Hierarchical Classification of Insects using a Combination of Resnet and VGG Networks

MSc Research Project
Data Analytics

Priyal Narendra Patil
Student ID: x20193394

School of Computing
National College of Ireland

Supervisor: Dr.Giovani Estrada

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Priyal Patil
Student ID:	x20193394
Programme:	Data Analytics
Year:	2021-2022
Module:	MSc Research Project
Supervisor:	Dr.Giovani Estrada
Submission Due Date:	
Project Title:	Hierarchical Classification of Insects using a Combination of Resnet and VGG Networks
Word Count:	
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Priyal Narendra Patil
Date:	15th August 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Hierarchical Classification of Insects using a Combination of ResNet and VGG Networks

Priyal Narendra Patil
x20193394

Abstract

Image processing had been exponentially developed in the past few decades. Alongside the need and applications had been also rising. In our research, we approached the agricultural sector for image processing pests. As pests are most harmful to the crops and plants at any stage of the plant it is essential to detect them early as well as identify them in early stages which can help to tackle them with adequate measures. Our approach showcased the identification of the pest among 102 classes using VGG16 and ResNet models which had attained accuracies of around 64 and 65.44% respectively. And one further investigation an unique method of training of model involving the AugMix technique and random noise in the training data helped to improve the accuracy of the ResNet50 and VGG16 hybrid model. With the proposed methodology model was able to achieve the accuracy of the 75% on the test dataset of the IP102.

1 Introduction

Agriculture sector is the most important domain for any country and it plays a vital role in the nation's economy. Various approaches had been tested as well as under development for advancement in this same domain. Latest approaches which are artificial intelligence and machine learning-oriented solutions had been effectively showcasing their best performance(Karar et al.; 2021).Moreover, 'precision agriculture' is a term given to the approaches for agriculture domain in which AI and ML especially, from the planning of sowing seed to harvesting of crops all solutions are covered.

Protecting the crops from external factors like temperature or humidity or water content in the soil or many other, becomes important for practicing precision agriculture.Controlling the parameters is also another challenge that is capital intensive. Rather than optimizing the existing process, development in protecting the crops and their planning can be an easier and most suitable approach with which we can effectively increase the yield production with minimum capital investment. Biggest obstacle to be tackled is the protection of the crops at any stage of their growth period from insects had been observed across the globe (Kasinathan et al.; 2021).Factors like change in temperature or rainfall, etc also produce a noticeable effect on pests which furthermore increases the damage.

A deeper study in the damages caused by pest to crops and its economic loss across countries provided an enough amount of data on which approaches based upon AI and ML can be implemented.During the study, it showcased that the time of detection of the pest as well as its prevention measure further implemented plays a vital role in the crop's yield (Yang et al.; 2021) .Some study showcased that early detection of the pest with its correct type can create a noticeable impact on the yield of crops as well as planning of the process from sowing seed to harvest using AI and ML had significantly improved the process and provided a better quality of crops with increased yield. Some countries which are leading producers of the crops includes China, India, US, and Brazil. A study showcased that nearly 20 to 40% of yield produces is wasted due to pest worldwide and a global economy loss of around US $ 70 billion. Most recent pest outbreak can be studied of 2019-2021 locust infestation in East Africa, the Arabian Peninsula and the Indian subcontinent had been affected. A study proposed the possible reason for it as a Cyclone Mekunu in 2018 at Arabian Peninsula which initially spread locust to Iran, Pakistan, and India and then East Africa. Although spread had been automatically controlled due to the COVID-19 pandemic and its restrictions but this recent event provided concrete evidence of the danger associated with the crop pest and its vulnerability.

Many studies had been conducted in the agricultural domain related to pest in which comparatively less technologies are implemented which are using AI and ML for providing the solution. Effective as well as autonomous model for providing the forecast, or process optimization in real time or inventory/raw material management are some applications on which recent studies had been focused.As the approach based upon AI and ML can also be as effective as other domains, pest identification can be an unexplored domain in which our study is focused.

To implement the preventive measures or corrective measure for the invaded pest, its identification as well as its classification among large varieties plays a vital role. For identification of correct pest using AI and ML approach can be beneficial and save the crops by damage of that pest. An automated process of classification and identification of the pest image feed to the AI and ML model which in result provides us an exact name of the pest so that user can proceed with its corrective measure for saving crops from that identified pest is achieved in this research. By utilizing the proposed model in this research, it also helps user to reduce the pesticides and chemical which are supplied on the general bases. Basically, only those chemicals and pesticides are needed to be supplied to crops which is identified specifically at that particular location of field/farm.
.

