SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I5P106

Volume 13 Issue 5, 62-77, May 2026
© 2026 Seventh Sense Research Group®

Original Article

Jetson Nano-Powered Smart Farm Prototype for Real-Time Pest Classification Using a Vision Transformer–SVM Hybrid Model

Sanjyot Thuse[1], Meena Chavan[2]

[1]*Department of Electronics and Communication, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India.*
[2]*Department of Electronics and Telecommunication, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India.*

[1]*Corresponding Author : mschavan@bvucoep.edu.in*

Received: 11 February 2026 Revised: 09 March 2026 Accepted: 08 April 2026 Published: 30 May 2026

Abstract - *Agriculture is a key source of income for a large section of the world's population. It has become a necessity to make prompt and accurate pest detection for overall crop protection along with sustainable farming. Several current research projects rely on CNN architectures or cloud-based systems, which may not be suitable for agricultural fields with limited resources. To address this gap, the paper proposes a novel hybrid ViT-SVM model for real-time pest identification. This combines Vision Transformer, having global feature extraction capabilities, with Support Vector Machine, which has robust decision limits. Extensive tests were undertaken to evaluate several feature extractor-classifier mixtures. These combinations include ResNet, DenseNet, and ViT, combined using both SVM and Random Forest classifiers. To guarantee statistical reliability, F1-score, recall, accuracy, and precision were the metrics used to assess performance. Macro, weighted averages, class-wise ROC-AUC, PR-AUC, confusion matrices, McNemar tests, as well as bootstrap confidence intervals were calculated to test the performance of the models. The ViT-SVM hybrid system achieved superior performance to all other systems because it reached 96% accuracy and achieved an F1-score while maintaining equal success rates in detecting visually similar yet less common pest species. The system was developed as a model that ran on a Jetson Nano human-operated robotic system known as AgroPestBot, which used an Android interface for real-time pest identification at field sites. The method delivers an effective edge-AI solution that combines advanced machine learning capabilities with agricultural technology that can be used in the field to deliver farmers instant, precise pest detection results. The study demonstrates how lightweight hybrid models can be applied to resource-limited environments while establishing base technologies for upcoming automated systems and improved, precise agricultural methods.*

Keywords - *Deep Learning, Jetson Nano, Pest Classification, Support Vector Machine (SVM), Vision Transformer (VIT).*

1. Introduction

Agriculture has an important role in a country's economy, especially in nations such as India, where a large portion of the people rely on it for a living and survival. The agricultural sector functions as a vital economic driver that contributes to social stability while creating employment opportunities and ensuring food security for rural areas. Furthermore, it contributes significantly to the GDP. India contributes roughly 16% to the gross domestic product (Gross Domestic Product) [1]. Agriculture and related sectors employ more than 58% economic India's workforce, making them the country's primary source of employment [2, 3].

Agriculture is essential for food security because it provides people with the necessary nourishment [4]. Pest infestations are a substantial threat to food security due to the fact that they cause significant crop losses and flavour deterioration. The issue has been intensified by the increased dependence on chemical-based pesticides, which has resulted in insect resistance. Ineffective pest control accounts for more than 40% of crop losses globally, highlighting the critical need for long-term solutions [5]. Pest infestations may significantly impair agricultural production, particularly for key crops such as grains [6].

Pest infestations can impact greenhouse-grown vegetables and flowers. Large-scale producers depend on the sale of high-quality vegetables. Most florists make their living primarily from the sale of high-quality flowers. Pest-contaminated vegetables and flowers may impair their daily revenue. Farmers use pesticides to protect crops, including cereals, vegetables, and flowers. Farmers are frequently

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

unaware of the effects pesticides have on growing crops, vegetables, and flower plants. Farmers are also not aware of the correct pesticide that can keep crops, vegetables, and flowers safe from pests [37]. Multiple pesticide types are sometimes employed, jeopardising the health of farmers and, consequently, consumers. Farmers are constantly exposed to pesticides, which may cause both short- and long-lasting conditions such as respiratory problems, Alzheimer's, Parkinson's, skin and eye problems, cancer, etc. [7, 8] Though consumers are usually less exposed to pesticides than farmers, issues related to health may still occur [9].

Farmers need education about specific pest names that damage their crop fields, vegetable fields, and flower fields because this knowledge enables them to properly describe their problems to pesticide suppliers and select appropriate pesticides for their needs. The farmers who live in rural areas face challenges because they cannot identify the name of the pest. Therefore, there is a need to develop a system that uses advanced technology to help farmers identify the names of pests.

The existing deep learning and machine learning frameworks built so far prove to give good accuracy results for pest classification, but are struggling with blurriness and noise in the image. Most of the studies focus on classification based on deep learning and machine learning. Very few studies focus on real-time, low-cost, and embedded field deployment using edge devices.

To address this gap, the study aims to develop a prototype named AgroPestBot using the Jetson Nano framework, which can efficiently identify the name of the Pest/insect that is causing damage to the crop so that farmers can approach the pesticide vendor and purchase the right pesticide. From the study, it is observed that for insects belonging to the same family, the same pesticide can be applied. This study proposes a revolutionary method for crop protection and productivity that uses cutting-edge technology, mobility, as well as a simple Android interface to provide accurate and rapid pest detection. Jetson Nano is used to deploy the model constructed using ViT and Support Vector Machine. At present, the system needs human intervention for navigation and operation, but it can be converted to full automation as a part of future work.

2. Related Work

Precision farming seeks to decrease crop damage and boost productivity; one crucial stage in this process is the identification or classification of agricultural pests. Machine Learning (ML) and Deep Learning (DL) are two examples of AI technologies that researchers have relied on more and more over the last decade to automate pest identification with remarkable accuracy. Researchers in this area examine the pros and cons of using deep learning algorithms and machine learning algorithms to detect agricultural pests, as well as their practical applications in the field. In order to get the greatest outcomes in terms of total pest classification, the study aims to develop a novel technique for a combined approach. Section 2.1 focuses on the classification of pests using ML and DL models.

3. Machine as well as Deep Learning Classifiers for Pest Classification

Different methods used to identify pests include feature extraction and classification. Section 2.1 begins with feature extraction techniques and ML models used for classification. This section extends with deep learning algorithms, which perform feature extraction as well as classification.

Traditional approaches use HOG and LBP as handcrafted descriptors to build their systems, which operate with classical machine learning models. In [24], the study assessed tomato pest identification through HOG and LBP features combined with SVM testing, which resulted in HOG achieving 97% accuracy. The model structure depends on manual feature creation, which results in a basic design but restricted capacity to handle different situations. The work in [25] studied GLCM, HOG, LBP, and SURF features, which were tested using SVM, KNN, and DT algorithms. The LBP SVM model achieved 81.02% accuracy, which demonstrates that traditional machine learning models maintain their performance when users develop high-quality features. The hybrid methods use a combination of segmentation and classification techniques to achieve better accuracy results. The spatial fuzzy C-means clustering method, which operates with MCSVM in [32], improves segmentation consistency but creates additional difficulties for system operation. The NDVI-based segmentation method in [33], which uses GLCM features together with SVM, achieved 97% accuracy in actual field testing, which demonstrates its adaptability but needs testing on different crop and pest types.

Deep learning systems automatically extract essential features that result in improved prediction accuracy. The hybrid model, which combines YOLOv3 with ResNet50 and VGG16, achieved 98% accuracy according to [10], but it cannot be used in real-time Internet of Things applications because of its high computational demands. EfficientNetB7 achieved 93.5% accuracy in [11], but its complex design prevents installation on embedded devices. The researchers achieved their detection improvement results through their work with DenseNet-121 because their current testing method does not provide numerical evaluation.[12] The study in document [13] tested nine different pre-trained CNN models, which achieved maximum accuracy rates of 99.7% to demonstrate the effectiveness of transfer learning. The modified ResNet50 system from [14] achieved 92% accuracy, but it encounters difficulties because objects show different postures and their movements are blocked. The attention-based CNN in [15] achieved 96.92% accuracy, yet upcoming vision transformer methods might exceed its performance. The current research study was focused on various deep learning algorithms that were developed to detect and identify insect pests automatically. The study results demonstrate that deep models achieve high accuracy but require extensive computational resources, while SVM models provide equal accuracy at reduced system demands, which makes them ideal for deployment in edge environments.

The paper presents classification study results through Table 1, which summarizes the findings of the algorithm. Classification of pests has recently been the focus of deep learning framework research. As per the survey, the number of pest detection systems has increased, with a number that are using optimisation and deep learning algorithms.

Various models for pest and disease categorization of plants are discussed in the literature review. These models range from deep learning architectures, which include CNN YOLOv3, ResNet, and machine learning approaches like SVM and Random Forest, which are standard machine learning methods. The main purpose of the literature survey in this paper is to find an efficient model that can perform well with edge devices, is lighter, and gives good accuracy. In Figure 1, we can see how different classifiers for pest categorisation stack up against one another. Although XceptionNet and ResNet-50 are among the most accurate deep learning models (up to 99.7 percent), they are also the most memory and GPU-intensive. On the other hand, Support Vector Machine (SVM) models generally provide better generalisation in noisy real-world settings, lower complexity, and robust performance overall. This is particularly true when paired with well-engineered methods for feature extraction, such as HOG, LBP, and NDVI segmentation, even though SVM-based models occasionally have slightly lower accuracy (97% in [24] and [33])-compared to models such as EfficientNetB7 or YOLO, the NDVI data + SVM and HOG + SVM methods obtained a 97% accuracy rate with a much-reduced computational burden. Another approach that successfully combines accuracy with deployability is the hybrid technique. Examples of such approaches include SURF/LBP with SVM and Fuzzy C-Means + SVM.

