International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

Pest Detection and Classification in Peanut Crops

Sushma D S[1], Mohammed Alqhama[2], Aravind M[3], Jayanth A B[4], Rakshith Kumar K[5]
[1]Rajeev Institute of Technology, Hassan, Karnataka, India.
[2,4]Rajeev Institute of Technology, AKTU Hassan, Karnataka, India.
[3,5]Rajeev Institute of Technology, UPES Hassan, Karnataka, India.
Emails:*rjsush007@gmail.com[1],mohammedalqama1@gmail.com[2],aravindm6361@gmail.com[3],babujayanthab@gmail.com[4],rakshishetty62@gmail.com[5]*

Abstract
Recent advancements in image processing have significantly improved pest detection and classification in peanut crops. Our study introduces an innovative approach that optimizes image features for accurate pest identification. Leveraging insights from successful image analysis methodologies, our model employs a tailored architecture for pest detection, segmentation, and classification tasks. By integrating dual branch segment representations and a dual-layer transformer encoder, we aim to enhance image representations and consolidate pest image segments of varying sizes. We evaluate our approach using three distinct pest datasets—Aphids, Wireworm, and Gram Caterpillar—ensuring comprehensive analysis and model validation. Prior to training, we preprocess the datasets extensively, employing feature extraction techniques and addressing image quality issues. We then apply normalization procedures to standardize the data for seamless integration into our model architecture. Our methodology focuses on extracting key features through self-attention mechanisms and standardized scaling processes to enhance predictive capabilities. Comprehensive experimentation demonstrates the superiority of our approach, outperforming established benchmarks in pest detection and classification with high accuracy rates. In summary, our study presents a novel framework that optimizes feature extraction and enhances predictive accuracy in pest detection and classification for peanut crops, addressing the unique challenges of agricultural pest identification.
Keywords*: CNN; peanut; moth flame optimization; Pest; vision transformer.*

1. Introduction
Agriculture plays a vital role in sustaining both human and livestock populations globally. The integration of environmentally friendly artificial intelligence (AI) and Internet of Things (IoT) technologies has expanded agriculture's role in clean energy generation. Additionally, farming serves as a primary source of natural substances used in the production of materials, chemicals, and pharmaceuticals. Despite a modest 15% increase in agricultural land between the 1960s and the early part of the 21st century, agricultural production tripled. This surge was attributed to the adoption of pesticides and fertilizers, as well as advancements in precision farming and the development of high-yielding crop and livestock varieties. However, recent trends indicate a slowdown in the rate of agricultural production growth, exacerbated by emerging challenges such as climate change, population growth, and rural-to-urban migration. The agriculture and food processing industry is pivotal in any nation and plays a significant role in enhancing the quality of rural and food products. In agricultural regions, the growth in food processing transformations is primarily driven by warehousing services and domestic market demands. Under specific conditions, it necessitates infrastructure, consistent equipment support, and workspaces regularly [1-4]. Pest infestation is a major challenge in the agriculture sector, leading to a decline in crop quality. Pests, microorganisms, and weeds cause significant yield losses, resulting in reduced market value for agricultural products. Discovering more efficient methods to achieve even small increases in productivity can mean the difference between

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

turning a profit or incurring losses. It is essential to understand the impact of pest infestation on crops, affecting their growth. Major cash crops significantly contribute to overall production. Pests are the primary cause of crop quality degradation and reduced crop efficiency. Therefore, monitoring and evaluating losses due to pests are crucial to ensure crop quality and security in agriculture. Peanuts are a versatile crop with significant nutritional value. As a primary source of oil and economic yield for our Government, its cultivation area is continuously expanding, making it the second largest cultivated crop in India. Both the nuts and the diverse field environments make peanut leaves susceptible to contamination by microorganisms. Microorganisms can spread rapidly through natural factors and have a high reproductive capacity [5]. The primary factor influencing their proliferation is the moisture content of peanuts during the seedling stage. Leaf diseases can reduce peanut yield and quality by destroying green tissue and chlorophyll in the leaves. Accurate identification of peanut leaf infections requires specialized knowledge, as they can easily be misdiagnosed solely based on visual observation. Therefore, timely diagnosis and treatment of peanut diseases are essential. The key to controlling peanut diseases is to promptly and accurately identify the type of disease and implement appropriate control measures. Various methods exist for detecting plant diseases in their early stages. The traditional method of visual observation with the naked eye is inadequate and unreliable for large crops. Hence, leaf disease detection is a significant area that offers numerous benefits in monitoring large crop fields. Peanut diseases can affect yield and quality by damaging the green tissue of the leaves. Controlling these peanut diseases involves promptly and accurately identifying the type of disease and implementing appropriate corrective actions in a timely manner. Leveraging the capabilities of advanced Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Machine Learning (ML), and Vision Transformer (ViT) algorithms, disease recognition is efficient, time-saving, and accurate. The current research focuses on predicting peanut diseases in real-time environments. Although machine learning and CNN algorithms are suitable for image classification, segmentation, and identification, the accuracy they achieve is not satisfactory. To achieve efficient results, this study proposes an Enhanced Vision Transformer Architecture (EViTA) for analyzing, classifying, segmenting, and identifying peanut pests based on images. In the EViTA method, input images are segmented into multiple segments for easier processing. These processed images are then encoded with positional information and fed into distinct transformer layers for accurate identification of peanut pest nature. Experiments were conducted using publicly available datasets containing Aphids, Wireworm, and Gram Caterpillar pests, which commonly affect peanut crops. The MFO algorithm is utilized in this work to extract features from the chosen datasets. The extracted features are then input into the Extra Arrangement Segment (EAS) block, containing the most impactful features for peanut growth. Finally, the EViTA method is fed with the extracted data to predict affected peanut crops and aid in increasing peanut crop growth. The significant contributions of the proposed model are as follows:

- CNN is employed to predict pest infections in peanut crops.
- MFO is utilized to enhance the prediction rate by selecting the most relevant features.
- MFO and state-of-the-art techniques are comprehensively evaluated.
- Experiments demonstrate that the proposed model outperforms other popular EViTA methods, highlighting the beneficial effect of integrating MFO with EViTA methods.

2. Method

2.1 Insect Dataset Description

For this study, three categories of insect datasets were collected. The primary insect dataset that significantly affects groundnut leaf is Aphids (IP102 Dataset), which comprises 42 types of pests

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

found in field crops. The next insect category considered is Wireworm (IP102 Dataset), consisting of 88 data points for training, 14 for validation, and 45 for testing purposes. The final dataset, Gram Caterpillar, was sourced from the Kaggle dataset, with 210 data points for training, 35 for validation, and 105 for testing perspectives. To enhance the representation of insect descriptions, image preprocessing techniques were applied to isolate insects from the original images before inputting them into the deep learning models. Specifically, RGB insect images were converted into grayscale images. Edge Detection was employed to identify edges in insect images and suppress noise. The distinct external patterns in identified pest images were then determined. Each pattern was bounded by four points (p, q, r, s), where (p, q) represents the upper-left corner of the bounding rectangle, and (r, s) represents its width and height. Subsequently, an upright bounding rectangle was established for each pattern. If the bounding rectangle containing the insect had a width and height greater than 50 pixels, the region of interest (ROI) of the insect was extracted using the coordinates (p, q, r, s) from the original RGB insect image. Finally, the processed insect image was obtained. Figure 1 illustrates the preprocessed sample insect images from Aphids (IP102 Dataset), Wireworm (IP102 Dataset), and Gram Caterpillar datasets.

Table 1 Peanut Pest Dataset Description

Dataset	Training	Validation	Testing	Total Images
Aphids	339	59	170	565
Wireworm	88	14	45	147
Gram Catepiller	210	35	105	350
Total Count	637	105	320	1062

The number of pest images in each dataset is outlined in Table 1. Furthermore, all processed insect images were resized to 227 x 227. Mathematical transformation techniques such as scaling, translation, rotation, and flipping were applied to augment the number of insect samples in the datasets.

2.2 Convolutional Neural Network (CNN)

CNNs are currently one of the most popular models and have demonstrated outstanding performance on numerous image classification tasks in various domains. The concept of weight sharing in CNNs facilitates efficient feature extraction from images and reduces the over fitting problem. A typical CNN architecture consists of convolutional layers, pooling layers, and fully connected layers [6]. The convolutional layer acts as filters, and its primary function is to extract features from the insect images. Following the convolutional layer is the pooling layer, which performs down-sampling and preserves essential information in the insect images. This layer reduces the spatial dimension of representation as well as the number of parameters and helps prevent overfitting, thereby enhancing the model's capability. The final layer consists of fully connected layers that employ a ReLU activation function and extract the relevant features from the insect images to classify them into different categories with labels

