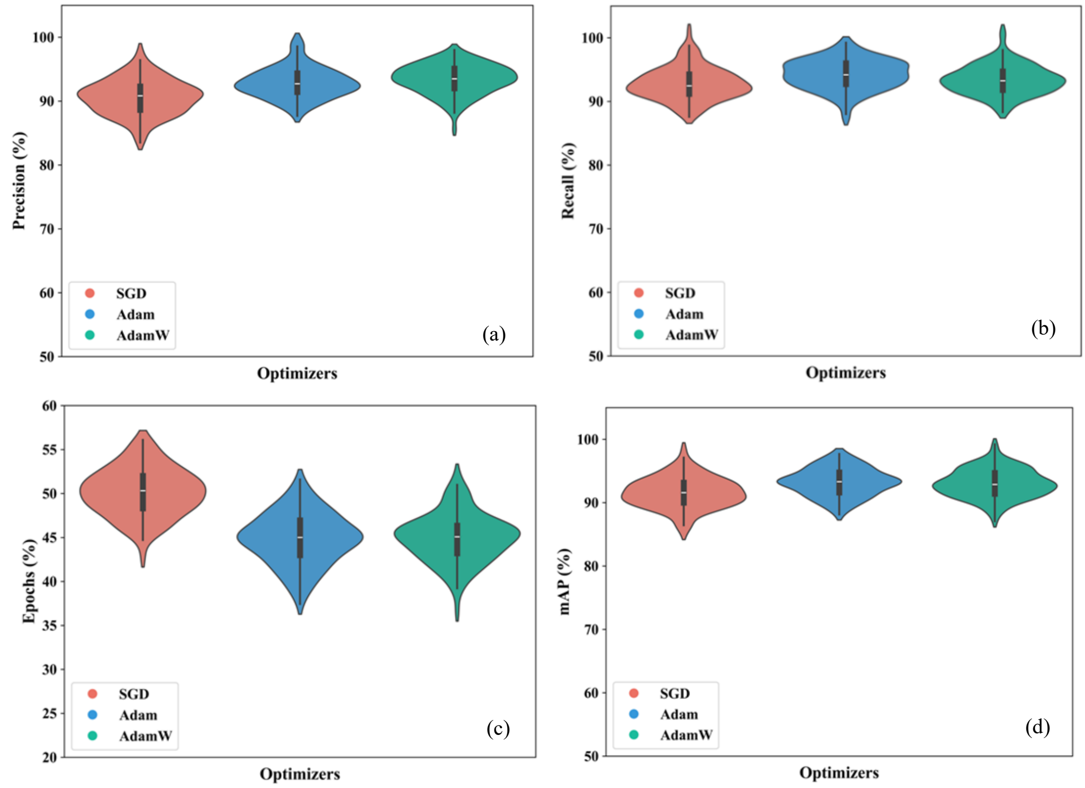

Figure 7. **Performance Analysis of different optimizers**

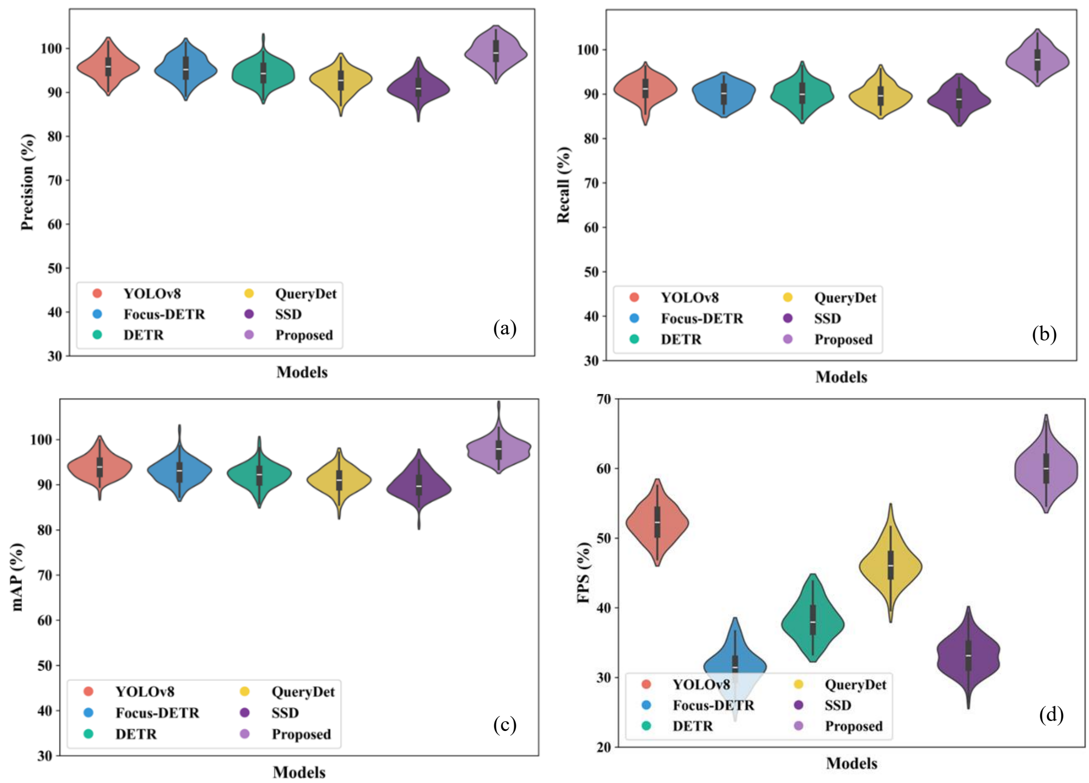

Figure 8. **Performance Analysis of different detection models**

The comparison of various detection techniques for both existing and proposed techniques was demonstrated in Figures 8a to 8d. The experimental findings demonstrated that the suggested model achieved improved recall, mAP, frames per second (FPS) and precision of 98.8%, 98.0%, 60 seconds and 97.9%, respectively. In Figure 8(a), the existing QueryDet achieved an average value of 93%. However, the existing model has over-fitting and interpretability issues. In Figure 8(c), the existing SSD and QueryDet achieved low mAP values of 90% and 91%. However, these existing models had high computational costs. The existing Focus-DETR and SSD attained low FPS scores of 31 and 33 sec in Figure 8(d). However, these existing technique's speeds are very slower and have high memory issues.

Figures 9(a) to 9(c) shows the comparison of different metrics like intersection over union (IoU), mAP and computation time. Compared to existing techniques, the proposed model attained higher values for all performance metrics. In Figure 9(a), the existing ResNet101 and VGG16 attained average IoU values of 85.3% and 92.5%, respectively. The existing DensNet121 and Inception attained high mAP values of 94.4% and 92.7% in Figure 9(b). However, these existing techniques were highly dependent on data. In Figure 9(c), the existing ResNet50 and ResNet101 achieved high computational times of 13.4 and 10.1 sec. Existing models, such as ResNet50, ResNet101, and DenseNet121, have lower IoU and mAP values due to inadequate feature extraction, which restricts object boundary identification. Furthermore, models such as VGG16 and Inception are prone to overfitting and exhibit a high dependency on data, which diminishes their ability to generalize effectively. SSD and QueryDet encounter difficulties in identifying small objects, which results in decreased accuracy and recall rates. Moreover, deep architectures like ResNet152 and DenseNet121 incur significant computational costs, rendering them less appropriate for real-time applications. The absence of sophisticated attention mechanisms further restricts feature prioritizing, resulting in suboptimal detection performance.