Table 1. Summary of classifier survey for pest categorization

Classification Techniques	Take away	Scope for improvement
HOG vs. LBP with SVM [24].	HOG surpasses LBP: 97% accuracy with handmade features.	Has no deep learning comparison and is limited to handmade features. System is not reliable since it depends on data.
GLCM, HOG, LBP, SURF with ML classifiers [25].	SVM with LBP has the greatest accuracy (81.02%) when examined across many measures.	Models based on hybrid or DL might potentially enhance performance.
Fuzzy C-Means + Multiclass SVM [32].	Spatial clustering increases segmentation, whereas hybrid models improve accuracy.	Complexity is high, and there is little confirmation on pest categorization.
NDVI Segmentation + SVM [33]	NDVI-based segmentation using GLCM features improves accuracy (97%).	Must be tested on a variety of crops and pests.
YOLOv3 + ResNet50 + VGG16 + AE-HHO [10]	High accuracy (98%) using IoT and a sophisticated hybrid feature pipeline.	Unfit for use in real-time due to excessive resource use and latency.
EfficientNetB7 [11]	Reached a level of pest detection accuracy of 93.5%	Rather cumbersome to implement on microcontrollers or edge devices
DenseNet121 + POA + MLP (AIDC-POADL) [12]	Displayed effective detection methods via optimization	Lack of quantitative measures; little opportunity to compare models
9 Pretrained DL models incl. ResNet, MobileNet [13]	Hit an accuracy level of 99.7 percent; top performers were ResNet-50 and XceptionNet.	For deployment to be feasible, energy and latency profiles are required.
Adversarial Attention Module [14]	Increased precision and recall by 5%; accuracy of 92%	Shape and occlusion sensitivity; potential for overfitting tests
CNN with attention mechanism [15]	Impressive 96.92% accuracy rate; outperforms most current techniques	It could be surpassed by more recent models that use ViT instead

Based on its performance, SVM serves as a very promising classifier. It is nearly as good as deeper designs. SVM may still do well even when its claimed accuracy is lower.

- SVM works effectively with small and medium-sized datasets, but deep models may overfit.
- SVM increases generalisation by identifying optimum decision boundaries with the greatest margins, particularly in noisy as well as overlapping feature fields.

- SVM can accurately categorise features such as HOG, LBP, or GLCM.
- Ideal for edge devices like Jetson Nano that need high real-time performance plus power economy.

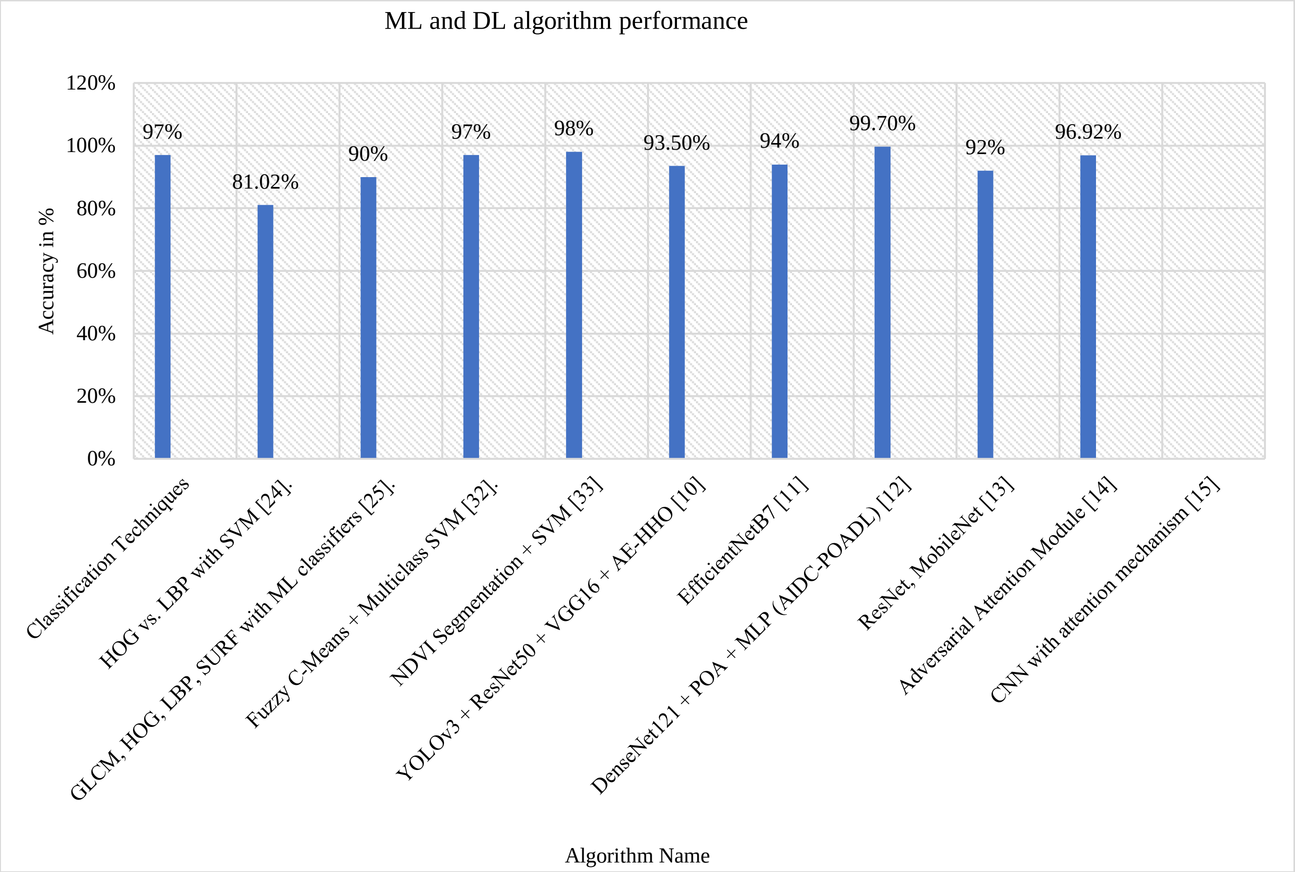


Fig. 1 Examining various classifier performance

4. Feature Extraction for Pest Classification

Section 2.1 in the paper concludes that though deep learning algorithms proved to perform better, they are still too heavy to deploy on edge devices like Jetson Nano. Thus, SVM, being lighter and performing well with a limited dataset, can be used as a classifier. Deep learning algorithms automatically extract important features from images, but are too heavy to be deployed on edge devices. Recent studies state that SVM is better in many cases. It is essential to identify a feature extraction technique that gives better results than the traditional feature extractors and generates strong and meaningful information. The relevance and quality of characteristics extracted from pest images are crucial for accurate pest categorisation in agricultural systems. To facilitate effective classification from raw visual data, feature extraction represents a crucial pre-processing step. Various pest features have been identified using modern deep learning algorithms in addition to form, colour, and texture-based descriptors. To better understand how different feature extraction strategies affect pest classification model performance, this section reviews the relevant research. Early pest classification studies relied heavily on hand-crafted feature descriptors such as BoVW, SIFT [26], ORB, LBP [27], SURF [28], GLCM [29], and HOG [30]. The methods can identify local texture and edge details, but their performance fails to extend beyond their intended operational boundaries because they cannot model whole spatial patterns, which makes them ineffective for distinguishing between pest species that look alike. Deep learning methods developed better solutions to this problem because they enabled systems to learn more advanced features, yet traditional CNNs still treat local patterns as their main focus, while they miss important differences between classes. Recent research investigated Vision Transformer–based systems because these systems provide superior capabilities to understand the complete image context. The GNViT model in [16] achieved 99.52% accuracy for groundnut pests, which demonstrates the Transformer's capabilities, yet their performance suffers from ImageNet pre-training limitations. The same results apply to SFA-ViT [17]. Employed scale, Pose, and factor attention mechanisms are offered an opportunity for enhancement of classification, but they exhibited perceptible sensitivity to small data samples, like IP102. A comparative study in [18] demonstrated that ViT achieved better results than CNN models, which included ResNet and AlexNet, when researchers used powerful data augmentation techniques. The hybrid method EViTA, which combined PCA and MFO [19], demonstrated better accuracy results, but the method still faced difficulties when operating on datasets of medium size. The research in article [20] tested ViT, which used 34000 images and showed better results than ResNet, yet the system still experiences operational issues during real-world applications. The study presented in this paper intends to solve the problem through a lightweight ViT–SVM hybrid system, which operates on a Jetson Nano-based mobile platform. The development of Vision Transformers as reliable feature extractors for pest categorization stems from their self-attention technology.

The networks acquire complete visual data from all images in the dataset. The visual presentation of agricultural pests improved through ViTs because these systems actively choose their concentration areas, while standard deep learning models employ static filters to detect only nearby patterns. The ability of pre-trained ViTs to work effectively with small or unbalanced datasets presents an advantage over traditional deep learning methods. Pest detection and other fine-grained classification problems are perfect fits for the ability to draw out a wealth of high-level, semantically rich data. Combining ViTs without lightweight machine learning in general classifiers like SVM and random forest allows for efficient and accurate classification, even on devices with limited computing power.