1.1 Motivation

With the rise of around 1.05% annually of population, demand of the food had also been rising rapidly. For preparation of any kind of meal, directly or indirectly crops are involved, 11% of globe's land surface is used for production of the crops makes it a 1.5 billion ha. For feeding the world's population major countries involved in the agricultural sector are belonging to developing or under developed except few. Enormous numbers of the quantity in which food is produced makes it a one of the most important and crucial sectors. A small downfall in this sector can create huge loss economical as well as materialistic loss. It can also lead to food scarcity or poverty in a notion if not handled correctly. Apart from the activities in which human are involved, natural climate actions like wildfire, global warming, climate change, increase in temperature, etc are other causes which affects the agricultural sectors although in these types of natural climatic actions it is not in the hands of humans but the activities involving humans can be controlled. Our main proposed for selecting this domain as with the help of the AI and ML model, if the identification of the pest by the user using an image can create huge impact of the yield production as well as selective amount of the pesticides to be used upon the identified pest can increase the quality of the produced yield. Our study proposes the model which can identify the image of the pest and tell specifically the name of the pest detected. Considering the effectiveness and its outcome, our proposed approach can be beneficial to large number of users as well as on future development it can also be integrated with various tools like identification of crop diseases, measures for detected categories, etc. For detection of the pests, our approach utilize dataset named as IP102 which comprises of 75000 images belonging to 102 categories of pest. Further details about the dataset is discussed below followed by research question which provide the intention behind the conducting research.

1.2 Dataset

IP102 dataset was proposed in CVPR2019 by the author (Gomes and Borges; 2022) and it contains large number of images to train Machine Learning Model it becomes very helpful. It comprises of 102 classes of the pest and nearly about 75000 images which are labeled make it a perfect dataset choice for implementing in various machine learning models.

On a closer study of the dataset, it is observed that it is highly imbalance as well as interclass variance (Mohsin et al.; 2022).Images collected from the internet and among 8 categories they are categorized as Mango, Rice, Corn, Wheat, Beet, Alfalfa, Vitis, and Citrus with 10, 14, 13, 9, 8, 13, 16, and 19 insects.

Author (Li et al.; 2022) had also described about different approaches for feature extraction with several baseline classification numbers for deep learning models such as YOLOv3, ResNet50, SVM, KNN, etc. Results showcased a maximum accuracy of nearly 51% is achieved.

1.3 Research Question

Here this research tries to find the solutions to the following research questions,

- How can different techniques be combined to improve the classification accuracy?
- How might data Augmentation impact the overall classification Accuracy?

2 Literature Review

2.1 Classification of the Dataset IP102

To automate the process of the classification of the pest through a machine learning approach, models are essential for training for that specific task. For the easiest way to make it user-friendly, the image might be an option. The author (Wu et al.; 2019) provided the collection and, making of the IP102 dataset which is 102 classes of pest images data containing about 75000 images. Out of 75000 images of the pest, 19000 images are provided with the bounding box and the labeled images which shall affect the model's performance greatly. It had been a hierarchical taxonomy of the data in which it is sub-grouped in their respective categories. The prepared dataset had been observed for inter and intra-class variance imbalance and it had been also described by the author. It had been publicly available for the user to utilize and improve.

The rise in population had also produced an immense rise in food demand. As the food demand rises, it had led us to increase the productivity of the cultivation and production of the crop which makes a pest issue so much crucial. Pest can cause damage to crops which may damage the growth of the crop. Early detection and mitigation it makes essential for farmers. (S.-Y. Zhou Su, n.d.) provided an approach through which the image of the pest can be classified easily without any lengthy process. The author provided a machine learning model which is based upon SqueezeNet. A machine learning model is trained upon the IP102 dataset consisting of 102 classes of pests and the primary function of this model is to classify the given image of the pest into 102 classes of the pest. It had been also compared with various models such as ResNet101, MobileNetV3-large, ShuffleV2, EfficientNet, etc. It has been observed that the proposed model by the author which is the combination of two blocks had showcased higher accuracy among all. It had reached about 52.32% of accuracy without any augmentation involved.

Considering the importance of the detection of the pest as well as classification of the pest in their respective classes at the correct time so the preventive measures can be taken to save crops, the author (Nanni et al.; 2022) showcased an ensemble CNN model approach which is based upon different topologies like ResNet50, ShuffleNet, DenseNet201, mobileNetV2, and GoogleNet. The author altered it by random selection of the class of the pest from the IP102 dataset which is used for training the models. For the scaling factor, two new Adam algorithms which are based upon DGrad had been introduced by the author in this prosed approach. Making all the models a set of 5 CNNs had been trained upon the two well-known datasets as IP102 dataset and Deng (small). A remarkable accuracy had been observed by the ensemble approach as 95.52% on Deng and for IP102 it had reached 73.46% . To ensure the robustness and the behavior of the model on different data, it has been also tested on the medical image data and observed its classification outcomes.

As the ensemble approach and two block approach had showcased a good performance on the pest classification, the author (Ren et al.; 2019) proposed the feature reuse residual block-based approach which had benchmarked with the IP102 dataset. IP102 dataset is the 102-classes dataset of the pest which is utilized for comparing the model in this research by the author. Stacking of the feature reuse block in the model proposed in this research has provided a significant improvement in the classification of the pest. Combination of the approach along with CIFAR-10 or 100 or SVHN had also showcased an improvement in the performance of the classification of the pest images but the overall observation by the author stated about the effectiveness of the proposed Approached.

As the training of models plays a essential role in the performance of the classification, the data on which the model has been trained is also important. The majority of the research had been utilizing the IP102 dataset which had been reported for the high of data among its classes so the author (Gomes and Borges; 2022) had provided an alternate approach to provide a solution for that problem. IP-FSL which is based on the IP102 dataset had been proposed by the author of this research. Alongside the testing of this new proposed data as well as a better network for classification of the images of pests a prototypical network had been proposed. Based upon Kullback-Leibler divergence measure accuracy of about 86.33% for adults and 87.91% for early stages pests had been observed by the author through its proposed network as well as its new pest dataset.