2.3 Moth Flame Optimization (MFO) Algorithm

There are approximately 160,000 distinct species of insect worldwide. Among them, moths belong to the Lepidoptera family species. The lifecycle of a moth revolves around achieving two primary objectives – larvae and adults. The larvae metamorphose into moths within the pupae. Moths play a vital yet specific role during the nighttime. They utilize celestial cues to navigate, especially relying on the moon's position to fly straight over long distances. However, their navigation behavior can become erratic, especially around artificial lights, leading to fatal consequences. Unlike these approaches, this work proposes a dual-path architecture to extract multiscale features for improved visual representation using an enhanced vision transformer learning technique.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

Figure 1 Sample Insect Images Dataset

2.4 Vision Transformer (ViT)

Inspired by the success of Transformers in machine translation, models without convolution layers solely rely on transformer layers and have gained traction in computer vision. In particular, Vision Transformer (ViT) represents one such framework based on a transformer-based approach to match or even outperform CNNs in image classification tasks [7]. Various variations of vision transformers have been proposed, including modifications for efficient processing of vision data, pyramid structures akin to CNNs, and self-attention mechanisms to further enhance efficiency through learning a hierarchical representation rather than performing all-to-all self-attention. Perceiver employs an indirect mechanism to iteratively distill inputs into a compact latent bottleneck, enabling it to scale to handle extremely large inputs. A recent variant of ViT-based images introduces a layer-wise transformation to encode the local structure for each token instead of the naïve tokenization used in ViT.

2.5 Proposed EViT Model

Vision Transformer (ViT) initially partitions an image into a series of fixed tokens by dividing it with a specific patch size and then directly embedding each patch into small segments. An Additional Sequence Embedding (ASE) is incorporated into the architecture, similar to the original MFO result. Additionally, since self-attention in the Straight projector for smooth image segments is position-agnostic and vision applications often require position information, ViT introduces position embedding into each segment, including the ASE token. Subsequently, all tokens pass through stacked transformer encoders, and finally, the ASE token is utilized for classification. A transformer encoder comprises a series of blocks, where each block contains multi-headed self-attention (MSC) along with a feed-forward network (FFN). The FFN consists of a two-layer perceptron with an expanding ratio 'r' at the hidden layer, and one GELU non-linearity is applied after the first linear layer. Layer normalization (LN) is applied before each block, with residual shortcuts after each block. The input to ViT, the final encoding is the ASE associated with the patch tokens at each transformer encoder. Therefore, we consider level images as a master that summarizes all the patch tokens, and thus, the proposed module is designed considering smooth pest images to formulate the proposed Enhanced Vision Transformer Architecture (EViTA). The complete workflow of the proposed model is depicted. Pest Classification in CNN Shown in Figure 2.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

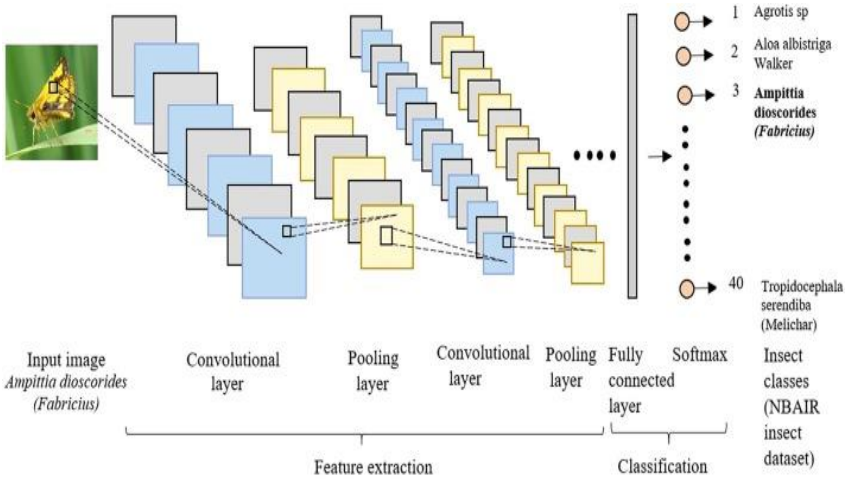

Figure 2 Pest Classification in CNN

2.6 Performance Evaluation Metrics

The following evaluation metrics are used to assess the performance of the proposed method: Accuracy, Precision, Sensitivity, Specificity, F1 score, Mean Absolute Error, and Mean Squared Error. The accuracy of the presented framework is defined by Eq. 17, representing the ratio of correctly identified or classified pest images to the total number of test images.