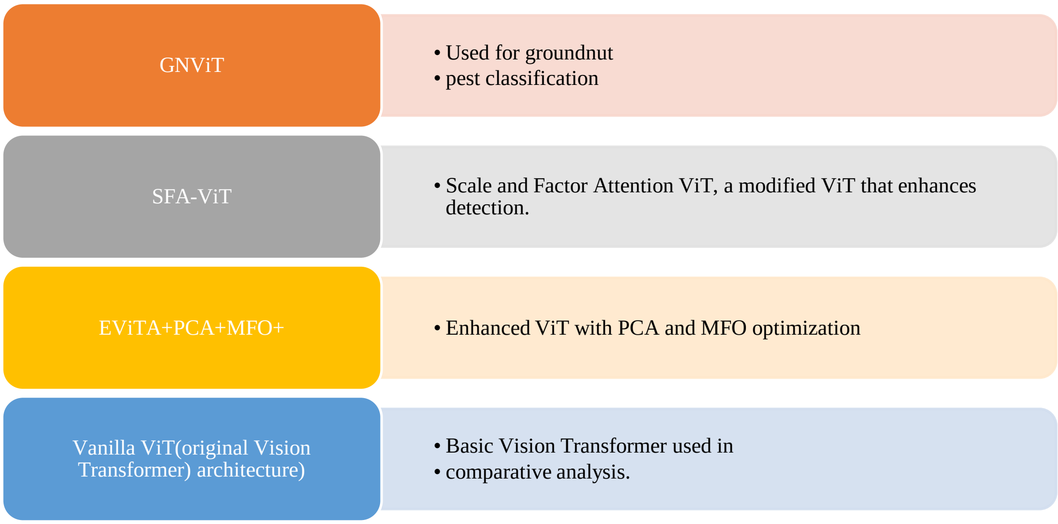


Fig. 2 Vision transform variants used for pest classification

The researchers' ViT modifications for pest categorisation are shown in Figure 2. Several modified Vision Transformer (ViT) design concepts have demonstrated enhanced pest detection performance using attention mechanisms or optimised techniques; however, these pipelines are usually complex and resource-intensive, which may make them unsuitable for real-time field deployment from edge devices. The mentioned architectural designs, GNVIT SFA-ViT and EViTA, serve as examples. The study integrates Support Vector Machine together with Vanilla ViT-Base model to create a system that extracts features efficiently while maintaining generalization capabilities. The previous studies in this area have examined different feature extractors and classifiers, which include the traditional HOG and LBP descriptors together with CNN models and modern Transformer architectures. The existing methods require high-performance computing systems for operation because they do not support agricultural field applications, which need real-time processing. The majority of models depend on deep networks or traditional techniques, yet only a few studies examine hybrid methods that achieve precise results while maintaining low resource consumption on embedded systems.

The primary accomplishment of this study was to develop a hybrid ViT-SVM model, which operates successfully on agricultural research embedded systems. The new method combines Vision Transformers' global feature learning capabilities with SVMs' precise decision boundary creation to achieve accurate pest detection on resource-constrained devices.

The project creates an operational prototype which uses a NVIDIA Jetson Nano to link the trained model with a mobile system and Android interface for all-time pest detection at actual field sites. The entire system shows how the proposed method operates in real-world conditions, which extend beyond its testing through algorithm experiments. The next section presents the process used to create the model, train it, and connect it to the overall system.

5. Methodology

The existing methods which rely on expert knowledge fail to deliver effective results for farmers because the existing methods which rely on expert knowledge fail to deliver effective results. Thus, farmers face difficulties with accurate and timely pest detection, which causes their crops to suffer major losses. An accurate, user-friendly, and field-portable pest categorization system is necessary. The research presents a novel embedded solution that combines NVIDIA Jetson Nano hardware with Vision Transformer (ViT) and Support Vector Machine (SVM) to enable real-time pest detection through automatic image analysis, which will assist farmers with their pest management activities. The suggested setup will accomplish multiple purposes, which have been defined as objectives.

- To implement a hybrid model combining Vision Transformer (ViT) for feature extraction and Support Vector Machine (SVM) for classification.
- To deploy the model on an embedded platform (NVIDIA Jetson Nano) for real-time pest identification in field conditions.
- To capture high-quality pest images using a camera integrated with the embedded system.
- To provide accurate and timely classification results to assist farmers in pest management.
- To ensure the system is efficient, portable, and cost-effective for practical agricultural use.

5.1. Dataset Description

The proposed method is evaluated using a pest image dataset containing ten distinct agricultural pest categories. The dataset contains images that show different types of pests, which include aphids, armyworm, ballworm, beetle, grasshopper, mites, red spider, sawfly, stem borer, and whitefly. The number of samples across different classes results in a situation where the classes show slight imbalance. All images are resized to 224 × 224 pixels and used with random rotation, horizontal flips, and scaling as basic augmentation methods to improve generalization while reducing overfitting. The dataset provides appropriate conditions for testing different models because it contains their ability to identify visually matching pests. The study used 3,742 training images and 1,323 testing images, while the remaining images served as validation to create a fixed 70:15:15 distribution for learning purposes, to assess model performance without bias, and to optimize model parameters. This dataset is more available for discriminative feature learning, so that an appropriate signaling strategy would be beneficial.

5.2. Evaluation Metrics

The evaluation process required multiple performance measures to establish an accurate and unbiased assessment. The system's overall reliability assessment uses accuracy measurement, whereas precision, recall, and F1-score metrics evaluate the system's ability to identify minority pest classes throughout the unbalanced dataset. Confusion matrices are analyzed to understand misclassification trends between visually similar pests. The classifier decision boundary analysis uses AUC-ROC curves to evaluate class separability. The combination of these metrics provides a complete performance assessment that exceeds basic accuracy assessment.

5.3. Experiment Setup

The tests needed to use deep learning models for efficient processing because these tests required execution on GPU-based environments. The proposed system uses a modern GPU that has 12 GB of memory and 16 GB of RAM to handle large feature extraction processes. The software environment used Python 3.12 with PyTorch, Torchvision, Transformers, and scikit-learn to build deep learning models and traditional machine learning classifiers. The deployment prototype was created using an NVIDIA Jetson Nano (4 GB RAM), which served as the main training environment. The framework uses Jetson Nano to identify pests during actual field testing according to [21-23]. The device functioned as a testing tool to determine whether the trained model could operate successfully in a real-time environment. The prototype currently functions through manual operation because users need to capture pest images before the system can process them for prediction purposes. The authors will investigate automation and real-time imaging system integration as part of their upcoming system enhancements.

The research study divided its dataset into three parts, which included 70% for training purposes, and 15% for testing, and 15% for validation testing to ensure proper data availability for model training while maintaining adequate data for unbiased testing. The study used three deep learning models, which included Vision Transformer (ViT-Base), ResNet-50, and DenseNet-121 as their pretrained deep learning architectures for feature extraction purposes. The researchers selected ViT for its ability to track remote connections through its global self-attention system, which identifies distant dependencies, while ResNet and DenseNet serve as established convolutional neural networks that effectively extract features from data.

The extracted features were subsequently input into two classifiers: Support Vector Machine (SVM) with an RBF kernel and Random Forest (RF). SVM establishes its reputation as a powerful tool for work in high-dimensional feature spaces, while Random Forest delivers strong decision-making abilities through its ensemble learning methods. The combination of multiple feature extractors with two different classifiers ensures a comprehensive evaluation of feature quality and classifier suitability.

The authors performed additional tests to find how feature extraction methods and classifiers affected their results. The authors assessed six model combinations through systematic evaluation. The six different model combinations, which included ViT-SVM and ViT-RF, ResNet-SVM and ResNet-RF, DenseNet-SVM, and DenseNet-RF. The experiments were conducted through identical dataset splits, which enabled them to maintain equal conditions for testing each model. The multiple performance metrics were also used to measure performance, which included accuracy, precision, recall, F1-score, and confusion matrices, while ROC curves provided a basis for evaluating class-wise separability and classifier confidence. The results obtained from these experiments were analysed and compared in the Results and Discussion section, which helps the readers to understand which model configurations deliver better pest classification performance.

5.4. System Architecture

The paper presents a real-time pest detection system that operates its complete detection system in agricultural environments. The system architecture of the novel AgroPestBot is proposed in Figure 3. The system uses a camera to take pictures of pests, processes them locally, and then uses a model that combines Vision Transformer (ViT) with Support Vector Machine (SVM) running on an NVIDIA Jetson Nano to categorise them.

The Android mobile app immediately displays the identified pest's name for easy reference. There may be some variation in the size, sharpness, and backdrop of the field photographs. Consequently, in order to make the model more resilient, preprocessing techniques including scaling, denoising, background removal, and information augmentation are used. To aid in visually differentiating species, a Vision Transformer extracts characteristics based on texture and high-level shapes.

In order to classify the pests, finally, an SVM classifier is fed these learnt attributes. The assessment results showed that the hybrid ViT-SVM model was 96% accurate and had an F1 score of 96%. Lightweight, field-powered inference is made possible after model training by exporting and deploying it on Jetson Nano. An Android App, integrated through Wi-Fi, allows farmers to monitor the captured image and view the recognized pest name in real time. A manually-operated robot prototype uses an Android app to traverse the region. The bot's fixed camera responds to human input by taking pictures of pests. Currently, real-time continuous pest detection is difficult due to the bot's restricted height and the tiny size of many agricultural pests. This means the robot can take still photos whenever needed and then run them by means of Jetson Nano's ViT-SVM classifying pipelines. Users are able to make a precise analysis on the mobile app, which displays the pest name. The system demonstrates potential for real-time field deployment, even if it is currently limited by hardware.

