3rd International Conference on Computational Sciences and Technologies

Deep Transfer Learning Inspired Automatic Insect Pest Recognition

Mohammad Kamran Khan and Muhammad Obaid Ullah
Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan
*Email: mohammad.kamran@students.uettaxila.edu.pk
†Email: obaid.ullah@uettaxila.edu.pk

***Abstract*—Agriculture is not only the source of living for many living organisms but is also the backbone of many economies. Since insect pest diseases knock down the growth and production of agricultural resources, there is a need to vanish the insects through pesticides after their accurate recognition. Fast and effective algorithms for insect pest recognition are now possible on account of improvements in computer vision and artificial intelligence (AI). The deep transfer learning models (DTLMs) based insect pest recognition system is proposed in this paper. Two deep transfer learning models i.e., InceptionV3 and VGG19 are applied in this work. Pre-processing is used to locate areas containing the most relevant features. The IP102 dataset launched in 2019 is used in this research and a comparison is carried out to check the performance of the applied models. Experimental results show that the applied models outperform existing insect classification algorithms in terms of accuracy on large datasets.**

***Index Terms*—Insect pest recognition, Deep Neural Networks, Transfer learning, Saliency map.**

I. INTRODUCTION

Agriculture is the most important component of Pakistan's economy. The majority of the population is dependent on this industry, either directly or indirectly. It accounts for half of the employed labor force and is the main source of foreign exchange profits, contributing around 24 percent of GDP [1]. As insect pests stifle the growth and output of agricultural resources, pesticides must be used to eradicate the insects after precise identification. Various methods based on machine learning have been proposed in the recent past to accurately classify insects/pests. In 2015, Limiao Deng et al. [2] used the Difference of Gaussian (DoG) as a pre-processing technique to enhance the images. Local Configuration Pattern (LCP) the algorithm was used for feature extraction and for classification a Support Vector Machine (SVM) was trained to achieve a recognition rate of 89%.

In 2015 Hulya Yalcin et al. [3] proposed that the use of pheromone traps prevent huge reproduction because male insects are drawn to the traps and are unable to mate with female insects. While there are more than fifty distinct species of insects in Turkish forests, the most commonly encountered insect in agricultural regions is the green worm (a.k.a Heli- coverpa armigera), for which governments spend millions of dollars. The goal of this study is to create a robust algorithm for detecting and recognizing insects in a pheromone trap under various lighting circumstances. The visual sequences captured by the camera are first segmented by background extraction and machine learning algorithms. We put our method to the test using image sequences of insects collected in a simulated trap environment, namely a sticky pad in a corner of ITU's Visual Intelligence Laboratory.Local binary patterns (LBP) outperformed the other three feature extraction methods in classifying the insects out of four.

In computer vision, a saliency map of an image shows how important a pixel is to the human visual system. In 2018, Limiao Deng et al. [4] proposed a computer vision and machine learning based technique. Firstly, saliency maps were extracted via the Saliency Using Natural Statistics Model (SUN). Invariant features were extracted using Scale Invariant Feature Transform (SIFT) which were subjected to SVM for classification on a custom dataset with 10 categories to achieve an accuracy of 85.5%. In 2019, a new dataset for pest recognition and classification was put forward in Xiaoping Wu et.al. [5] The dataset comprised of 75000 images categorized into 102 classes with approximately at least 250 images per class. The author also evaluated the dataset for various machine learning approaches with different feature sets e.g., SIFT, Speeded up Robust Features (SURF) and Gabor using SVM and K-nearest Neighbour (KNN) as classifiers. Deep learning (DL) based models including Alexnet and Resnet were also trained and evaluated.They also summarized both machine learning and deep learning algorithms performance. In the past few years, deep learning models have out-paced traditional methods in computer vision such that, like the past state of pest recognition, involved meticulously creating hand- crafted feature extractors. Deep learning provides a hands- off approach that allows us to automatically learn important features directly from the image data. In 2019 Md. Imran Hossain et al. [6] proposed an insect recognition technique to avoid product loss of crops due to insect diseases in Bangladesh. They used a custom dataset with 6 classes and a total of 1600 images to train a Convolutional Neural Network (CNN). The Model was composed of 4 convolutional layers, two fully connected layers, and a dropout layer with a ratio of 25% to avoid over-fitting problems. 93.4% accuracy was reported by the author on validation data. The research aimed to help the farmers to get an immediate solution to the antidote of disease.