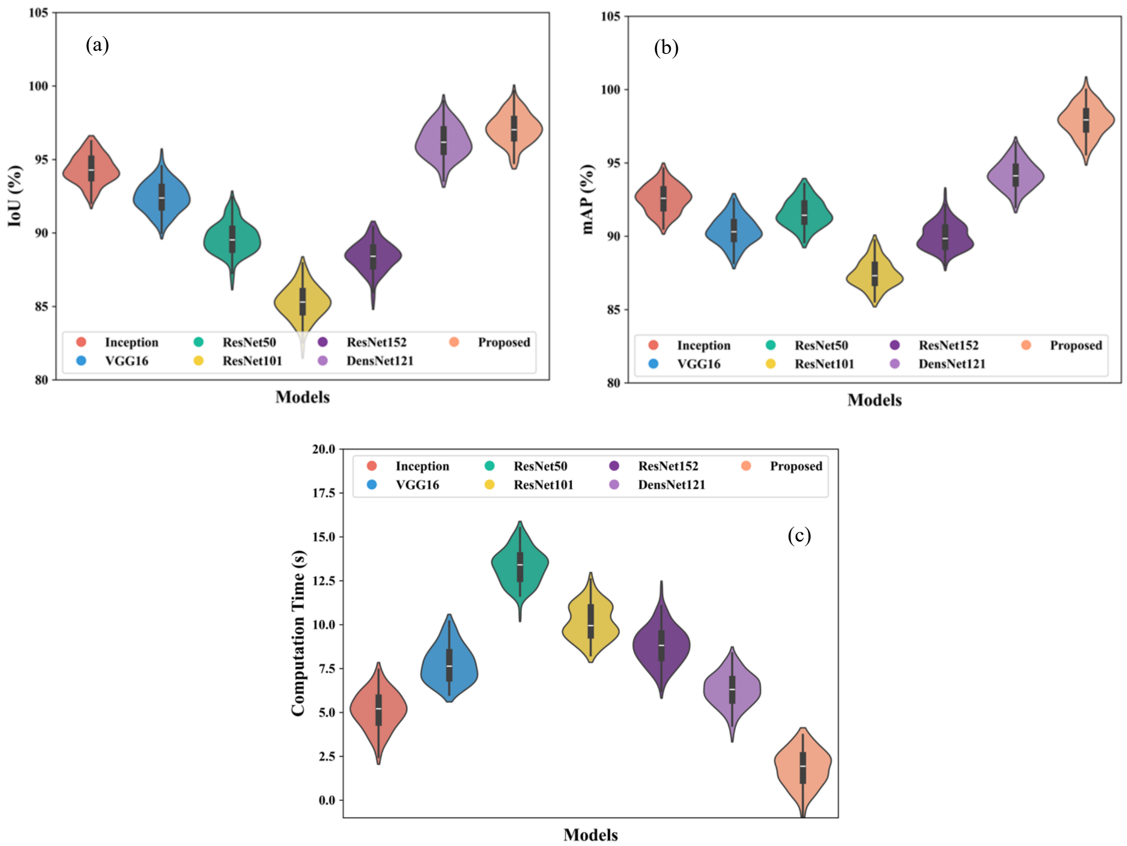

Figure 9. Comparison Analysis of existing and proposed techniques

The connection between time and performance measure or accuracy percentage is portrayed in Figure 10. It seems to contrast the convergence rate and long-term performance of two distinct models or approaches. The yellow curve points to a model that performs marginally better in the end but takes longer to stabilize. The blue curve shows a model that converges more quickly computationally but gives up some ultimate accuracy in the process. The training accuracy is represented in the blue curve, and the testing accuracy is shown in the yellow curve.

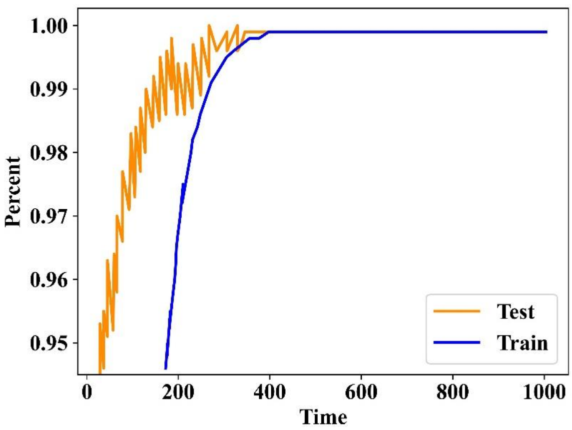

Figure 10. Training and Validation Accuracy Graph

Cross-Entropy Graph: The suggested model has a lower cross-entropy loss than the baseline. Figure 11 shows that it is superior in terms of both learning speed and ultimate performance. This highlights how the feature pyramidal network with the channel attention mechanism helps to increase the accuracy of pest identification. In this graph, the proposed model demonstrates a consistently lower cross-entropy loss compared to the baseline model YOLO, indicating that it learns the features more effectively. The rapid decline in loss during the initial training epochs signifies faster convergence, meaning the model quickly adapts to the dataset. The ultimate lower loss value demonstrates its capacity to generalize better and reduce misclassification errors.

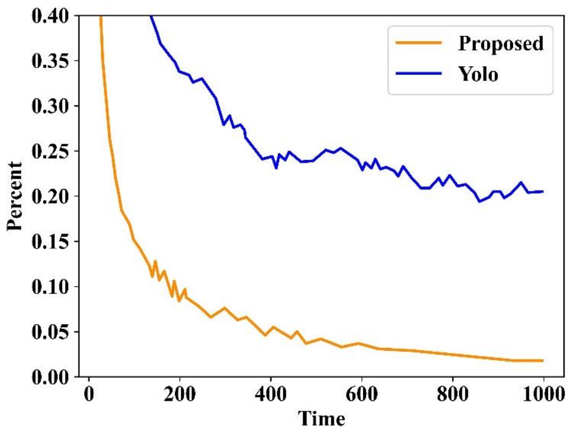

Figure 11. Cross-Entropy Graph

Current models such as YOLOv8, Focus-DETR, DETR, QueryDet, and SSD present certain limitations that adversely affect their accuracy, precision, recall, and F-score. Although YOLOv8 is known for its speed, it encounters difficulties in accurately detecting objects within cluttered environments, resulting in misclassifications. Despite their effectiveness, transformer-based attention mechanisms are a major component of Focus-DETR and DETR, which results in slower inference times and higher computational demands. QueryDet, while proficient in managing detection queries, is prone to overfitting and exhibits diminished generalization capabilities, which negatively impacts recall and precision. Due to its fixed grid-based methodology, SSD is a single-stage detector that sacrifices accuracy, making it less effective at detecting smaller objects. The proposed model addresses these issues by incorporating a feature pyramidal network alongside a channel attention mechanism, thereby enhancing feature extraction across various scales. By improving object localization and classification, this development successfully reduces false positives and negatives. Additionally, its optimized design maintains real-time performance while striking a balance between computational efficiency, leading to improved mAP, precision, recall, and F-score.