5.5. Workflow of the Proposed System

The AgroPestBot's process is shown in Figure 4. There are three stages to the process of pest classification. The novel hybrid model (ViT-SVM) is trained and tested in Phase I. The chassis, which contains the power source Jetson Nano, and the camera, is designed and constructed during Phase II. The third and final stage involves showing the mobile interface together with navigational features, the recorded images, and the outcome as a pest name in a window.

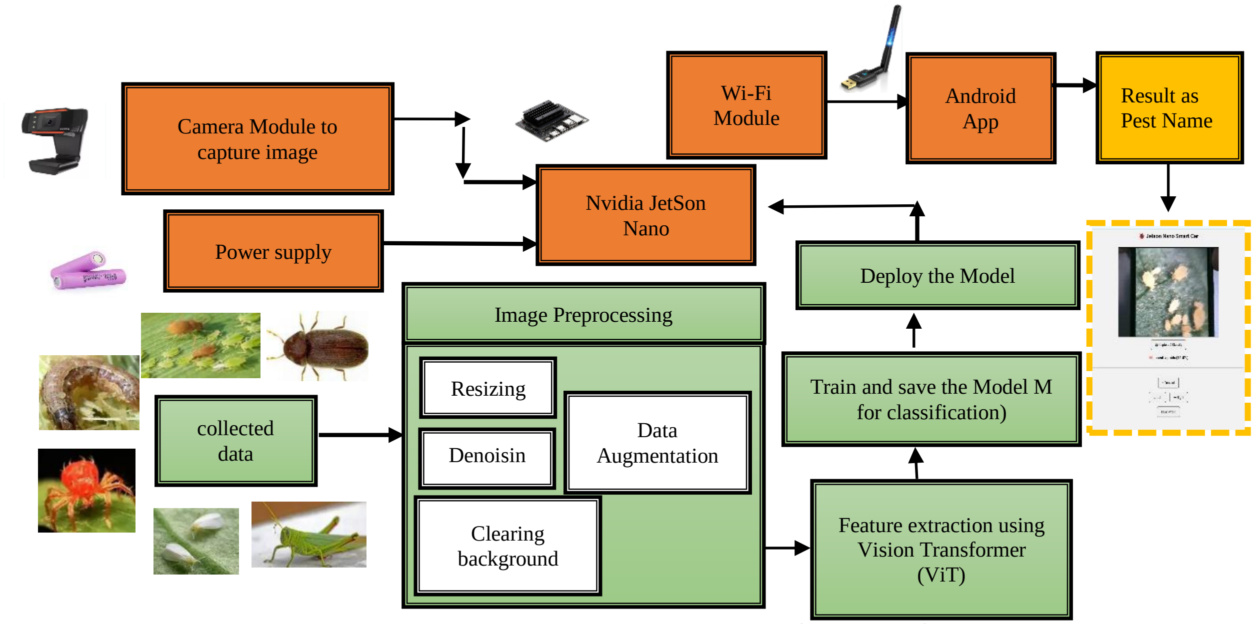


Fig. 3 Nvidia Jetson Nano-based agropestbot architecture for pest classification

6. Results and Discussion

The results show that a hybrid model is the best option when there is a lack of available data, when there is a need for practical deployment across edge computing devices, and when there are hardware limitations. The authors conducted a series of tests to evaluate the performance of different feature extractor–classifier combinations in order to select the hybrid technique that precisely identified the pest. The various combinations were carefully chosen from the literature survey. These combinations included DL and ML algorithms like ResNet and DenseNet, with SVM. Vision Transformer (ViT) with RF and SVM. ResNet and DenseNet with Random Forest. Consistent with previous research, the findings suggest that classic CNN-based feature extractors like ResNet and DenseNet have a hard time handling certain pest classes, often resulting in very poor recall for under-represented groups, including stem borer, red spider, and armyworm. Particularly struggling to generalise across classes, DenseNet-SVM achieved an overall accuracy of just 37.8 percent. The ViT-based models demonstrated much better results because both ViT–RF and ViT–SVM delivered high accuracy results across most pest identification tests. The overall accuracy of ViT–RF reached 93%, but the system failed to identify visually similar pests because it struggled with red spider and stem borer identification. The ViT–SVM hybrid achieved its highest results through 96% accuracy, which maintained strong class balance across different categories. The model showed high precision and recall results for almost all pest types while accurately detecting small and visually challenging classes, which included mites, sawflies, beetles, and ball worms. To further analyse classification behaviour, both confusion matrices in Figure 5 and ROC plots in Figure 6 were generated. The ViT-SVM confusion matrix showed fewer misclassifications than other models, which was especially true for pest classes that usually share visual similarities. The results from the hybrid models that this study conducted during their feature extraction and classification work are shown in Table 2 and Figure 5. The study uses both raw measurements and weighted average measurements of three metrics, which include Precision and Recall, and F1-score, to assess the different hybrid algorithms. The overall trend demonstrates that the model consistently achieves better results than both major and minor groups. The weighted average method evaluates each class according to its sample weight, which results in a performance metric that accurately represents actual class distribution within the dataset. When both metrics are used, the classifier's performance on unbalanced datasets becomes easier to assess.

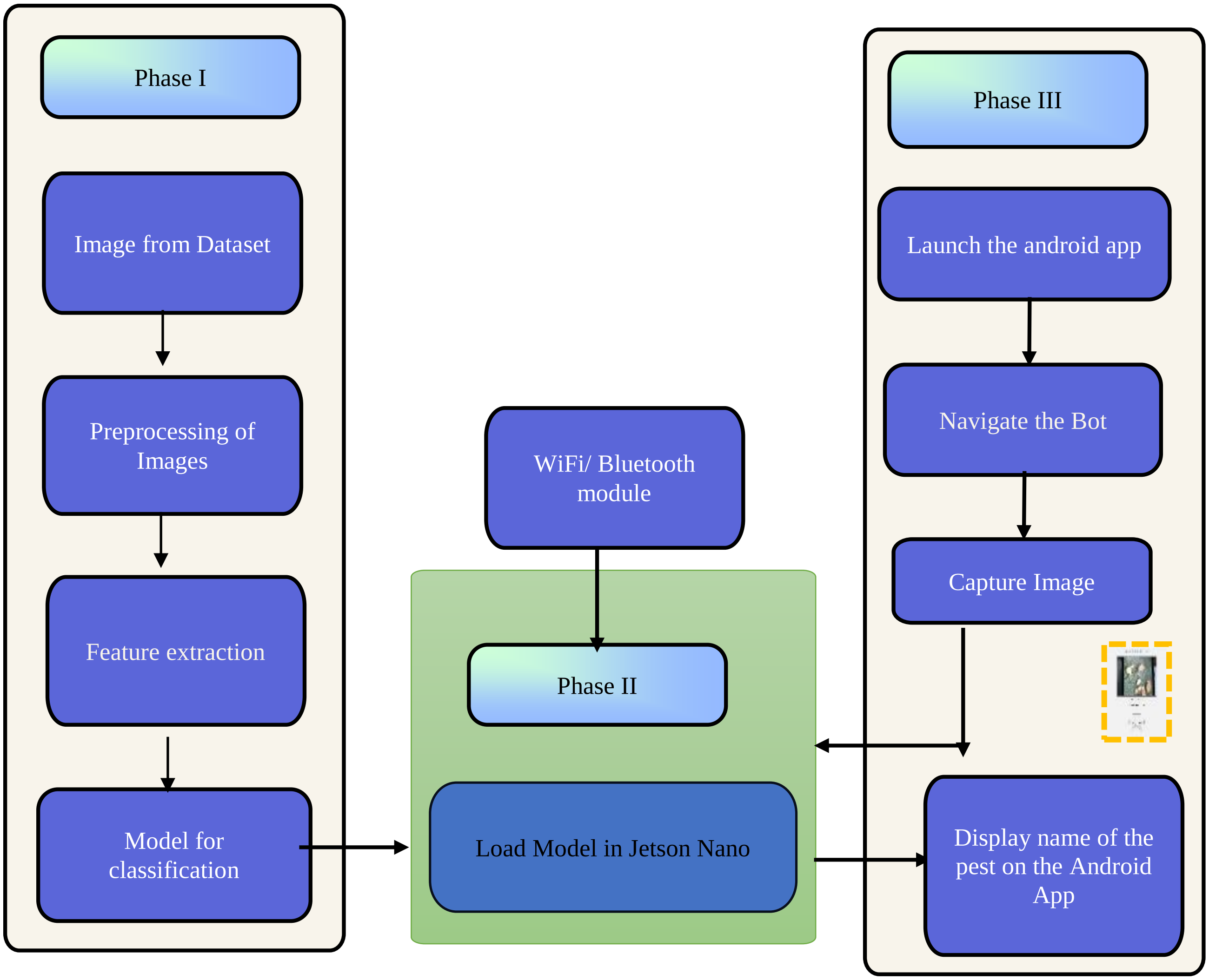
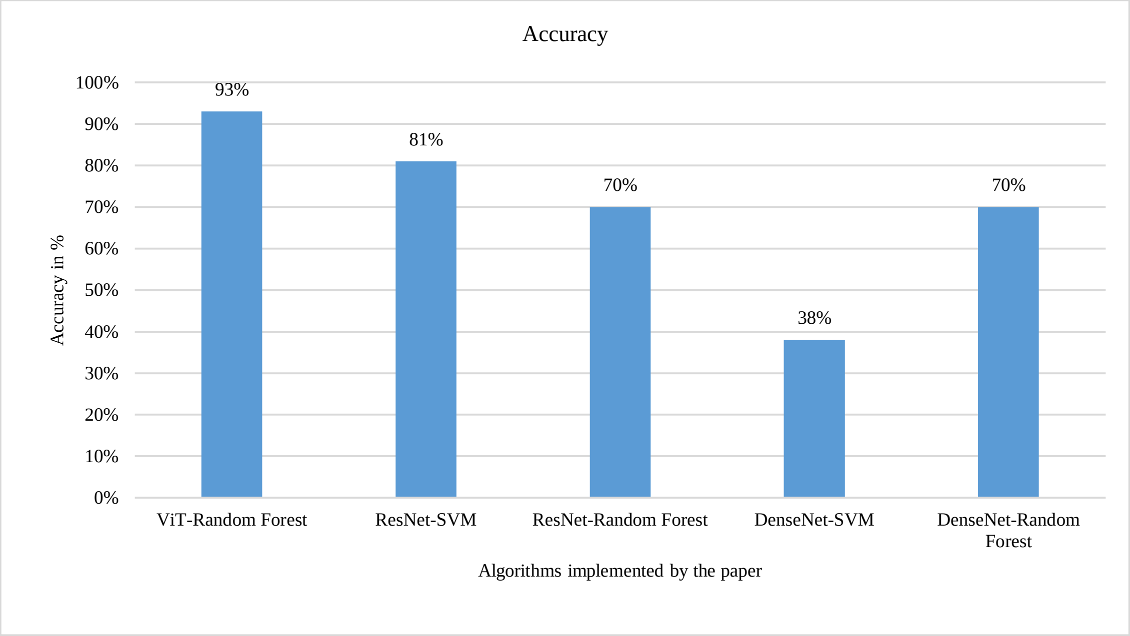