2.2 Region of Interest Detection and Attention mechanism

Early detection of the pest in the crops can create a great impact on the crop's growth and production. For its detection, an image-based approach had been selected. Selecting the pest from the given image makes it a difficult task. The region of interest makes an essential part of the detection or classification of the pest. The author (Bhadane et al.; 2013) provided a machine vision-based approach to detect the region of interest using feature extraction by a machine learning model. The author also proposed a system to take the infected crops' images and process them for training the model. Majorly research had been focused upon the region of detection and the infected area computing by the model.

The testing of the approach had been conducted on the IP102 dataset divided into small and large parts. By using the proposed methodology, the model attained about 92.43% of accuracy for the small IP102 dataset whereas 61.93% for the large IP102 dataset (Nanni et al.; 2020).

The early detection, classification, and monitoring of the pest play a vital role in the crop's yield which is needed to take care of at a very early age of the crops. As the identification of the pest plays a much as important role as its classification or detection because once the pest is classified then only the proposed preventive measure can be taken. For the classification or detection approach machine learning and machine vision-based approach had been discussed in this research(Ranganathan Engineering College Institute of Electrical and Electronics Engineers) .The uniqueness of this approach as described by the author is the color-based image segmentation which makes the method efficient enough to detect the pest in any environment background image. Simulating the method had provide effective outcomes which were performed on Otsu's methods and edge detection.

As the appearance and the outside features are much more similar in some species of the pests it makes the machine learning model difficult to distinguish between them. The author (Liu et al., 2019) provided an approach based upon a region-based end-to-end approach named as PestNet classifying the pest in the multi-class classes. Proposed PestNet consists of three blocks which are Channel Spatial Attention (CSA) which is fused into a Convolution neural network (CNN) responsible for feature extraction followed by a Region Proposal network and lastly a Contextual Region Of interest layer for enhancing the accuracy of the overall model and classification. The proposed model had been tested upon MPD2018 which is a multiclass pest dataset containing more than 8oK of images with labeled pests of about 580K by an agricultural expert. Mean average Precision of about 75.46% had been attained.

The labeled dataset and its size of it play a vital role in the accuracy and performance of the model. The author (Wang et al., 2021) had proposed an AgriPest dataset which is specifically developed for the machine learning model to train from the detection and classification of the pests from the raw images. Over the seven years, AgriPest had captured about 49.7K images from four crops which shall be a total of 14 species of pests. Manually annotation of each image provided authenticity making the image data of about 264.7K of bounding boxes in the location of the pests. It had been structured such that it is easy for the machine learning model to utilize and train itself making it one of the most valuable and important datasets for the pests.

2.3 Hierarchal Classification for multiclass classification

Classification of the object or any kind of value had been a challenging task with the help of machine learning and artificial intelligence. As the number of classes increases the more difficult it becomes for the machine as well as humans to classify it. An alternate approach was developed which can be hierarchically distributed among the classes. In this research , the author (Silva-Palacios et al.; 2017) proposed a new method to classify the object with the hierarchy-based approach. A classifier that is based upon the hierarchy method is used in this approach which learns a tree-like hierarchy while classifying. For evaluating the proposed model some datasets are used to perform the classification which showcased promising outcomes as expected by the author and able to classify efficiently.

The author (Aly; 2005) provided the supervised classification algorithm which shall be a model trained on a labeled dataset. A survey on the classification methods had been showcased by the author. Some algorithms which are been used for the classification of the multiclass or binary classification is been studied and compared. Majorly, the author focused on the different techniques for the multiclass classification and its algorithm used as well as issues faced by them. (Raiyan Mohsin, 2022b) As the research was focused upon the IP102 dataset as the pest classification through machine learning and machine vision approach the author provided the deep learning approach. From 75000 images of the pests which were classified into 102 classes are used. Models like EfficientNetB5, InceptionV3, VGG19, ResNet50, and DenseNet121 are implemented upon the LIME-based XAI framework which is an explainable artificial intelligence framework. Among the five different models, DenseNet121 is performing best with a classification accuracy of about 46.31% to 95.36% .

For the classification and detection of the pest, seven different models were utilized by the author to compare them with each other and provide the classification's best outcome for the pest(Ayan et al.; 2020). Used seven models for classification are VGG-16, VGG-19, ResNet-50, Inception-V3, Xception, MobileNet, SqueezeNet. Initially, the models are used with the D0 dataset which is of 40 classes and the ensemble of the outcome provided the correct class of the input image. All the models are also fine-tuned with the pest datasets for their classification. GAEnsemble had achieved the highest classification among all others and reached about 98.81% which also justifies the robustness of the model. For the IP102 dataset which consists of about 102 classes had been achieved about 67.13% of the classifying accuracy.

For the multiclass classification labeled set of organized data had been a very difficult task. In this research, the author (D. Zhou et al., 2011) proposed a hierarchical Support Vector Machine (SVM) based approach which shall classify each node from the classifiers at its ancestors. Introducing the regulation that forces the normal vector to classify at each node using hyperplane orthogonal is been showcased in this approach. For validating the proposed approach, the text-based classification is showcased by the author.