A CNN model consisting of attention blocks and a ResNet backbone was proposed by Yoon Jin Park et al.[7] was trained

ISBN # 978-969-23372-2-9

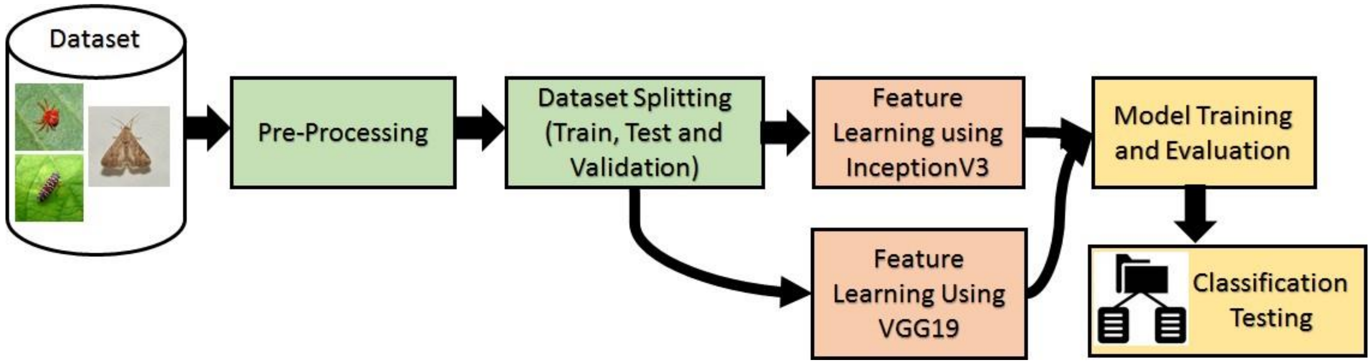

Fig. 1. Block diagram of the proposed methodology.

on an insect, database gathered from Atlas of Living Australia (ALA)3. The dataset consists of 123 insect categories with approximately 35000 images gathered in indoor and open-air environment. The results show that the proposed CNN model attained higher accuracy than other techniques. In 2018 Denan Xia et al. [8] proposed a multi-classification model based on the convolutional neural network (CNN) to solve the problem. The first 16 convolutional layers of the VGG-19 model were used to extract features and a Region Proposal Network (RPN) was used to classify and localize the insect. Large pictures were obtained by photographing open environments and collecting images online through Baidu and Google. Following that, image pre-processing and data augmentation were used to create their own dataset, dubbed’’MPest”. In 2018 Maxime Martineau et al. [9] turned the pest assignment into an image-based pattern recognition issue by recognizing the insect from a photo to make it easier. To solve this challenge, state-of-the-art deep convolutional architectures are used in this research. VGG-16 instances pre-trained on ImageNet 1000 are used in the fine-tuning CNN studies. In addition, they discovered that only training VGG16’s final convolutional block is sufficient to generate a virtually optimum solution. As a result, ImageNet-1000 characteristics can be concluded to be adequately generic up to a high level of abstraction. In 2020 Enes Ayan et al. [10] proposed that insects are a major cause of considerable losses in wheat, rice, corn, sugarcane, soybeans, chickpeas, and potatoes, among other crops. On the publically available D0 dataset with 40 species, seven distinct CNN models that had already been trained were used adjusted, and re-trained with the help of proper transfer learning and finetuning methodologies. Later, to improve classification performance, the best three CNN models (MobileNet, Xception, and InceptionV3) were combined by the maximum sum of probabilities method. Two datasets were used in the procedure, which shows the best accuracy for small dataset (95.15%) and less accurate for the IP102 dataset (67.13%).