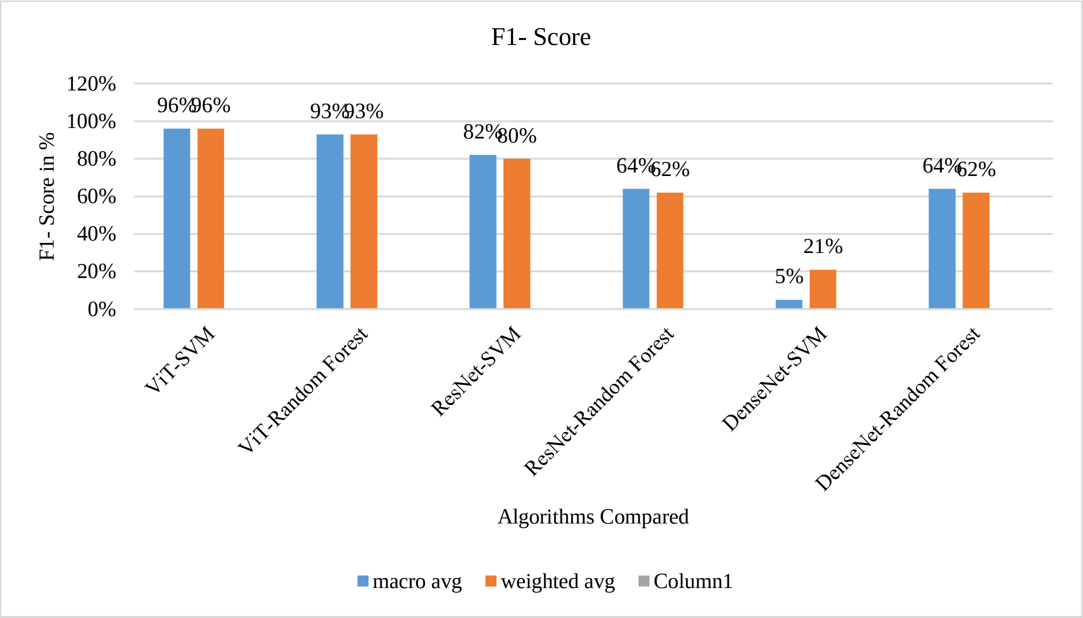
Fig. 4 Software flow diagram for the proposed system

Table 2. Comparison of the implemented Hybrid algorithms

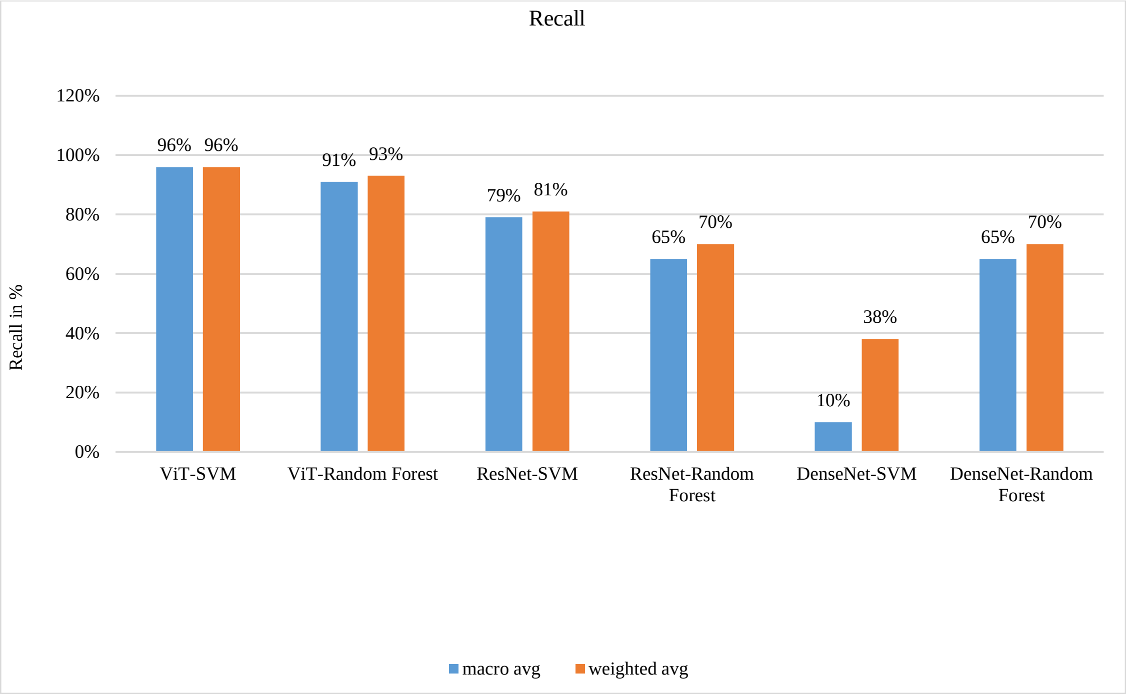
Algorithms	**Performance Parameters**						
	Accuracy	**F1 Score**		**Recall**		**Precision**	
		macro avg	weighted avg	macro avg	weighted avg	macro avg	weighted avg
ViT-SVM	96%	96%	96%	96%	96%	97%	96%
ViT-Random Forest	93%	93%	93%	91%	93%	97%	94%
ResNet-SVM	81%	82%	80%	79%	81%	90%	83%
ResNet-Random Forest	70%	64%	62%	65%	70%	75%	66%
DenseNet-SVM	38%	5%	21%	10%	38%	4%	14%
DenseNet-Random Forest	70%	64%	62%	65%	70%	75%	67%



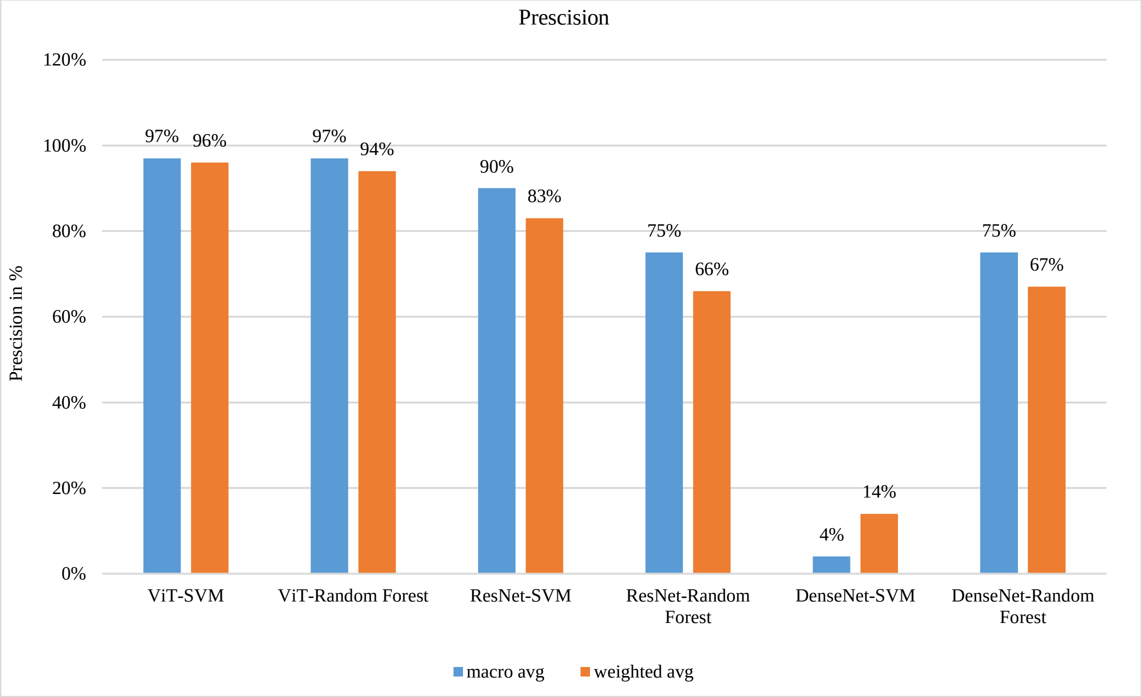
(a)