4-4- Performance Evaluation of IP102 Dataset

Figures 12(a) to 12(e) shows the performance analysis of the suggested and current approaches in terms of precision, recall, mAP, IoU, and calculation time. In Figure 12(a), the existing C3M-YOLO and SSD300 attained average precision values of 93.4% and 92.5% respectively. The existing YOLOv3-Tiny and C3M-YOLO attained low recall values of 86.4% and 78.8% in Figure 12(b). However, these existing techniques had a high computational load and required large datasets. In Figure 12(c), the existing YOLOv5s and YOLOv3-tiny achieved high mAP values of 93.0% and 86.9%.

But, these models are computationally expensive. The existing SSD300 attained a lower IoU value of 90.2% in Figure 12(d). The existing YOLOv5s attained a high computational time of 8.8 sec in Figure 12(e). Comparing to other existing techniques, the proposed method outperformed better, and it can detect pests accurately. The comparison of existing and proposed methods for the IP102 dataset is illustrated in Table 2.

Table 2. Performance Analysis of existing and proposed methods for IP102 dataset

Model	Precision (%)	mAP (%)	IoU (%)	Computation Time (s)
C3M-YOLO	93.4	78.8	89.4	6.8
YOLOv5s	94.3	89.3	85.8	8.8
SSD300	92.5	91.4	90.2	5.9
YOLOv3-tiny	94.1	86.9	93.8	4.6
Proposed	**97.8**	**95.9**	**97**	**2.5**

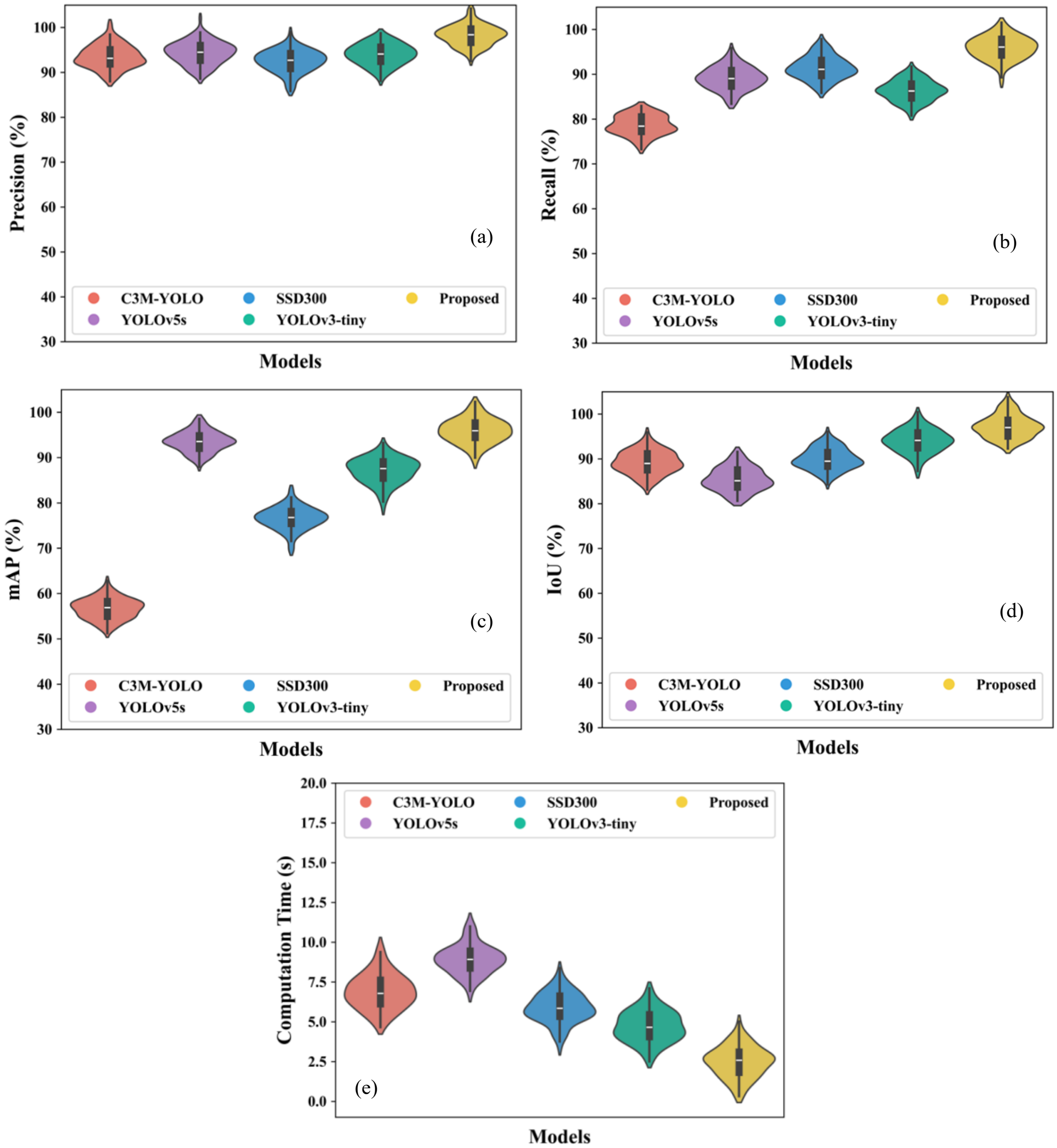

Figure 12. Performance analysis of precision, recall, mAP, IoU and computation time