In 2020 Fuji Ren et al. [11] developed a new and simple structure, called feature reuse residual block that mixes features from a residual block’s input signal with the residual signal. By learning half of a feature and reusing half of it, each feature reuse residual block improves representation capacity. The feature reuse residual network (FR-ResNet) was created by stacking the feature reuse residual block and tested on the IP102 benchmark dataset. In 2020 Nour Eldeen M.khalifa et al. [12] a deep transfer learning-based insect pest recognition system is proposed. In this study, the IP102 insect pest dataset was chosen. The IP102 dataset, which contains 75000 images and they used only 27500 images of 102 insect pest classifications, is one of the largest datasets for insect pests and was launched in 2019. Some DTLMs used in this research which is AlexNet, squeezNet, and GoogleNet, and the augmentation technique is used to make a powerful model and solve the overfitting problem. They evaluate the model by different parameters like F1 score, accuracy, and precision. Only the alexnet shows 89.33% testing accuracy for using a subset of the IP102 dataset.

In 2020 Qi-Jin Wang et al. [13] proposed a large dataset named is Pest24 that contains over 25000 annotated images taken in a field by pest trap and imaging system. Pest24 involves a total of 24 types of common pests, the majority of which devastate field crops. For the detection of pests, they used different DL techniques, and obtained promising results for actual-time monitoring field crop pests. In 2019 Xiaoping Wu et al. [5] proposed a survey and employ two classifiers for recognition problems and the dataset is available at https://github.com/xpwu95/IP102. They make the dataset and pre-trained models freely available. On the IP102 dataset, the SVM and KNN classifiers performed well in classification using several assessment criteria. The representations are grouped into two categories: handmade and deep features. In 2020 Loris Nanni et al. [14] presented a classifier based on the combination of saliency approaches and CNN was described. Saliency methods are popular image processing algorithms that highlight the image’s most important pixels. They achieved high accuracy for small datasets and low accuracy for large which is 92% and 62% respectively. They used distinct techniques to extract saliency maps from the images in our dataset. In future research, we’ll see if other saliency methodologies can help us improve the outcomes even more.

In this paper, a deep learning approach based on transfer learning for insect pest recognition has been proposed.

Fig. 2. Sample images from state-of-the-art IP102 insect pest recognition dataset

Firstly, pre-processing was done using fine-grained saliency maps. The pre-processed images were split into three sets namely: training, validation, and testing. Using the advantages of transfer learning, two deep learning models were trained including VGG-19 and InceptionV3 pre-trained on imageNet i.e., the largest image classification dataset with 1000 classes. For models training, the IP102 dataset has been utilized in this study as it is the most widely used challenging and diverse dataset for insect pest recognition. Most of the studies utilizing this dataset are using subsets of this dataset. However, in this technique, models under study were trained and evaluated on a complete dataset.

The rest of this paper is organized as follows. In Section II, the proposed methodology is described in detail. In Section III, the results and discussion are explained. In the end, Section IV concludes the paper.

II. PROPOSED METHODOLOGY

Using different Deep Transfer Learning Models for insect recognition tasks achieved high comparative accuracy. A complete block diagram of the proposed methodology is shown in Fig. 1.

A. Dataset Selection

IP102 dataset contains 102 classes/species of insects having over 75000 images of both environment images of field and lab images. Some of the class images are more occluded and cluttered environment, unbalanced data, and having intra-class problem as some of the images in classes has different appearance due to their life-cycle. Therefore it is a challenging insect pest recognition dataset and publicly available in [5].Sample of some classes of the dataset are shown in Fig. 2.

B. Pre-Processing

The preprocessing is so important because the region of interest and background are merged and we have to use a

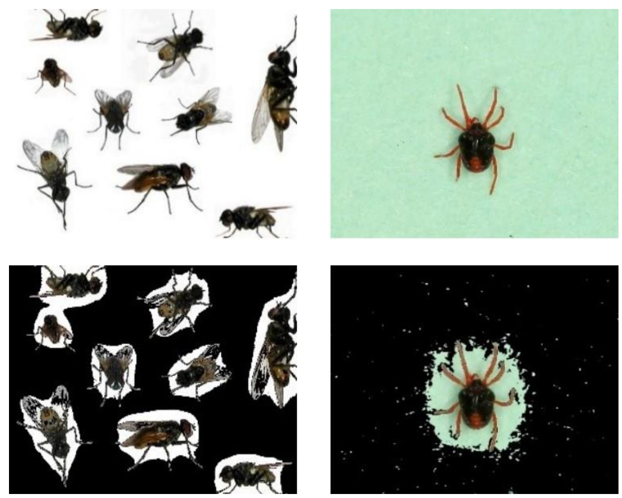

Fig. 3. Samples images of dataset before pre-processing (top row) and after pre-processing (bottom row)