(b)



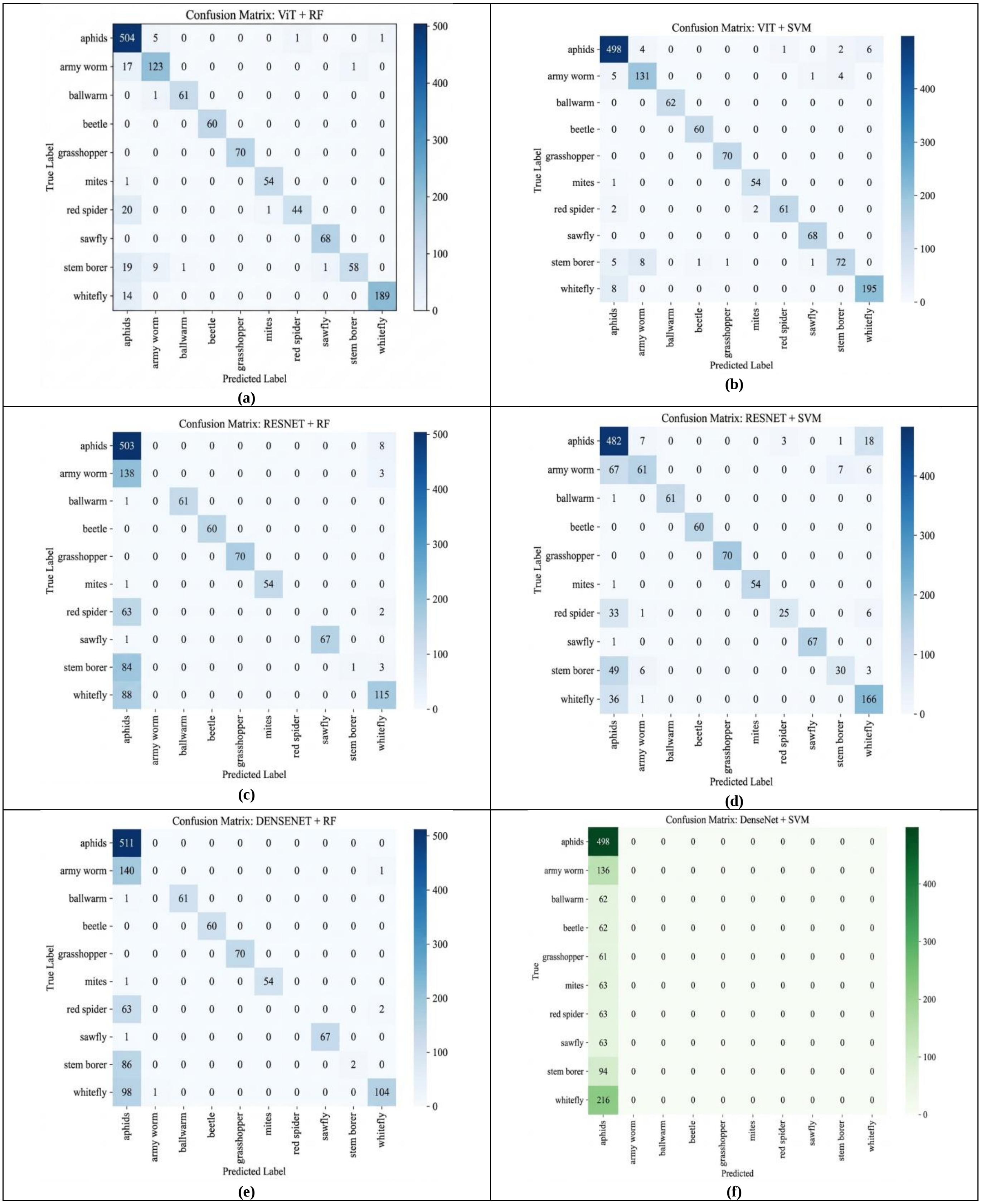
(c)



(d)

Fig. 5 (a) and (b) Accuracy and F1- Score for algorithms, (c) and (d) Performance parameters for Hybrid algorithms.

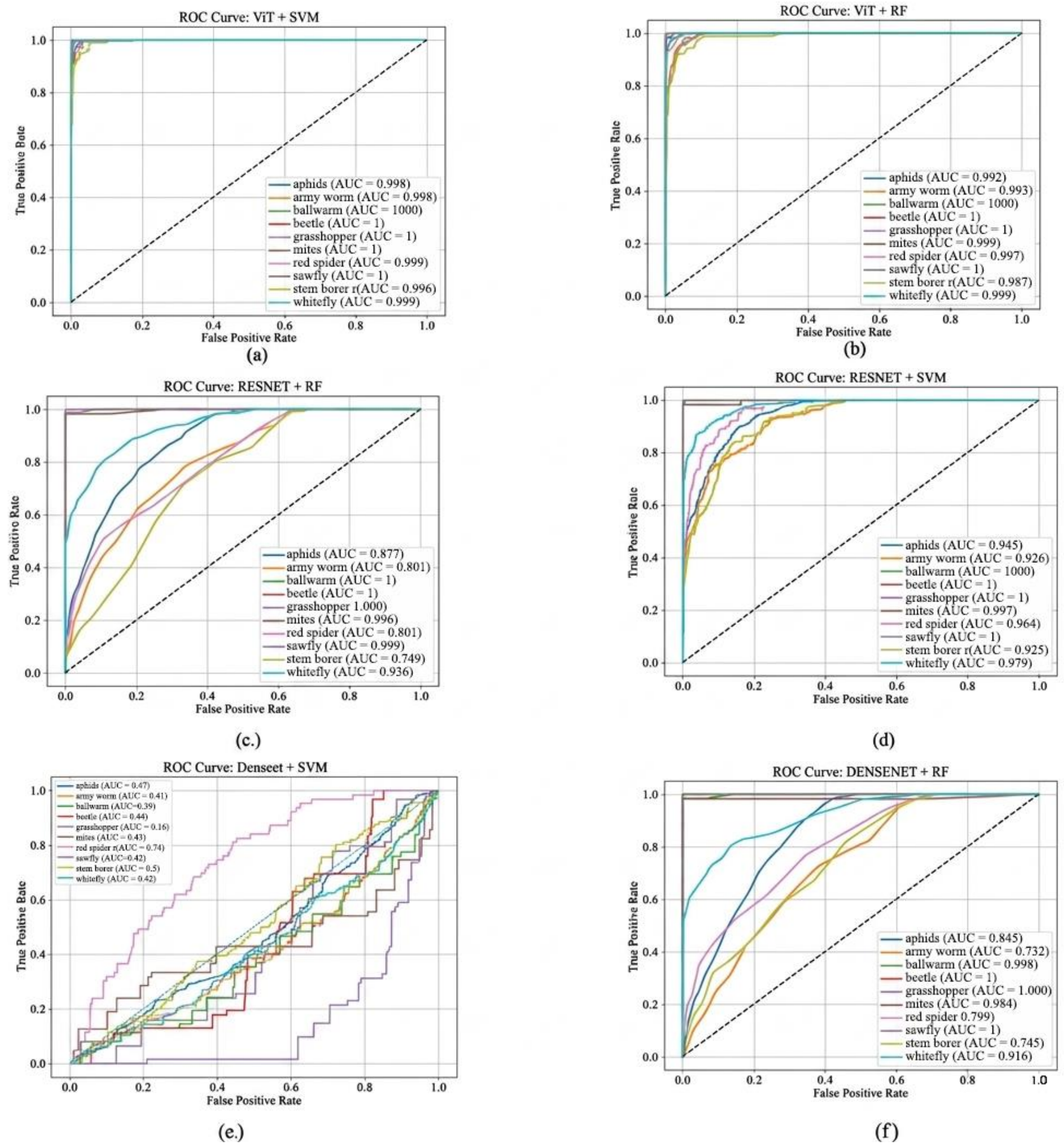
The performance comparison drawn from Table 2 and Figure 5 shows that ViT-based hybrid models consistently outperform all other approaches. The highest accuracy and balanced macro-weighted scores across all metrics were achieved by ViT-SVM. ResNet-based hybrids perform moderately. DenseNet-based combinations present unstable behavior, which results in significantly decreased macro-weighted averages. Overall, ViT-SVM stands out as the most dependable hybrid model, which delivers consistent performance across different tasks.



(a) **(b)** **(c)** **(d)** **(e)** **(f)**

Fig. 6 Confusion matrix plotted for Hybrid algorithms, (a) ViT-RF (b)ViT-SVM (c) ResNet-RF (d) ResNet-SVM (e) DenseNet-RF, and (f) DenseNet-SVM.