Existing models, including C3M-YOLO, YOLOv5s, SSD300, and YOLOv3-tiny, exhibit lower mAP, IoU, precision, and recall, alongside increased computational times. This is primarily due to their insufficient feature extraction capabilities and inadequate multi-scale representation. These models face challenges in detecting small, overlapping, and intricate objects, which results in diminished IoU and lower precision-recall metrics. The high computational demands of models such as YOLOv5s and C3M-YOLO lead to extended processing durations, adversely impacting their real-time performance. Furthermore, the lack of attention processes and feature pyramid networks limits their ability to focus on relevant object regions. The proposed model overcomes these limitations by combining feature pyramidal networks and channel attention, resulting in improved localization, detection accuracy, and computing efficiency. It delivers higher mAP, IoU, precision, and recall while also reducing computational time, making it more robust and efficient.

The outcomes of the model's training process were also recorded and stored in a file. Tensorboard documentation of the training's accuracy scalar graph is displayed in Figure 13. The validation accuracy is shown by the blue-coloured line, while the training accuracy is represented by the orange-coloured line. After 400 training cycles, the model had already achieved 100% accuracy for both the training and validation data. This suggests that neither overfitting nor underfitting is present in the trained model.

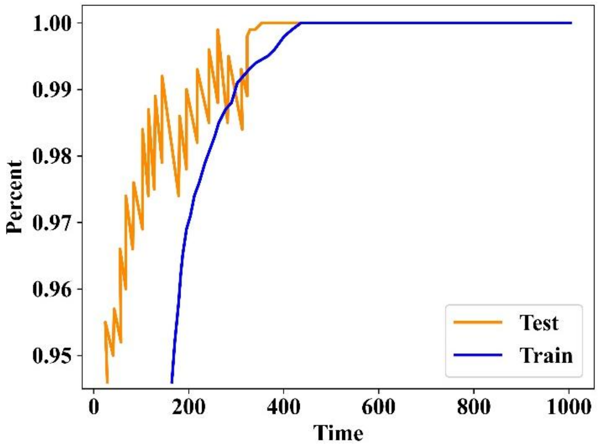

Figure 13. **Training and Validation Accuracy Graph**

Cross-Entropy Graph: The training cross-entropy graph is shown in Figure 14. The variance between probability distributions is measured using cross-entropy. It calculates the deviation of the prediction from the actual distribution. The yellow line displays the cross entropy of the training data, whereas the blue line represents the validation data. The comparison of different parameters and memory usage is shown in Table 3. Table 4 depicts the performance analysis for various pre-processing techniques.

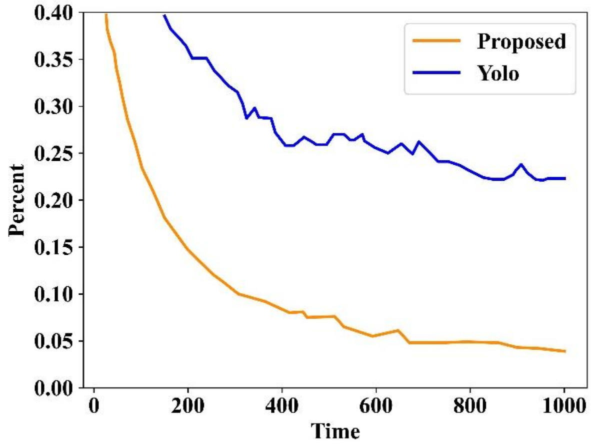

Figure 14. **Training cross-entropy graph for existing and proposed**

Table 3. **Comparison of various parameters**

Model	**Memory (MB)**	**Parameters (kB)**
FPN	45	224.5
TOOD	32.2	159.8
SSD300	37.2	185.6
YOLOv3-tiny	17.9	890
YOLOv5	14.5	728.5
Proposed	4	**100**

Table 4. Performance analysis of various pre-processing models

Filtering models	Noise reduction	SSIM	PSNR
Gaussian filter	Moderate	0.82	28.5
Median filter	Low	0.78	27.6
Bilateral filter	Low	0.87	30.12
Up-weGaf	High	0.91	34.56

The proposed IMpc-PyrYOLO model achieves a computation time of 2.5 seconds on the IP102 dataset, significantly outperforming existing models such as YOLOv3-tiny (4.6s), SSD300 (5.9s), YOLOv5s (8.8s), and C3M-YOLO (6.8s). This 60% reduction in processing time compared to YOLOv5s and 46% improvement over YOLOv3-tiny demonstrates its efficiency. Additionally, the model maintains superior performance in 97.8% precision, 95.9% mAP, and 97% IoU despite using the lowest memory of 4MB and parameters of 100kB.