Fig. 4. VGG19 architecture

preprocessing technique to extract the Region of interest (ROI). Generally, the dataset images have a cluttered and occluded environment. Once training and testing of different deep learning models for the recognition, the data must be preprocessed so that the most relevant features are located. In this proposed task the fine-grained saliency is used for preprocessing and it is the most innovative part of the proposed work. This method calculates saliency based on center-surround differences. High-resolution saliency maps (heat maps) are generated in real-time by using integral images.

ISBN # 978-969-23372-2-9

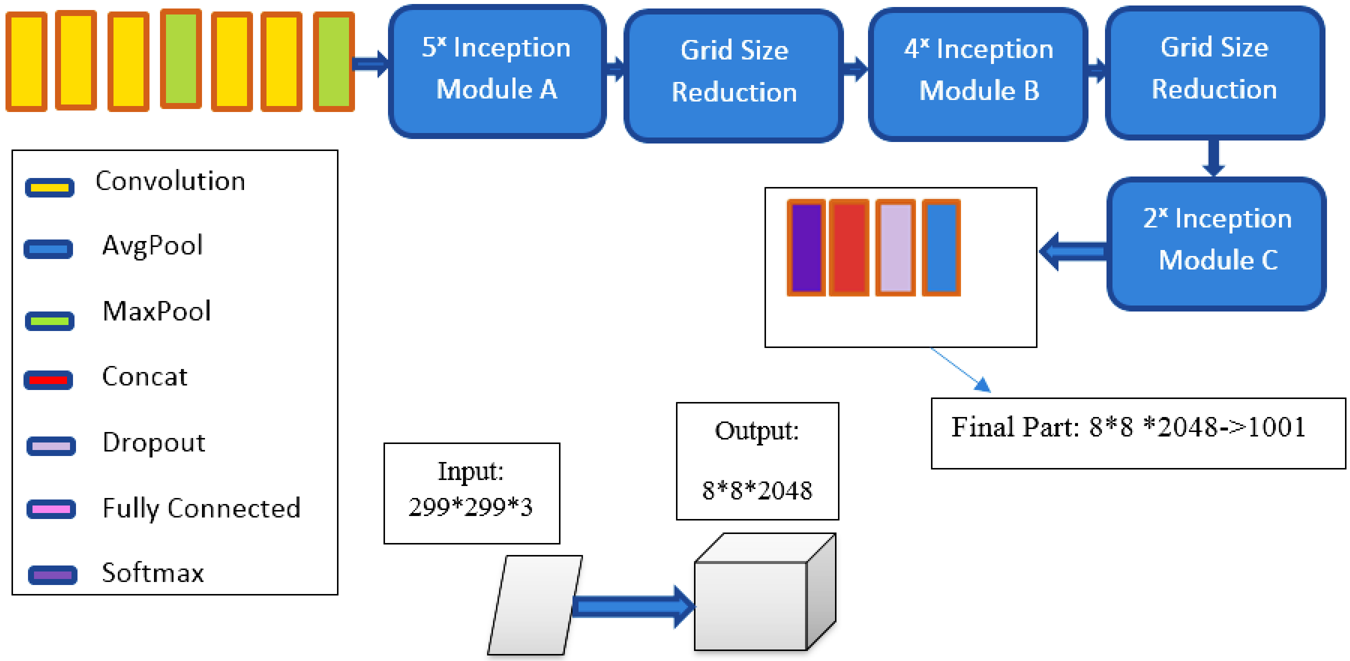

Fig. 5. A general InceptionV3 architecture

The distance of each pixel to the rest of the pixels in the same frame is calculated as follows:

$$SALS(I_s) = \sum_{j=1}^{N} |I_s - I_j| \qquad (1)$$

Where I_j is the value of pixel j (0,255), and its expanded form can be written as:

$$SALS\ (I_s) = |I_s - I_1| + |I_s - I_2| + \ldots + |I_s - I_N| \quad (2)$$

Here N is the current frame's total number of pixels .Then we can further restructure the formula as:

$$SALS(I_s) = F_n \sum_{j=1}^{N} |I_s - I_n| \qquad (3)$$

Here F_n is the frequency of I_n. Fine-grained saliency is found better than saliency methods used in [14]. Fig.3 shows some preprocessed images of the IP102 dataset.

C. Deep Transfer Learning Architectures

Traditional image processing techniques produced acceptable results and performance when it came to detecting insect pests utilizing insect images. As deep learning has transformed the field of artificial intelligence, Image classification is a type of computer vision and the detection and recognition of objects Learning at a deeper level is the most recent technology that has made significant impact advancements in the field of artificial intelligence [12], [15]. While Transfer Learning (TL) is a machine learning technique in which a model created for one job is utilized as the basis for a model on a different task. The CNN model is a type of deep artificial neural network often employed in image analysis. By treating an image as a volume, it learns spatially linked features. It contains several volume specialized layers that alter the volume of an image in various ways. Much of the processing for classifying an image is done by a convolutional layer, within a convolutional layer, there is a succession of kernels that slide or convolve over an image volume. One of the most notable advantage of CNNs is that as their training progresses, these kernels can recognize textures, forms, colors, and other visual properties [16].

Recently in a few years, the competition's image detection/classification task accuracy rate was considerably improved because of numerous developments in deep convolutional neural networks. The annual challenges improved significantly when CNN pre-trained models were used. Many pre-trained models were introduced as AlexNet, VGG-16, VGG-19, GoogleNet, ResNet, Xception, InceptionV3, and DenseNet. In this proposed work, it is shown how to recognize insect pests using deep transfer learning models. The IP102 dataset contains both types of lab and field images. For the automatic insect pest recognition problem, this work presented different deep transfer learning techniques like VGG-16, Alexnet, Resnet50, InceptionV3, and VGG-19. The training and testing session proved that InceptionV3 and VGG-19 gives better results for the proposed task. Inception-v1 is a deep convolutional architecture that was first introduced as GoogLeNet in [17]. The Inception architecture was later modified in several ways, the first of which was the introduction of batch normalization (Inception-v2). In the third iteration, which will be referred to as InceptionV3 in this paper, new factorization ideas are presented. Fig. 5 shows the inceptionV3 architecture. The inceptionV3 model is modified through different parameters like convolution factorization two 3*3 convolutions are replaced by one 5*5 convolution

ISBN # 978-969-23372-2-9

TABLE I

PERFORMANCE COMPARISON OF PROPOSED TECHNIQUE WITH EXISTING DEEP LEARNING METHODS USING IP102 DATASET

Ref.	Year	Dataset	Methods	Accuracy*
[14]	2020	IP102	Ensembles (AlexNet, GoogLeNet, ShuffleNet, MobileNetv2, DenseNet201).	61.93 %
[10]	2020	IP102	GAEnsemble (CNN models, InceptionV3, Xception, and MobileNet)	67.13 %.
[11]	2019	IP102	Feature reuse residual network (FR-ResNet)	55.24 %
[5]	2019	IP102	Alexnet	28 %
			GoogleNet	40 %
			VGGNet	48 %
			ResNet	49%
This work	2021	IP102	InceptionV3	81.7 %
			VGG19	80 %
*Some works use a subset of IP102 but accuracy of these works is based on larger IP102 dataset.				