2.4 Data Augmentation Techniques to improve the test accuracy

When it comes to image classification, data augmentation becomes the most important factor, especially when performing object recognition-related problems. Most of the real-world data may have been imbalanced and it greatly affects the accuracy.

AugMix is a data augmentation technique for improving the robustness and uncertainty proposed by (Hendrycks et al.; 2019). In his article, he shows how a small change in the data distribution can hamper the performance of the model. He shows that introduction of the distortion in the training images can fail to generalize the unseen images. He proposes AugMix methods over methods such as CutMix and Mixup where images from multiple classes are merged like collage or blended. However complaints about the computational cost and training time. In contrast to those techniques, AugMix applies simple augmentation techniques selected randomly on an image and makes a new image by overlaying it on top of another.

In most of the research, or the existing solutions, data augmentations are a must especially when the multi-class classification problem is considered. The Author (Shorten Khoshgoftaar, 2019) Compares the different approaches for data augmentation. There are several approaches to it, mainly classified as the Basic manipulation and Deep learning approach. The basic idea is to increase the training images so that models can learn from more information. This basic image manipulation consists of Color Space transformation, Geometric Transformation, kernel fitters, etc. which are most widely used since are easy to implement and can be easily inserted into the existing training pipeline. However Deep learning approaches on the other side methods such as Adversarial training, Neural Style Transfer, GAN data augmentation, etc are harder to implement. And consumes additional resources. The analysis presented by Shorten Khoshgoftaar shows that Cropping from geometric transformation was able to increase the accuracy of the model.

The Author (Perez and Wang; 2017) investigated the effectiveness of the data augmentation methods for deep learning they also like (Shorten Khoshgoftaar, 2019) investigates the 2 approaches traditional one and deep learning. And they compared the techniques on the problem of classification of the imagenet dataset. And the analysis shows an increase in the validation accuracy when any form of data augmentation is applied. They showed that the traditional approach is much better than the GAN approach. This shows that the additional efforts do not provide enough benefits as compared to the traditional ones. However, they have also proposed a novel approach called "Neural Augmentation" which had showed significant improvement. However, they have tested on the single dataset and on the single structure "SmallNet". But this study certainly enforces the belief of using traditional approaches for data augmentation.

Also, noise can help to improve the performance of the model, and that is what a GAN model does it takes the random noise as the input and generates an image using just enough features which can help to determine its classes. (Han et al.; 2019) combines the Noise to image GAN and the Image to Image GAN, in prior a noise is given as input to generate the additional training image later an image is given as input and the model generates another image. They have demonstrated a novel approach to tumor detection datasets and other medical images.

As discussed in this section, the deep learning approach does improve the result but also requires additional resources and is more complex to integrate into the existing training pipeline. And also, (Perez and Wang; 2017) shows that the traditional methods just work great out of the box. So looking at that, (Hendrycks et al., 2019) propose a simple approach that can help to increase the test/validation accuracy if there are large differences between the training and test/validation accuracy. Their main goal was to improve the Robustness and improve performance during the uncertainty. Here authors also discuss the impact of the noise in the training data and show that it does not improves the performance in the real world since the model learns the distribution of that random noise that had been added. In the AugMix approach images are Stretched or compressed on the X and Y axis, rotated as well as other warps had been added. And generates an image that is quite similar to the original one in the proposition. And showed consistently low errors on the prediction. And the same approach had been used in this work with the addition of noise in the training.

3 Methodology

For this research, Deep learning approaches with transfer learning mechanisms are explored and multi-class classification problems has been divided into multiple multi-class classification problems. As discussed previously IP102 dataset possess unique challenges, it consists of 102 imbalanced classes of different insects. However, it covers only 8 different plants. Because each plant is infested by some specific type of insects that may not be found on other plants.

In this research 9 models are used for the classification of the given image into the 102 different classes. This Methodology can be called divide and conquer or hierarchical classification.

Finding and designing the structure of a deep learning model can be very time consuming and tedious, since there are an infinite number of structures and requires an adequate amount of resources, as well as significant time, which goes into the training of each model. So instead of wondering and testing different models, in this research ResNet50 and VGG 16 structures have been used since these structures have a reasonable size and have shown better performance in previous studies.

The research scope was further reduced by deciding that Level 1 and Level 2 Models can be different models or can be the same models but All Level 2 Should be of same structure. I.e. level 2 models can either be a ResNet or VGG. The same goes for Level 1 It can be either ResNet or VGG.

Particularly this research investigates the different combinations of the Hybrid models by taking combinations in the level 1 and level 2 models. And also investigates the different data augmentation as well as processing methods which may help to increase the test accuracy.
Following are different approaches which are used in this research,

1. Level 1: ResNet50 and Level 2: ResNet50
2. Level 1: VGG16 and Level 2: ResNet50
3. Level 1: ResNet50 and Level 2: VGG16
4. Level 1: ResNet50 and Level 2: VGG16 Training With Augmix approach with noise in training data
5. Cascade Training Of SubClass Classifier (VGG16 as well as ResNet50)

Following is the research framework,

In subsequent sections, information about the steps and approaches used in this research will be explained.