The confusion matrices for all hybrid models appear in Figure 6(a) shows that the ViT + RF model achieves high precision across most pest categories-particularly aphids, armyworm, grasshopper, and whitefly-with only minor confusion in visually similar classes such as red spider and stem borer. (b) illustrates that the ViT + SVM model delivers the most consistent and accurate performance, with very few misclassifications and notably high recognition rates for aphids, armyworm, red spider, sawfly, and whitefly. The model achieves its best performance through its ability to differentiate between very similar patterns according to its minimal cross-confusion measurement, which demonstrates the model's superior performance compared to all other models it tested. (c) shows that the ResNet + RF model performs well for several classes, including aphids, armyworm, beetle, grasshopper, and sawfly, but exhibits confusion for red spider, stem borer, and whitefly. The ResNet + SVM model shows strong performance with common classes, but the system fails to recognize visually similar classes, which leads to moderate mismatches according to the findings of (d). The DenseNet + RF model shows strong accuracy results with main classes; however, it fails to identify red spider, stem borer, and whitefly, which are less common classes. The (f) statement shows that the DenseNet plus SVM model predicts almost all samples as aphids because it suffers from severe class imbalance, which prevents accurate identification of other pest species. The confusion matrix analysis shows that the ViT plus SVM hybrid model provides the strongest pest identification performance through its most accurate and balanced classification results.



(a) (b) (c.) (d) (e.) (f)

Fig. 7 ROC curves plotted for various hybrid algorithms, (a) ViT-RF, (b)ViT-SVM, (c) ResNet-RF, (d) ResNet-SVM, (e) DenseNet-RF, and (f) DenseNet-SVM.

The ROC curves provide an intuitive method that does not require thresholds to measure how well each hybrid model can identify pest classes. The AUC metric provides visual and quantitative information that the model uses to demonstrate its performance. The ViT-based hybrids deliver the strongest class separability according to the results that Figure 7 displays. The ViT + RF model maintains consistent high AUC scores, which range from 0.987 to 1.000, because its detection curves stay close to the top-left area, which shows reliable performance at extremely low false-positive rates. The ViT + SVM model proves its ability to identify different classes, which approach perfect accuracy, according to its AUC score, ranging between 0.995 and 1.000, because it shows outstanding performance during real-time multi-class pest detection tasks. The ResNet + RF and ResNet + SVM models show different results because they achieve high class separability for some classes, but they fail to separate additional classes. The DenseNet-based hybrid systems show their most unpredictable performance because DenseNet + RF shows average results with unpredictable AUC results, and DenseNet + SVM reaches AUC values between 0.4 and 0.5, which shows it does not work for this task.

The ROC plots establish that transformer-based feature extractors, together with classical machine-learning classifiers, demonstrate the highest stable discriminative ability for all pest classes. The evaluation requires more than performance metrics, which include accuracy, precision, recall, F1-score, and AUC, because both models demonstrate strong performance. The metrics show model performance, yet they do not demonstrate whether the detected differences reach statistical significance or result from sampling errors. The study examines statistical significance between ViT-RF and ViT-SVM, which represent the best-performing models, to validate their performance enhancements. The comparison between the two hybrids became necessary because they both achieved high performance metrics, and statistical testing proved essential to determine which model produced consistently superior results.

The probability-based analysis process demonstrated the ViT-SVM strengths, which extended beyond its accuracy analysis. The model outperformed ViT-RF with macro-ROC-AUC 0.5686 and PR-AUC 0.1775, which provided superior ranking abilities and trustworthy confidence assessment. The AUC results demonstrated particular effectiveness in distinguishing between red spider and aphid classes. The bootstrap confidence intervals for both ROC-AUC and PR-AUC measurements showed minimal overlap between the two models, which proved that ViT-SVM delivers more consistent results through multiple evaluation tests. The McNemar test returned a p-value of 1.0000 because both models achieved the same final results. The extended probability assessment demonstrates that ViT-SVM provides better accuracy in identifying visually similar pest categories.

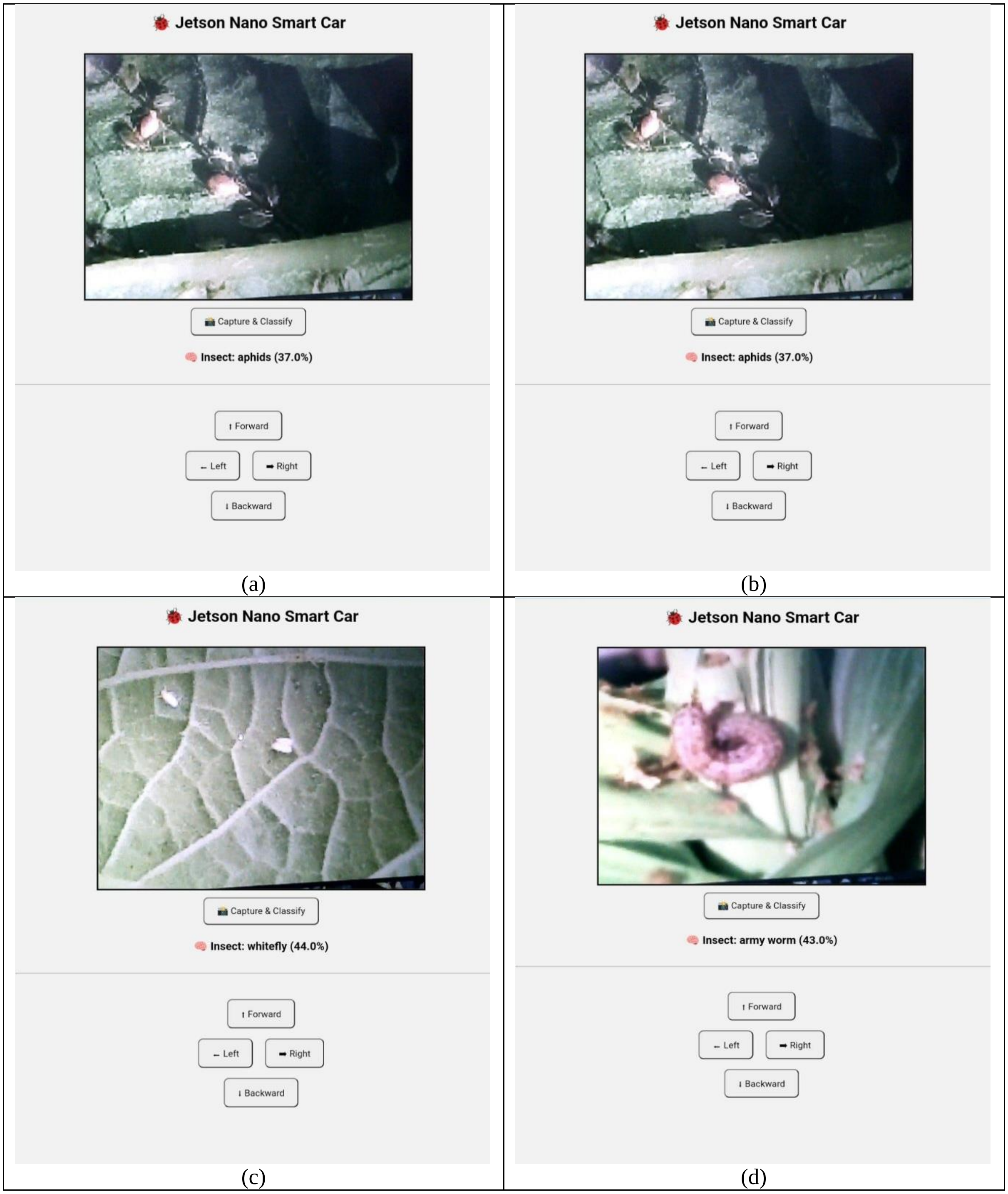
Based on these results, the suggested framework uses Support Vector Machines (SVMs) for classification and Visualisation (ViT) for feature extraction; it is designed to be deployed on a robotic platform powered by NVIDIA Jetson Nano for the purpose of real-time pest detection in agricultural settings. We validate its applicability for real-time field application by successfully deploying the trained ViT-SVM model on a Jetson Nano-based mobile prototype. The embedded system can detect pests locally, without the need for cloud access, when combined with an Android interface. By including this, we can see that the suggested method works well in controlled studies and would be a good fit for actual farming settings.

The final model's performance was confirmed using real-time field images taken in agricultural situations. Pests were photographed on several crops, including okra, cauliflower, roses, hibiscus, and fruit-bearing plants like mango. Particularly noticeable were pests like Mango Stem Borer around the base of the trees. One step toward a completely integrated pest detection system is Agro Pest Bot, which has built-in categorisation capabilities.

A height-adjustable robotics platform that can adjust to different crop structures would be a great feature for future advancements to include in order to make deployment under dynamic agricultural settings more successful. The performance of the model may be improved by increasing the size and diversity of the pest dataset. While the present prototype takes pictures using a standard webcam, a smartphone would provide better results in terms of precision and image quality.

The user receives real-time feedback on the categorised pest name using the Android app. Figure 8 shows the outcomes of the field testing conducted using the mobile interface. With its suggested ViT-SVM-based AgroPestBot, the aforementioned research successfully addresses the pressing issue of real-time pest detection in agriculture. In addition to computational enhancements, the robotic implementation driven by the Jetson Nano guarantees mobility and real-time field usability. Their connected Android app offers farmers an easy-to-use interface. Improving a bot's use in the field might be as simple as designing an arm that can be adjusted to fit various plant and crop heights. Furthermore, features such as muddy and uneven terrain are crucial for improving the bot's mobility and dependability in actual agricultural settings.

Based on the related work and literature survey, it is clear that recent research is moving toward hybrid classification approaches. Deep learning methods have overcome many limitations of traditional feature extraction techniques and have delivered strong classification results.



Insect: aphids (37.0%)

Insect: aphids (37.0%)

Insect: whitefly (44.0%)

Insect: army worm (43.0%)

(a) (b) (c) (d)

Fig. 8 Real-time categorisation outcomes by recognising Pests in the field photographs alongside their respective identities from left to right: (a) Aphids, (b) Grasshoppers, (c) Whitefly, and (d) Armyworm.