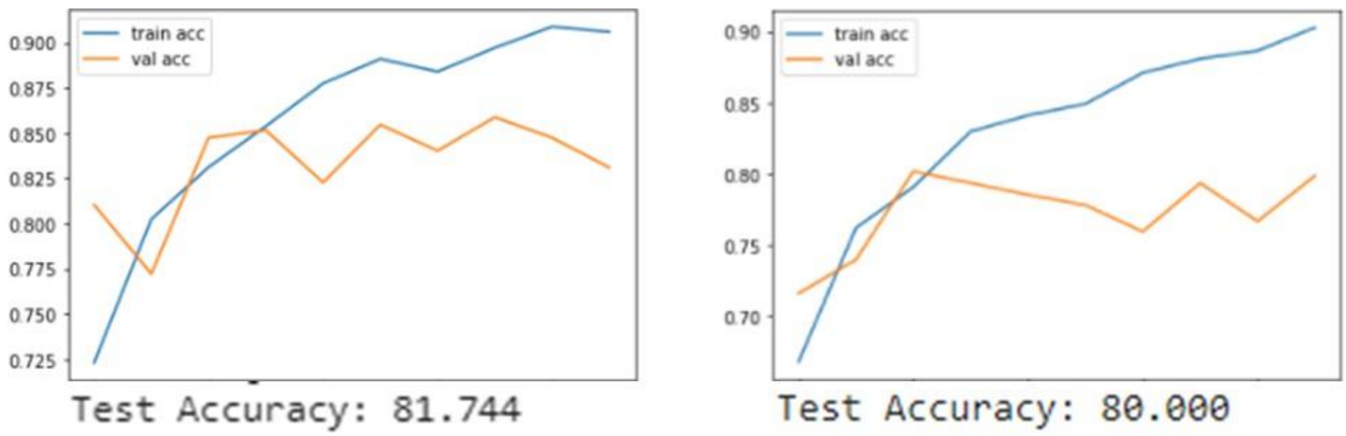

Fig. 6. The graph of training and testing accuracy of InceptionV3 and VGG19 models

Which reduced the number of parameters by 28% without affecting the efficiency. Instead of utilizing two auxiliary classifiers on the top of the final 17*17 layer, only one auxiliary classifier is employed and the feature map reduction is traditionally done through max pooling, like in AlexNet and VGGNet. However, either max-pooling followed by Conv layer is too greedy, or Conv layer followed by max-pooling is too expensive. An effective grid size reduction is proposed here [17]. Similarly, the VGG is a deep CNN used to classify images. AlexNet was released in 2012, and it improved on traditional Convolutional neural networks. We can assume VGG as a successor to AlexNet, but it was developed by a different group at Oxford University called the Visual Geometry Group, hence the name VGG. It carries and improves on some of its predecessors' ideas, and it uses deep Convolutional neural layers to improve accuracy. The VGG19 model is a variation of the VGG model that has 19 layers in total (16 convolution layers, 3 fully connected layers, 5 MaxPool layers, and 1 SoftMax layer). Fig. 4 shows the architecture of VGG19. VGG19 is also a deep CNN model used to classify images.

D. *Training and Testing*

This work is conducted on a local computer that uses window 8.1 pro operating system and equipped with google colaboratory GPU. Google colaboratory permits anyone to execute python code via browser. While the total number of images is over 75000, divided into training data, test data, and validation data according to the information given in the dataset in text form. The image processing deep learning frameworks Tensorflow and Keras are employed. In both the training and testing dataset, we employed categorical mode with three-channel color mode and a batch size of 32. The model provided in the study performs admirably. After 20 epochs, the inceptionV3 employed in this study achieved 90.5% accuracy on the training data set and 81.7% accuracy on the testing dataset. On the IP102 dataset, this corresponds to a high level of accuracy. Similarly when the VGG19 is employed the training accuracy and testing accuracy are 90% and 80% achieved respectively.

III. RESULTS AND DISCUSSION

In comparison to the previous papers, this work provides the highest level of accuracy. To characterize the performance, use the formulas below for the proposed architecture's classification. Table 1 shows the comparison of results.

$$\text{Accuracy}=\frac{T_P+T_N}{T_P+T_N+F_P+F_N} \quad (4)$$

Where T_P is True Positive, which is a set of all the predicted classes which are accurately classified as their True classes; T_N is True Negative, which is set of all the predicted classes that are not classified to their irrelevant classes; F_P is False Positive, which contains all the predicted classes incorrectly classified as an actual class; and F_N is False Negative, which is set of all the predicted classes that are not classified to their actual classes. The training and testing accuracy of the proposed models are shown in Fig.6. It achieved the highest training and testing accuracy to the state of the art techniques. The training and validation accuracy of the inceptionV3 model are represented by two lines of blue and orange color on the left graph, respectively, while the training and validation accuracy of the VGG19 model is represented by two lines of blue and orange color on the right graph.