3.1 Data Pre-processing

In this research pre-trained models and existing structures are used, and these structures require images to have a width and height of 224 pixels Since the dataset contains different sizes of images, these images should be processed such that they can be compatible with the structure.

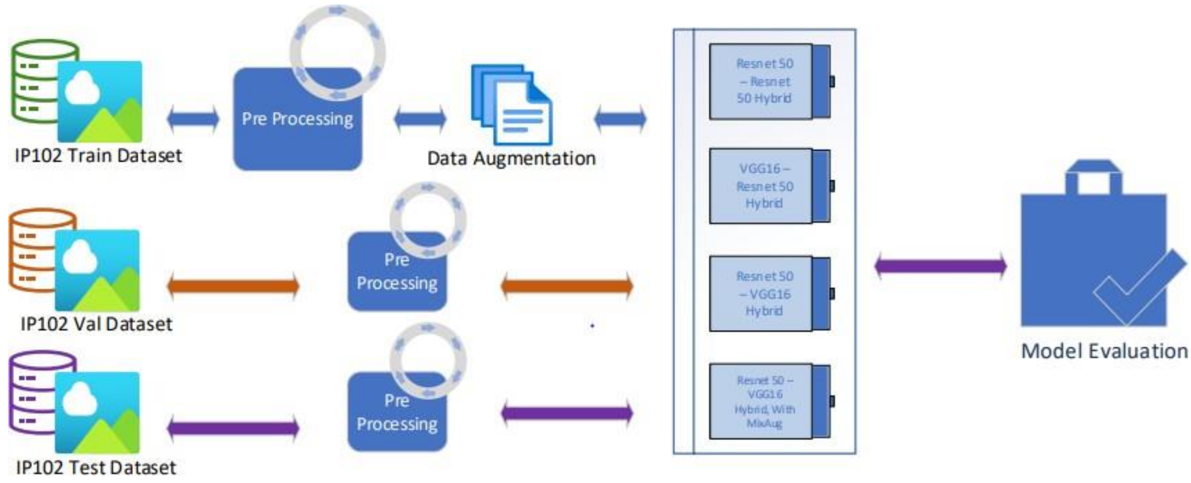

Figure 1: Research Framework

The following 3 steps were performed during the pre-processing of the images.

- Resize the image to 256x256
- A random Crop of 224x224
- Normalize image

These steps ensure that the data is compatible with the existing structures.

Considering the IP102 dataset resizing, cropping, and normalizing of the provided input image had been focused which is explained as follows:

3.1.1 Resize

This part basically consists of enlarging or diminishing the length and breadth of the provided image which generated a new image that could be utilized for the training of the model with both, the original and generated image.

3.1.2 Crop

This function is the cutting of the existing image into another piece from it. Basically, the original image is been cut into any dimension of the image which results in the generation of the new image and can be used for the training of the new image. A combination of the resize of the image as well as cropping of the image also be used which shall depend on the user as well as provided dataset.

3.1.3 Normalized

This function provided support to the model to learn faster the image. Image is normalized in order to provide the best computing image to the model from which the model can learn its maximum features which leads to the increase in the accuracy by the same image as the original image. Keen purpose of the normalization of the image while performing the training of the model is to make efficient computation by reducing the value between 0 to 1.

This step also plays a essential role in the preparation of the dataset as raw images gathered for preparing the dataset shall be in any form as well as contained the noise and improper capturing of the image can be diminished under this step. A noticeable rise in the model's accuracy can be observed by performing normalization of the image.

In our proposed solution combining the steps, resizing, cropping, and normalizing had been implemented which provided better performance and can be observed results of the proposed model.

3.2 Data Augmentation

As the name suggests the creation of a new image or data from the provided data or image which can be used for the training of the machine learning model is basically known as data augmentation. (Y. Li et al., n.d.)As the machine learning model requires a large dataset which might not be possible for every situation or domain, in order to increase the efficiency and performance of the model to be trained augmentation of the data is performed. Considering the hypothesis that the data might look like the available data by varying any parameter of it the data augmentation is been carried out. Various parameter is available to be augmented in the given data, like for an image, size, width, orientation, color, shape, etc are there which can be varied and the resultant image can be fed for the training of the model.

3.3 Transfer Learning

It can be explained as storing the gained knowledge when one problem is been solved and approaching another problem, (Huang et al., 2022)using the stored knowledge and apply over it as problem is related, is basically known as transfer learning. (Khan Ullah, n.d.)In terms of machine learning and artificial intelligence language, it can be related with the help of pre-trained model to solve a new problem by gaining knowledge and trained to increase the chances of the new predictions. Classification of the image is the domain in which transfer learning had been widely used and we can also observe its application in our day-to-day life activities.

There are 2 methods to utilize the pre-trained models to perform transfer learning which are as follows, 1. Fine-tuning 2. Using Deep Features

3.3.1 Fine Tuning

In the fine-tuning approach, weights of the all the layers of the structure are updated during the training process. Those pre-trained weights are used as the initialization weights. And from that point weights are updated. On the ImageNet dataset ResNet50 and VGG16 were pretrained. The fine-tuning approach works the best with the multiclass classification approach.