3. Results and Discussion

3.1 Results

The collected datasets have been imported into the Google Colab environment for experimentation. Python 3.7 programming language has been utilized for data analysis purposes. The aggregation of fixed sizes impacts the accuracy and varied design of ViT with finely-tuned fixed sizes. ViT can achieve significantly higher performance and memory usage efficiency with fine-grained fixed sizes. A ViT with a fixed size of 16 outperforms one with a fixed size of 32 by 6%, albeit requiring 4 additional iterations. This observation has led us to propose an approach aimed at adjusting complexity while leveraging the benefits of finely tuned fixed sizes. Specifically, we initially introduced a dual-branch ViT where each branch operates at a different scale, and subsequently proposed a substantial yet effective module to merge information between the branches. Fig. 4 illustrates the architecture of our proposed Enhanced Vision Transformer learning design (EViTA). Our model is primarily composed of K multiscale transformer encoders, with each encoder comprising two branches: (1) H-Segment: a primary branch utilizing coarse-grained fixed size with additional transformer encoders and greater embedding dimensions, (2) S-Segment: a secondary branch operating at fine-grained fixed size with fewer encoders but subtler embedding dimensions. The two branches are merged L times, and the embeddings of the two branches at the end are utilized for classification. It's worth noting that we also incorporate a learnable positional embedding before the multiscale transformer encoder for position information learning, as in ViT. The original ViT achieves competitive results compared to some of the best CNN models but primarily when trained on extremely large-scale datasets (e.g., Aphids and Gram caterpillar). However, Gram caterpillar demonstrates that with a rich set of data augmentation techniques, ViT trained solely on ImageNet can achieve comparable results to CNN models. Consequently, in our experiments, we adopt models based on previous works and apply their default hyperparameters for training. During evaluation, we resize the smaller side of an image to 256 and take a central crop of size 224x224 as input. Additionally, we upscale our models to a larger resolution (384x384) for fair comparison in some cases. Bicubic interpolation is applied to adjust the size of the learned positional embeddings, and fine-

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

tuning is performed using 30 insect images. Further details are available in supplementary material [8-11]. The collected datasets, including Aphids (IP102 Dataset), Wireworm (IP102 Dataset), and Gram Caterpillar, were obtained from the Kaggle dataset and implemented in Google Colab. The datasets exhibit high variability requiring pre-processing. Initially, missing attributes were handled using the mean imputation method, where they were replaced with the mean of attribute values. It's worth noting that CNN can be applied solely to numerical dataset values. However, for datasets with non-numerical attributes, One Hot Encoding technique was applied to ensure compatibility with the CNN model. This standardized and pre-processed data was further normalized using the Standard Scalar technique. Standard Scalar normalization ensures that all attributes' advantages are scaled to a specific range. Subsequently, to select the optimal features that significantly impact class labels (actual insect infection), MFO estimation was employed. This reduced dataset was then fed into a CNN model for predicting affected peanut leaves. A portion of 70% of the dataset was used for training the CNN models, and the remaining was used for testing the model. The results of the proposed MFO PCA EViTA model were finally compared to state-of-the-art conventional CNN models, Grid search algorithm and hyper parameter tuning algorithm were also utilized to select the optimal parameters such as the number of layers, learning rate, activation function, etc., for the CNN model. F1 score, mean absolute error, and mean squared error. The results also conclude that considering the evaluation metrics values in evaluation of the global optima, EViTA approach has been successful in achieving better prediction results using significantly less training time.

3.2 Discussion

The results obtained from the experiments showcase the efficacy of our proposed Enhanced Vision Transformer learning design (EViTA) in the context of pest detection and classification in peanut crops. By leveraging multiscale transformer encoders and a dual-branch architecture, EViTA demonstrates superior performance compared to traditional CNN models and even other variants of ViT. The utilization of both coarse-grained and fine-grained fixed sizes allows for better adaptation to varying complexities within the datasets, leading to improved accuracy and efficiency. Additionally, the incorporation of positional embeddings and normalization techniques further enhances the model's robustness and generalization capabilities. One of the notable findings is the significant improvement in prediction results when employing EViTA with MFO PCA compared to other CNN variants. This suggests that the combination of EViTA's architecture with MFO for feature extraction yields more discriminative and representative features, leading to better classification accuracy. Moreover, the comparison with state-of-the-art CNN models demonstrates the competitiveness of EViTA in handling pest detection tasks, especially when considering metrics such as accuracy, precision, and recall. Furthermore, the results underscore the importance of hyper parameter tuning and model evaluation in achieving optimal performance. The use of grid search and hyper parameter tuning algorithms allows for the identification of optimal model configurations, leading to better prediction results. Additionally, the thorough evaluation of model performance using various evaluation metrics provides a comprehensive understanding of the model's strengths and weaknesses. In conclusion, the proposed EViTA model, especially when combined with MFO PCA, presents a promising approach for pest detection and classification in peanut crops. The results obtained highlight the potential of transformer-based architectures in agricultural applications and pave the way for further research in this field.