I. CONCLUSION

Agriculture is the source of food and money. But insect pest diseases adversely affect its expansion and productivity. Before the adult stage, the detection of insect pests through advanced computer and artificial intelligence algorithms helps avoid losses. This research has presented an insect pest recognition system based on (DTLMs). Before training, fine- grained saliency is used to extract ROI from a cluttered environment. Comparatively higher accuracies of 81.7% and 80% are achieved using inceptionV3 and VGG19 models respectively. In future work, different saliency strategies can be explored to improve the large datasets. More correct segmentation techniques can also be applied to enhance performance. This approach of automatic recognition would eliminate the need for agricultural experts to personally analyze insect types, which are time-consuming, less accurate, and expensive. Future studies might include developing an android system that can recognize the insect and deliver a better antidote for the attack.

REFERENCES

[1] Government of pakistan, *Pakistan Bureau of Statistics*, 2021.

[2] L. Deng and R. Yu, "Pest recognition system based on bio-inspired filtering and LCP features," in *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pages 202-204,2015

[3] H. Yalcin, "Vision based automatic inspection of insects in pheromone traps," in *2015 Fourth International Conference on Agro-Geoinformatics (Agro-geoinformatics)*, pages 333-338, 2015.

[4] L. Deng, Y. Wang, Z. Han, and R. J. B. E. Yu, "Research on insect pest image detection and recognition based on bio-inspired methods," vol. 169, pages 139-148, 2018.

[5] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, and J. Yang, "Ip102: A large-scale benchmark dataset for insect pest recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8787-8796, 2019.

[6] M. I. Hossain, B. Paul, A. Sattar, and M. M. Islam, "A Convolutional Neural Network Approach to Recognize the Insect: A Perspective in Bangladesh," in *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages. 384-389, 2019.

[7] Y. J. Park, G. Tuxworth, and J. Zhou, "Insect classification using Squeeze-and-Excitation and attention modules-a benchmark study," in *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3437-3441 2019.

[8] D. Xia, P. Chen, B. Wang, J. Zhang, and C. J. S. Xie, "Insect detection and classification based on an improved convolutional neural network," vol. 18, no. 12, pages 4169, 2018.

[9] M. Martineau, R. Raveaux, C. Chatelain, D. Conte, and G. Venturini, "Effective training of convolutional neural networks for insect image recognition," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 426-437 2018.

[10] E. Ayan, H. Erbay, F. J. C. Varçın, and E. i. Agriculture, "Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks," vol. 179, pages 105809, 2020.

[11] F. Ren, W. Liu, and G. J. I. A. Wu, "Feature reuse residual networks for insect pest recognition," vol. 7, pages 122758-122768, 2019.

[12] N. E. M. Khalifa, M. Loey, and M. H. N. J. J. T. A. I. T. Taha, "Insect pests recognition based on deep transfer learning models," vol. 98, no. 1, pages 60-68, 2020.

[13] Q.-J. Wang *et al.*, "Pest24: A large-scale very small object data set of agricultural pests for multi-target detection," vol. 175, pages 105585, 2020.

[14] L. Nanni, G. Maguolo, and F. J. E. I. Pancino, "Insect pest image detection and recognition based on bio-inspired methods," vol. 57, pages 101089, 2020.

[15] D. Rong, L. Xie, Y. J. C. Ying, and E. i. Agriculture, "Computer vision detection of foreign objects in walnuts using deep learning," vol. 162, pages 1001-1010, 2019.

[16] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE conference on computer vision and pattern recognition*, pages 3642-3649 2012.

[17] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1-9 2015.

ISBN # 978-969-23372-2-9