3.3.2 Using Deep Features

This is also referred to as the freezing method, where the weights of the structure fails to updated during the training on new dataset. Instead, a new adapter layer or structure is added along with the output of those models and that adapter layer is trained. It is basically taking the Pretrained model's output as input to the new models (which is the adapter layer) and training it. This may help in more simpler problems like binary classification etc.

In this research, a preliminary test of each of those showed very poor performance while utilizing those pre-trained models as the deep features for this application, so it was discarded and not considered for utilization.

3.4 Cascade Transfer Learning

It is the unique approach to training the Level 2 or Subclass classifier models. So as discussed previously, Fine tuning of the Pretrained models can improve the performance as well as bring down the training time. Here, we have assumed that features learned in one subclass may be helpful for the other subclass classification model. For example wights of a model trained on the classification of insects on the Rice could be helpful for training model for the Wheat. Here not much mix and match was performed instead a sequence was followed while training the 8 models at level 2 for the sub-class classification.

In Figure 3, First, the model for the Rice is trained where the initial weights were of the pre-trained model from Image net. Once the model is trained for the Rice, its weights are used as the initial weights while training for the Corn. And this is repeated until the last 8th model is trained.

3.5 Hyperparameters

In deep learning, and since using transfer learning, the number of hyperparameters is reduced and sometimes its values are limited by the availability of the resources. Due to transfer learning, the model structure-related hyperparameters are out of the scope, since an existing structure has been used. So remaining is the Optimizer Hyper Parameter which are learning rate, Epochs, and Batch size. Here the Batch Size is restricted by the memory size, most general is 64 but not all the systems can store such an amount of data. Whereas the Epochs really depend on how much time we have, and since the model can be retrained from where is it left and store models, it does not require to set any specific values. And learning rate does affect the learning, but is quite often kept dynamic, and updated by monitoring either validation measure or training measure like accuracy. In saying that in deep learning there is not much room for hyperparameters or performance improvement. That’s why many studies focus on either the structure of the model or, data related process to improve the accuracy of the deep learning models.

Also Unlike machine learning, in deep learning, with a complex problem, the hyper-meters tuning process becomes much lengthier. And offers little to no improvements. So most of those parameters were kept default throughout the research which is mentioned in the following table.

Table 2 Hyperparameters

Hyperparameters	Description	Default Values
Learning Rate (lr)	This parameter controls that how many weight should be updated based on the error. And is updated every epochs	0.0001
Epochs	Number of Times entire iteration of weight update process should be performed	10-30 (Also checked up to 100)
Batch size	Number of Samples should be considered before updating the weights of the model	32-64 (Depending on available Memory)

Table 1: Hyperparameters

4 Design Specification

In this research as discussed previously instead of investigating and finding a new deep learning structure, an existing structure was chosen. ResNet50 and VGG16 were considered for the classification of the IP102 dataset.

4.1 ResNet50

As the name suggests it is a variant of a ResNet model consisting of 50 layers stacked up. There are in total 48 Convolution layer with 1 Maxpool and 1 Average Pool layer(Ding et al., 2022). With the input as X, it is fed into the weight layer with relu as an activation function followed by the weight layer which is represented as F(x) and further added with the identity x which makes the output as F(x) + x in which relu activation function is used a typical representation of it is as under:

Here, the skipping approach is also known as the: highway network". This technique provides us a window through which the user can provide a condition that if the model is degraded or its performance had been diminished then the process shall skip the path and carry the further iterations. Various layers of the ResNet had been tested and on a comparative analysis, we can observe that whether increasing the layer or decreasing its direct effect can be seen.

It was also observed that with the error of only 3.57% ensemble of ResNet had won ILSRC 2015 classification competition on which the COCO dataset was used. Apart from ResNet50 various model like ResNet34, ResNet101, and ResNet152 had been also popular who could attain up to 11.3 bn FLOPS by ResNet152.

4.2 VGG16

A convolution neural network consisting of the 16 layers dedicatedly made for image processing. (Deepa and Chokkalingam; 2022)It consists of the layers as convolution + ReLU followed by max pooling, fully connected + ReLU, and SoftMax. The architecture of the VGG16 is shown as below

For convolution layer 1 it is having a fixed size of 224*224 RGB image which is passed. (Deepa and Chokkalingam; 2022)As the image is passed from the different layers the filters play the role and 3*3 to 1*1 receptive fields are used. A linear transformation is implied which shall transform the image into a single layer for 1*1. Lastly, spatial pooling and max pooling are carried out, and lastly, the softmax layer is used which makes the model a fully connected layer configuration(X. Yang et al., 2021).