However, these models are often complex, computationally expensive, and require large training datasets. To address these issues, the present work combines machine learning with transfer learning using ViT-based feature extraction. This combination creates a lighter and more efficient model that performs well even with limited data and is more suitable for deployment on edge devices for real-time pest classification. The effective balance between accuracy, efficiency, and reduced computational cost helped achieve better results compared with several existing methods reported in the literature. This work presents a lightweight hybrid ViT–SVM model optimized for resource-constrained, field-level pest recognition. Unlike prior studies focusing only on algorithmic performance, the proposed system integrates embedded Jetson Nano hardware, a mobile application interface, and a manual-driven robot, AgroPestBot, for real-time, in-field pest detection. The study used class-wise ROC-AUC, PR-AUC, and McNemar tests together with bootstrap confidence intervals to validate the model's performance, which showed better probabilistic discrimination and stronger ability to separate different classes. The research work delivers its main scientific and practical value through the combination of three elements, which include new hybrid modelling methods and their implementation in deployed systems for actual field use.

6. Conclusion and Future Scope

The research introduces AgroPestBot, which serves as a prototype system that uses a hybrid ViT-SVM model for detecting pests in agricultural fields. The research demonstrates that ViT-SVM technology outperforms standard CNN models through its ability to achieve precise results and equal performance across all categories and dependable probability assessments of visually identical pests. The model achieves 96% accuracy, 96 % F1-score, 96% recall, and 97% precision for 10 pest classes. The Jetson Nano–powered prototype, combined with its Android interface, provides pest identification tools that can be used in locations where resources are limited. The existing system operates manually because it depends on human operators, which shows that more automated features should be added to improve its performance in outdoor environments. The upcoming research will work on three main objectives: to create automatic functions for Agro Pest Bot, develop its ability to move through different types of agricultural environments, add features that enhance image capture, and create a broader collection of pest data to improve machine learning. The system will become more practical for actual farming operations because its improvements will enhance both its ability to perform tasks and its capacity to handle various farm settings.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Funding Statement

No Funding

Acknowledgments

We want to express our heartfelt gratitude to Dr. Ganesh Bansode, Assistant Professor, Entomology at the Zonal Agricultural Research Centre, Pune, for working with us in the field and helping capture real crop and pest images. His guidance and active involvement greatly enhanced the practical depth and field relevance of this research.

References

[1] K. Gireesan, *From Agricultural Co-operatives to Farmer Producer Companies: Analysing the Transition of Co-Operativism in India*, Routledge, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Alok K. Sharma, A. Rajawat, and V. Gupta, "An Analytical Study of Contribution of Agriculture Sector in Growth of Indian Economy," *Jai Maa Saraswati Gyandayni An International Multidisciplinary e-Journal*, vol. 7, no. 1, pp. 08-18, 2021. [CrossRef] [Publisher Link]

[3] S.D. Talekar, and Yonas Dubale, Economic Reforms and Agricultural Growth in India: Issues and Challenges, Economics, Agricultural and Food Sciences, 2020. [Online]. Available: https://www.semanticscholar.org/paper/Economic-Reforms-and-Agricultural-Growth-in-India%3A-Talekar-Dubale/a3abc93b7b17858e8f99eec065cffa9de13cf5f2

[4] J. L. Quaresma, The Influence of Agriculture for a Country, Communication and Digital Media, vol. 1, no. 1, 2020. [Online]. Available: https://comopolo.com/index.php/comdig/article/view/2

[5] P.S. Soumia et al., *Entomopathogenic Microbes for Sustainable Crop Protection: Future Perspectives*, Advances in Sustainable Agriculture and Farming Systems, Springer, Singapore, pp. 469-497, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Mifftha Yaseen et al., *Insect Pest Infestation During Storage of Cereal Grains, Pulses and Oilseeds*, Health and Safety Aspects of Food Processing Technologies, Springer, Cham, pp. 209-234, 2019, [CrossRef] [Google Scholar] [Publisher Link]

[7] Amelia Apriyuni et al., "Analysis of the Use of PPE on the Health Risks of Farmers Spraying Pesticides," *Journal of Educational Innovation and Public Health*, vol. 2, no. 3, pp. 94-114, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mayang Dwi Octavia, and Susilawati Susilawati, "Analysis of the Use of Personal Protective Equipment on the Health Status of Pesticide Spraying Farmers," *Healthy People: Journal of Public Health*, vol. 2, no. 3, pp. 328-337, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Christos A. Damalas, and Spyridon D. Koutroubas, "Farmers' Exposure to Pesticides: Toxicity Types and Ways of Prevention," *Toxics*, vol. 4, no. 1, pp. 1-10, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[10] B. Prasath, and M. Akila, "IoT-based Pest Detection and Classification Using Deep Features with Enhanced Deep Learning Strategies," *Engineering Applications of Artificial Intelligence*, vol. 121, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sandhya Devi Ramiah Subburaj et al., "Efficient Pest Detection through Advanced Machine Learning Technique," *Current Agriculture Research Journal*, vol. 12, no. 3, pp. 1127-1134, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Karla Zayood, and Rama Asad Nadweh, “Automated Insect Detection and Classification Using Pelican Optimization Algorithm with Deep Learning on Internet of Enabled Agricultural Sector,” *International Journal of Advances in Applied Computational Intelligence*, vol. 7, no. 1, pp. 50-62, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Asaf Raza et al., “Automatic Classification and Detection of Insect Pests Using Deep Transfer Learning,” *2024 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] R. Prabha, and K. Selvan, “Modified RESNET50 with Attention Module for Detection and Classification of Pests in Vegetable Crops,” *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 61, no. 2, pp. 150-169, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[15] S. Senthil Pandi et al., “A Deep Learning Framework with Attention Mechanism for Accurate Detection and Classification of Crop Pest Management,” *2024 4th International Conference on Computer, Communication, Control & Information Technology (C3IT)*, Hooghly, India, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] P. Venkatasaichandrakanth, and M. Iyapparaja, “GNViT-An Enhanced Image-Based Groundnut Pest Classification Using Vision Transformer (ViT) Model,” *PLOS ONE*, vol. 19, no. 3, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] M. Xie, and N. Ye, “Multi-Scale and Multi-Factor ViT Attention Model for Classification and Detection of Pest and Disease in Agriculture,” *Applied Sciences*, vol. 14, no. 13, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Gabriel Savio de Lima Mota et al., “Classifying Pests in Crop Images Using Deep Learning,” *Proceedings of the 18th Workshop on Computer Vision (WVC)*, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] P. Venkatasaichandrakant and M. Iyapparaja, “Pest Detection and Classification in Peanut Crops Using CNN, MFO, and EViTA Algorithms,” *IEEE Access*, vol. 11, pp. 54045-54057, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Nikita Agarwal et al., “An Improved Deep Learning Model Implementation for Pest Species Detection,” *International Conference on Artificial Intelligence: Towards Sustainable Intelligence*, Pune, India, pp. 119-131, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Agus Kurniawan, *Introduction to NVIDIA Jetson Nano*, IoT Projects with NVIDIA Jetson Nano, Apress, Berkeley, CA, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] NVIDIA, Get Started with Jetson Nano Developer Kit, NVIDIA Developer, 2026. [Online]. Available: https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit.

[23] NVIDIA, TensorRT Open-Source Software, 2019. [Online]. Available: https://github.com/NVIDIA/TensorRT

[24] Gayatri Pattnaik, and K. Parvathi, “Automatic Detection and Classification of Tomato Pests Using Support Vector Machine Based on HOG and LBP Feature Extraction Technique,” *Progress in Advanced Computing and Intelligent Engineering*, vol. 2, pp. 49-55, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Gayatri Pattnaik, and Kodimala Parvathy, “Machine Learning-Based Approaches for Tomato Pest Classification,” *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 20, no. 2, pp. 321-328, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Abdul Amir Abdullah Karim, and Rafal Ali Sameer, “Image Classification Using Bag of Visual Words (BoVW),” *Al-Nahrain Journal of Science*, vol. 21, no. 4, pp. 76-82, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[27] T. Ojala, M. Pietikäinen, and T. Maenpaa, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[28] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, “SURF: Speeded Up Robust Features,” *European Conference on Computer Vision*, Graz, Austria, pp. 404-417, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[29] S. Murakami, H. Homma, and T. Koike, “Detection of Small Pests on Vegetable Leaves Using GLCM,” *American Society of Agricultural and Biological Engineers*, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[30] Tao Liu et al., “Detection of Aphids in Wheat Fields Using a Computer Vision Technique,” *Biosystems Engineering*, vol. 141, pp. 82-93, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[31] Arti Prasad et al., “A Detailed Survey on Awareness, Knowledge and Practice of Pesticides Used Against Various Vegetables, Fruits and Cereal Crops Grown in and Around Udaipur Region of South Rajasthan, India,” *Bulletin of Pure and Applied Sciences- Zoology*, vol. 42A, no. 1, pp. 43-63, 2023. [Google Scholar] [Publisher Link]

[32] Santosh Kumar Sahu, and Manish Pandey, “An Optimal Hybrid Multiclass SVM For Plant Leaf Disease Detection Using Spatial Fuzzy C-Means Model,” *Expert Systems with Applications*, vol. 214, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[33] Thierry Tchokogoué et al., “A Robust Segmentation Method Combined with Classification Algorithms for Field-Based Diagnosis of Maize Plant Phytosanitary State,” *Journal of Intelligent Systems*, vol. 33, no. 1, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]