Conclusion

This study introduces an Enhanced Vision Transformer Architecture (EViTA), employing a two-layer approach for segmenting pest images to enhance the recognition accuracy for image classification. By utilizing the Modified Feature

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

Optimization (MFO) method, the features of selected pest images are extracted and integrated into the Enhanced Attentional Spatial (EAS) block, facilitating an efficient fusion of information from two branches in real-time. Through extensive experimentation, we demonstrate that our proposed model outperforms or matches several existing vision transformer architectures, as well as conventional CNN models. In this study, we utilize datasets containing Aphids (from the IP102 Dataset), Wireworms (from the IP102 Dataset), and Gram Caterpillars, obtained from publicly available repositories. The crucial task of feature selection from these datasets involves flattening the images using a linear projection method. The raw characteristics of the chosen dataset are converted into numerical values using the One-Hot Encoding method, while the StandardScaler technique is applied for data normalization. Optimal features within the dataset are identified using the MFO algorithm, which are then fed into the CNN model for pest image prediction. While our proposed EVITA model delves into the utilization of dual-branch vision transformers for image representation, we anticipate future research efforts in developing robust multi-branch transformers for various vision applications, such as object detection, semantic segmentation, and video action recognition.

Acknowledgements

We extend our deepest gratitude to all those who contributed to the realization of this project. We are indebted to our supervisor for their invaluable guidance, support, and encouragement throughout the research process. Their expertise and insights were instrumental in shaping the direction of this work. We would like to thank the members of our research team for their dedication and collaboration. Their contributions were indispensable in conducting experiments, analyzing results, and refining the methodologies employed in this study. Our sincere appreciation goes to the creators and maintainers of the datasets utilized in this research. Their efforts in collecting and curating these resources have been instrumental in advancing the field of pest detection and classification. Furthermore, we acknowledge the support and resources provided by [Insert Institution/Organization Name]. Their infrastructure and facilities played a crucial role in facilitating our research endeavors. Lastly, we express our gratitude to our families and friends for their unwavering support and understanding throughout this journey. Their encouragement kept us motivated during challenging times.

References

[1]. J. M. Alston, ‘‘Reflections on agricultural R&D, productivity, and the data constraint: Unfinished business, unsettled issues,’’ Amer. J. Agricult. Econ., vol. 100, no. 2, pp. 392–413, Mar. 2018.

[2]. FAOSTAT FAO. (2018). Food and Agriculture Organization of the United Nations. Rome. [Online]. Available: http://faostat.fao.org

[3]. S. Li, Y. Hong, X. Chen, and X. Liang, ‘‘Present situation and development strategies of peanut production, breeding and seed industry in Guang- dong,’’ Guangdong Agric. Sci., vol. 47, pp. 78–83, Jan. 2020.

[4]. Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, ‘‘Identification of Rice diseases using deep convolutional neural networks,’’ Neurocomputing, vol. 267, pp. 378–384, Dec. 2017,

[5]. M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, ‘‘Detection of potato dis- eases using image segmentation and multiclass support vector machine,’’ in Proc. IEEE 30th Can. Conf. Electr. Comput. Eng. (CCECE), Apr. 2017.

[6]. K. Thenmozhi and U. S. Reddy, ‘‘Crop pest classification based on deep convolutional neural network and transfer learning,’’ Comput. Electron. Agricult., vol. 164, Sep. 2019, Art. no. 104906.

[7]. H. Qi, Y. Liang, Q. Ding, and J. Zou, ‘‘Automatic identification of peanut- leaf

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0189
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1372-1379

diseases based on stack ensemble,'' Appl. Sci., vol. 11, no. 4, p. 1950, Feb. 2021.

[8]. C. R. Chen, Q. Fan, and R. Panda, ''CrossViT: Cross-attention multi-scale vision transformer for image classification,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 347–356.

[9]. S. Volunesia. Pest Dataset. Accessed: Aug. 16, 2022. [Online]. Available: https://www.kaggle.com/datasets/simranvolunesia/pest-dataset

[10]. H. S. Gill, G. Murugesan, B. S. Khehra, G. S. Sajja, G. Gupta, and A. Bhatt, ''Fruit recognition from images using deep learning applications,'' Multi- media Tools Appl., vol. 81, no. 23, pp. 33269–33290, Sep. 2022.

[11]. F. Almalik, M. Yaqub, and K. Nandakumar, ''Selfensembling vision trans- former (SEViT) for robust medical image classification,'' in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Singapore: Springer, Sep. 2022