4.3 Hybrid Structure

As discussed earlier here the problem is devised on smaller problems and solved using multiple models. Here models are used in a hierarchical fashion. The following diagram depicts the flow of the information,

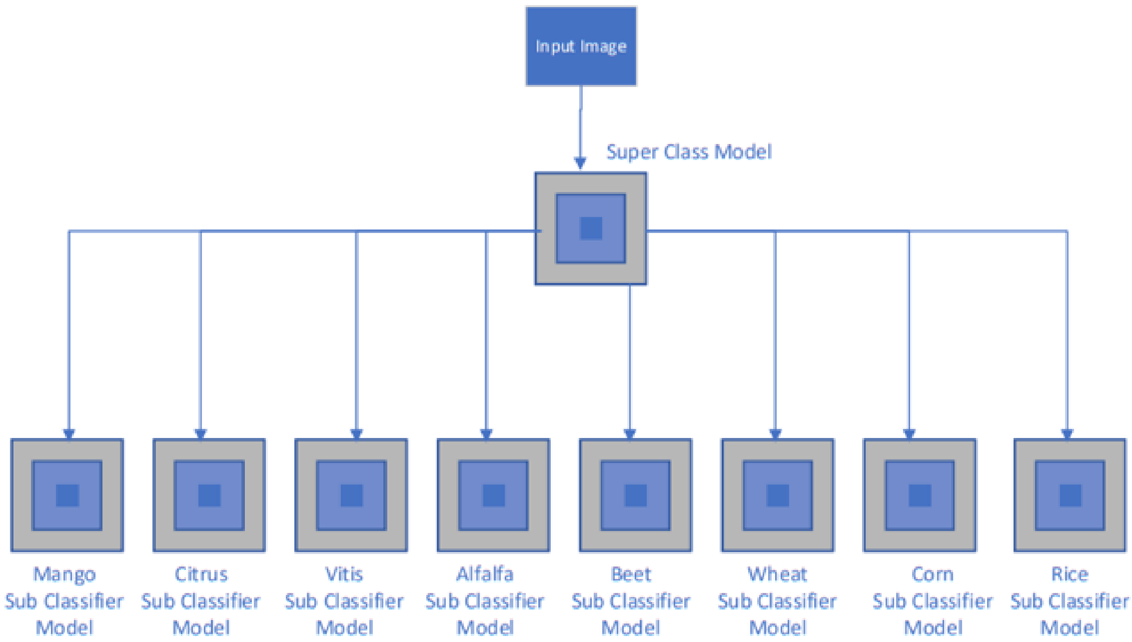

Figure 2: High Level Model Structure

To make this hybrid structure, each of the 9 models was trained individually. To speed up, the training process as well as to keep it simple. The combined all of these 9 models would require quite large resources which might not be accessible. The following table shows the size of the single ResNet50 and VGG 16 models when their weights are stored on a storage device.

5 Implementation

In this research, the IP102 dataset was used which is a benchmark dataset for insects' classification. The dataset is pre-separated into the train, validation, and test. So the train test split does not require. This deep learning model is implemented in the Python 3 Scripting language and the PyTorch library is used for the deep learning. In addition to that Sk-learn, Numpy, Matplotlib, etc. was used to facilitate other function.

Structure	Size in (MB)
ResNet 50	~ 90
VGG16	~528

Table 2: Structure Storage space requirement

Since it's a deep learning model, the training process consumes a lot of time when done on the CPU. A GPU should be used to train deep learning, for that reason a "Google Colab" a service by google is used where one can run the Python Notebooks on the servers which can have the access to GPU. First, a model for Superclass is trained, and then all models for each of the subclass are trained and their weights are stored on a storage device to evaluate later.

A Hybrid model is emulated by first passing the image to the superclass model and selecting the subclass model based on the prediction of the superclass and the final class predicted.

6 Evaluation

There are a total of 6 different models which were mentioned previously and are tested on the test data, In this section, the Classification report for the 8 superclasses is shown. Because the results vary based on the Classification accuracy of the superclass.

6.1 Both ResNet50

In this method, all 9 models were created using the ResNet50 Structure. This model was used as the initial model, where every parameter was kept default like the one shown in Table 2. Following is the classification report of the superclass.

	precision	recall	f1-score	support
Alfalfa	0.71	0.70	0.70	3123
Beet	0.72	0.60	0.65	1330
Citrus	0.77	0.76	0.76	2192
Corn	0.79	0.79	0.79	4212
Mango	0.79	0.78	0.79	2927
Rice	0.85	0.84	0.84	2531
Vitis	0.78	0.85	0.82	5274
Wheat	0.70	0.65	0.68	1030
accuracy			0.78	22619
macro avg	0.76	0.75	0.75	22619
weighted avg	0.77	0.78	0.77	22619

Table 3: Classification Report of All ResNet50 Model

The above figure shows the classification report where it shows that the F1 Score for the Beet and the Wheat class is the lowest as compared to others.All others are having F1 scores above the 70% . So this again can be because those 2 have few images as compared to the other classes. And the classification report or the confusion matrix for the final 102 classes is not shown here, in which the F1 score was ranging from 20% to 90% . And has a 57% micro accuracy. Half of the problem can be solved by improving the performance of this superclass only. Since the sub-class will have much fewer samples. And overall accuracy is about 65% .