High PSNR and SSIM values indicate better noise suppression with minimum distortion. Up-weGaf is selected for pre-processing due to its superior ability to balance noise reduction and feature preservation, which is crucial for accurate pest detection. Unlike traditional Gaussian and median filters that can blur fine details or distort texture, the Up-weGaf approach improves image clarity while retaining critical edge information, which helps to enhance the model's ability to differentiate pest features. Compared to computationally expensive methods such as non-local means (NLM) and wavelet denoising, the filter utilized in the proposed approach provides efficient noise suppression with minimal processing overhead, making real-time agricultural applications more effective. Table 5 represents a comparative analysis of some other existing models with the proposed approach [37] for the IP_Rice pest dataset. Table 6 represents the comparative analysis of the existing and proposed model for the IP_102 dataset [37]. Table 7 represents the ablation study of the proposed model.

Table 5. Comparative analysis of existing and proposed for IP_Rice pest dataset

Models	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
ResNet 50	83.41	82.95	83.02	83.17
VGG16	84.70	83.69	84.09	84.39
MobileNet	83.19	82.43	82.62	82.89
Ensemble	83.41	82.95	83.02	83.17
Proposed	**97.9**	**98.8**	**98.0**	**98.23**

Table 6. Comparative analysis of existing and proposed for IP_102 dataset

Models	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Ensemble	65.83	64.85	65.10	65.35
Proposed	**97.8**	**96**	**98.12**	**98.56**

Table 7. Ablation study of the proposed model

Models/Dataset	IP_RicePests			IP_102		
Performances	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)
Baseline YOLO	90.4	91.2	90.8	96.32	95.12	94.87
YOLO+ESA	92.2	94.0	93.9	96.99	96.65	95.12
YOLO+FPN	95.4	94.7	94.6	97.21	97.0	95.62
Proposed	**98.2**	**95.8**	**96**	**98.56**	**97.8**	**96**

4-5- Discussion

In natural environments, identifying pests in rice leaves is essential for automated ecological monitoring systems and smart agriculture. Real-world situations involving plant leaves result in complex foreground fluctuations, loud background clutter, and an uneven distribution of classes. The experimental findings indicate that the proposed model significantly surpasses current detection models in metrics such as mAP, IoU, precision, recall, and computational efficiency. Existing models, including C3M-YOLO, YOLOv5s, SSD300, and YOLOv3-tiny, demonstrate deficiencies in object localization, suboptimal feature representation, and elevated computational demands, which adversely affect their accuracy and efficiency. In particular, models like ResNet50, ResNet101, and DenseNet121 are characterized by longer computational times, rendering them inadequate for real-time applications. The proposed model employs feature pyramidal networks combined with channel attention, enabling it to proficiently extract and enhance features across various scales. The mAP and IoU metrics reveal that the proposed model attains superior detection accuracy while imposing a reduced computational load. Furthermore, the cross-entropy graph illustrates a quicker convergence rate and lower loss values, signifying enhanced learning efficiency. These advancements render the proposed model particularly suitable for real-time and large-scale object detection tasks, effectively addressing the shortcomings of existing approaches. The proposed method IMpc-PyrYOLO can achieve higher precision and recall performances with the help of the transformer structure's ability to capture long-range dependence information. Despite the lack of non-maximal suppression and a manually constructed anchor, IMpc-PyrYOLO outperforms traditional networks such as YOLOv5 and YOLOv7. In order to better suit real-world application circumstances, the model's detection capabilities will be improved for tiny target pests in intricate backdrop environments in the future.

5- Conclusion

In this paper, a novel IMpc-PyrYOLO model is introduced with an FPN network-based YOLO framework enhanced by an ECA mechanism for pest detection in rice leaf images. The proposed model addressed the limitations of existing DL models by enhancing feature representation and reducing computational complexity. The proposed model is evaluated using various performance factors such as precision, recall, f1-score, computational time, IoU, and mAP. Experimental results demonstrate that IP_RicePests attains a precision of 95.8%, recall of 96% and F1-score of 95.9%, whereas IP102 attains a higher precision of 97.8%, recall of 96%, mean Average Precision (mAP) of 95.9% and Intersection over Union (IoU) of 97% with processing time of 2.5 seconds. These results indicate that the proposed model significantly outperforms the conventional approach in both efficiency and computational time. Furthermore, the integration of attention processes improved the model's capacity to distinguish between visually identical pest species, contributing to better detection performance. The model's real-time efficiency makes it a promising solution for smart agricultural applications that require real-time monitoring and early intervention to avoid pest outbreaks. Despite being optimized, the suggested model has certain drawbacks, such as a complexity that makes it difficult to deploy on less sophisticated devices in agricultural environments. The dependence solely on RGB imaging restricts detection capabilities in difficult situations, indicating that multimodal fusion techniques, such as thermal, hyperspectral, or acoustic sensing, may improve detection reliability. Future research will prioritize model compression, transfer learning, and transformer-based architectures to enhance both efficiency and accuracy. Additionally, the integration of IoT-based cloud platforms for extensive deployment and real-time pest monitoring will be investigated. These innovations will aid in the creation of an intelligent pest detection system, promoting precision agriculture and sustainable crop protection methods.

6- Declarations

6-1- Author Contributions

Conceptualization, S.R. and S.Y.; methodology, S.R.; software, S.R.; validation, S.R., S.Y., and S.F.A.R.; formal analysis S.R., S.Y., and S.S.; investigation, S.R.; resources, S.R.; data curation, S.R.; writing—original draft preparation, S.Y. and S.S.; writing—review and editing, S.R., S.Y., and S.F.A.R.; visualization, S.R., S.Y., and S.S.; supervision, S.Y. and S.F.A.R.; project administration, S.Y. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available in the article.

6-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Ali, M. M., Bachik, N. A., Muhadi, N. 'Atirah, Tuan Yusof, T. N., & Gomes, C. (2019). Non-destructive techniques of detecting plant diseases: A review. Physiological and Molecular Plant Pathology, 108, 101426. doi:10.1016/j.pmpp.2019.101426.

[2] Bari, B. S., Islam, M. N., Rashid, M., Hasan, M. J., Razman, M. A. M., Musa, R. M., Nasir, A. F. A., & Majeed, A. P. P. A. (2021). A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework. PeerJ Computer Science, 7, 432. doi:10.7717/PEERJ-CS.432.

[3] Lin, F., Guo, S., Tan, C., Zhou, X., & Zhang, D. (2020). Identification of rice sheath blight through spectral responses using hyperspectral images. Sensors (Switzerland), 20(21), 1–16. doi:10.3390/s20216243.

[4] Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., Wang, F., Zhou, M., & Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. Sensors (Switzerland), 20(3), 578. doi:10.3390/s20030578.

[5] Khaki, S., Pham, H., Han, Y., Kuhl, A., Kent, W., & Wang, L. (2021). DeepCorn: A semi-supervised deep learning method for high-throughput image-based corn kernel counting and yield estimation. Knowledge-Based Systems, 218, 106874. doi:10.1016/j.knosys.2021.106874.

[6] Hu, Y., Deng, X., Lan, Y., Chen, X., Long, Y., & Liu, C. (2023). Detection of Rice Pests Based on Self-Attention Mechanism and Multi-Scale Feature Fusion. Insects, 14(3), 280. doi:10.3390/insects14030280.

[7] Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors (Switzerland), 17(9). doi:10.3390/s17092022.

[8] Wang, G., Sun, Y., & Wang, J. (2017). Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. Computational Intelligence and Neuroscience, 2017(1), 2917536. doi:10.1155/2017/2917536.

[9] Dong, Q., Sun, L., Han, T., Cai, M., & Gao, C. (2024). PestLite: A Novel YOLO-Based Deep Learning Technique for Crop Pest Detection. Agriculture (Switzerland), 14(2), 228. doi:10.3390/agriculture14020228.

[10] Li, W., Zhu, T., Li, X., Dong, J., & Liu, J. (2022). Recommending Advanced Deep Learning Models for Efficient Insect Pest Detection. Agriculture (Switzerland), 12(7), 1065. doi:10.3390/agriculture12071065.

[11] Bhagawati, R., Bhagawati, K., Singh, A. K. K., Nongthombam, R., Sarmah, R., & Bhagawati, G. (2015). Artificial neural network assisted weather-based plant disease forecasting system. International Journal on Recent and Innovation Trends in Computing and Communication, 3(6), 4168–4173.

[12] Jia, X., Jiang, X., Li, Z., Mu, J., Wang, Y., & Niu, Y. (2023). Application of Deep Learning in Image Recognition of Citrus Pests. Agriculture (Switzerland), 13(5), 1023. doi:10.3390/agriculture13051023.

[13] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. doi:10.1016/j.compag.2018.01.009.

[14] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. doi:10.1109/TPAMI.2016.2577031.

[15] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 779–788. doi:10.1109/CVPR.2016.91.

[16] Yang, S., Xing, Z., Wang, H., Dong, X., Gao, X., Liu, Z., Zhang, X., Li, S., & Zhao, Y. (2023). Maize-YOLO: A New High-Precision and Real-Time Method for Maize Pest Detection. Insects, 14(3), 278. doi:10.3390/insects14030278.

[17] Zhang, M., Yang, W., Chen, D., Fu, C., & Wei, F. (2024). AM-MSFF: A Pest Recognition Network Based on Attention Mechanism and Multi-Scale Feature Fusion. Entropy, 26(5), 431. doi:10.3390/e26050431.

[18] Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. International Interdisciplinary PhD Workshop, IIPhDW 2018, 117–122. doi:10.1109/IIPHDW.2018.8388338.

[19] Taylor, L., & Nitschke, G. (2018). Improving Deep Learning with Generic Data Augmentation. Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018, 1542–1547. doi:10.1109/SSCI.2018.8628742.

[20] Nazki, H., Lee, J., Yoon, S., & Park, D. S. (2018). Synthetic data augmentation for plant disease image generation using GAN. In Proceedings of the Korea Contents Association Conference, 459-460.

[21] 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. doi:10.1109/ACCESS.2022.3189676.

[22] Li, S., Wang, H., Zhang, C., & Liu, J. (2022). A Self-Attention Feature Fusion Model for Rice Pest Detection. IEEE Access, 10, 84063–84077. doi:10.1109/ACCESS.2022.3194925.

[23] Mendigoria, C. H., Concepcion, R., Bandala, A., Alajas, O. J., Aquino, H., & Dadios, E. (2021). OryzaNet: Leaf Quality Assessment of Oryza sativa Using Hybrid Machine Learning and Deep Neural Network. 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2021, 1–6. doi:10.1109/HNICEM54116.2021.9731957.

[24] Ali, F., Qayyum, H., & Iqbal, M. J. (2023). Faster-PestNet: A Lightweight Deep Learning Framework for Crop Pest Detection and Classification. IEEE Access, 11, 3317506. IEEE Access. doi:10.1109/ACCESS.2023.3317506.

[25] Kukreja, V., Sharma, R., Vats, S., & Manwal, M. (2023). DeepLeaf: Revolutionizing Rice Disease Detection and Classification using Convolutional Neural Networks and Random Forest Hybrid Model. 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023, 1–6. doi:10.1109/ICCCNT56998.2023.10306530.

[26] Bharanidharan, N., Sannasi Chakravarthy, S. R., Rajaguru, H., Vinoth Kumar, V., Mahesh, T. R., & Guluwadi, S. (2023). Multiclass Paddy Disease Detection Using Filter-Based Feature Transformation Technique. IEEE Access, 11, 3322587. doi:10.1109/ACCESS.2023.3322587.

[27] Ananyasreya, P. P., & Subhija, E. N. (2022). Paddy Pest Classification Using Deep Learning Based Features. Proceedings - 2nd International Conference on Next Generation Intelligent Systems, ICNGIS 2022, 1–5. doi:10.1109/ICNGIS54955.2022.10079825.

[28] Tenriola, A., Azis, P. A., Kaswar, A. B., Adiba, F., & Andayani, D. D. (2025). Identifying Rice Plant Damage Due to Pest Attacks Using Convolutional Neural Networks. Jurnal RESTI, 9(1), 10–19. doi:10.29207/resti.v9i1.6125.

[29] Chaurasia, H., Arora, A., Raju, D., Marwaha, S., Chinnusamy, V., Jain, R., & Haque, M. A. (2024). PanicleDet: a deep learning-based model for detection of panicle stages in paddy. Neural Computing and Applications, 37(4), 2663–2673. doi:10.1007/s00521-024-10746-6.

[30] Tan, S., Xie, Q., Zhu, W., Deng, Y., Zhu, L., Yu, X., Yuan, Z., & Chen, Y. (2025). Deep learning and hyperspectral features for seedling stage identification of barnyard grass in paddy field. Frontiers in Plant Science, 16, 1507442. doi:10.3389/fpls.2025.1507442.

[31] Hussain, A., & Balaji Srikaanth, P. (2024). Leveraging Deep Learning and Farmland Fertility Algorithm for Automated Rice Pest Detection and Classification Model. KSII Transactions on Internet and Information Systems, 18(4), 959–979. doi:10.3837/tiis.2024.04.008.

[32] Song, H., Yan, Y., Deng, S., Jian, C., & Xiong, J. (2024). Innovative lightweight deep learning architecture for enhanced rice pest identification. Physica Scripta, 99(9), 96007. doi:10.1088/1402-4896/ad69d5.

[33] Suryanarayana, G., Chandran, K., Khalaf, O. I., Alotaibi, Y., Alsufyani, A., & Alghamdi, S. A. (2021). Accurate Magnetic Resonance Image Super-Resolution Using Deep Networks and Gaussian Filtering in the Stationary Wavelet Domain. IEEE Access, 9, 71406–71417. doi:10.1109/ACCESS.2021.3077611.

[34] Wu, F., Yang, W., Xiao, L., & Zhu, J. (2020). Adaptive wiener filter and natural noise to eliminate adversarial perturbation. Electronics (Switzerland), 9(10), 1–14. doi:10.3390/electronics9101634.

[35] Cheng, L., Li, J., Duan, P., & Wang, M. (2021). A small attentional YOLO model for landslide detection from satellite remote sensing images. Landslides, 18(8), 2751–2765. doi:10.1007/s10346-021-01694-6.

[36] Zhou, T., Canu, S., & Ruan, S. (2021). Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. International Journal of Imaging Systems and Technology, 31(1), 16–27. doi:10.1002/ima.22527.

[37] Li, Z., Jiang, X., Jia, X., Duan, X., Wang, Y., & Mu, J. (2022). Classification Method of Significant Rice Pests Based on Deep Learning. Agronomy, 12(9), 2096. doi:10.3390/agronomy12092096.