6.2 Level 1 VGG16 – Level 2 ResNet50

Since the performance of the superclass should be increased here in this approach a different structure was utilized to do so, however, it fails to improve it which can be seen from the following classification report of the superclass.

	precision	recall	f1-score	support
Alfalfa	0.69	0.65	0.67	3123
Beet	0.68	0.57	0.62	1330
Citrus	0.74	0.71	0.73	2192
Corn	0.70	0.80	0.75	4212
Mango	0.78	0.72	0.75	2927
Rice	0.80	0.83	0.81	2531
Vitis	0.76	0.80	0.78	5274
Wheat	0.64	0.57	0.60	1030
accuracy			0.74	22619
macro avg	0.72	0.70	0.71	22619
weighted avg	0.74	0.74	0.73	22619

Table 4: Classification Report of VGG16-ResNet50 model

There is a difference of 4% in the result which will not improve the result. Since not all structure behaves the same for all scenarios VGG is then used for sub-class.

6.3 Level 1 ResNet50 – Level 2 VGG16

In this approach, the superclass classification model was kept the ResNet50 as the first one but the sub-class classification models are considered with the VGG16. And it had the same Classification report for the 8 classes but had slightly different outcomes when it came to the accuracy and the F1 measure of the 102 classes. The overall accuracy is about 64% .

However, each of the ResNet50 and ResNet50+VGG16 models is then tested by adding noise and gaussian noise respectively.

6.4 Both ResNet50 Adding noise

Here the same model was used as was in the first, just during the testing a small amount of noise had been added to the images and the accuracy of it dropped to 53% from 65% . Which certainly was a reduction in performance.

6.5 Level 1 ResNet50 – Level 2 VGG16 Adding Gaussian Noise

In the other model, Instead of the noise, images were blurred by a very small amount, there was little to no impact on the performance of the model.

6.6 Both ResNet50 With Cascade Training

Here models are trained in the cascaded fashion as described in the previous section. However, it did not improve the performance. This can be seen from the classification report of the 102 classes.

6.7 Level 1 ResNet50 – Level 2 VGG16 with Cascade Training

Similar to the previous one, here all the VGG16 are used in Level 2 and also had the same result as the previous one.

accuracy			0.64	22619
macro avg	0.59	0.55	0.56	22619
weighted avg	0.64	0.64	0.64	22619

Table 5: Classification Accuracy of model with Cascade Training of ResNet+VGG16

6.8 Level 1 Resnet50 – Level 2 VGG16 Training with AugMix and Random Noise

To arrive at this proposed methodology, classification results were visually investigated as well, following is the ReNet50 + ResNet50 model (i.e. All ResNet models)

Figure 3: Result of ResNet50 + ResNet50

From the above First, the column is of the correctly predicted class whereas the second one is incorrectly predicted images. This shows that images , where the insects are visible very clearly, can be easily detected however the ones which is not clearly visible are not classified correctly. And this result seems to be consistent with other models as well.

Now the following image shows the same model but when tested with the random noise,

In the above figure, the results show quite different compared to the previous one. Where the first image of “Locustoidae” from the Alfalfa superclass was misclassified to

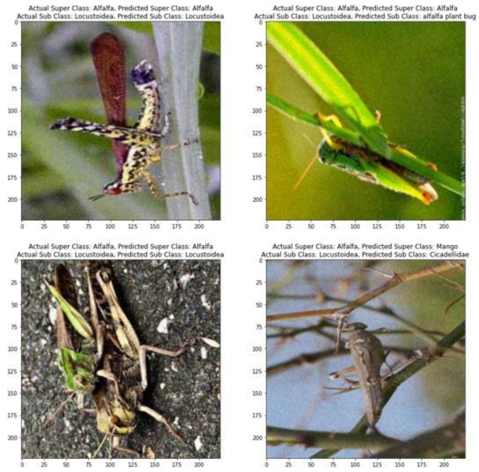

Figure 4: Classification result if ResNet50+ ResNet50 on noisy images

other classes when tested without the noise. And also, in the above image first column is of correct classification and the second for the misclassified images. But when images from the first rows are observed the other one had superclass classified correctly but miss classified into something else.

And both of are having same actual class names. When observed closely they both have different visual and physical features. Locustoidae. And here all 4 images are from the same sub-category “Locustoidae” Despite that all have different features. Because Locustoidae is a kind of family of the different insects, Grass hoppers. This is the very reason the classification from the images for the insects is very difficult. Because despite having the same family they all have different features and these features may also intersect with other categories. For example, the second image in the first row. Visually it does share the features of the “alfalfa pant bug” it is green and found on green colored something (can be plant or grass).

This behavior showed hope that training images with added noise and with augmentation can help to generate images that may have enhanced features. So here models are trained with the AugMix and the noise in the training image, and during the testing, images were not added the noise. And following classification report show the improvement in the classification accuracy of the superclass.

	precision	recall	f1-score	support
Alfalfa	0.96	0.92	0.94	3123
Beet	0.93	0.93	0.93	1330
Citrus	0.95	0.96	0.95	2192
Corn	0.97	0.95	0.96	4212
Mango	0.96	0.95	0.96	2927
Rice	0.96	0.97	0.96	2531
Vitis	0.94	0.97	0.96	5274
Wheat	0.93	0.93	0.93	1030
accuracy			0.95	22619
macro avg	0.95	0.95	0.95	22619
weighted avg	0.95	0.95	0.95	22619

Table 6: Classification Report of ResNet50+VGG16 Hybrid model trained with AugMix and Noise

Which certainly helped to improve the classification of the 102 Classes which can be seen from the following classification report.

accuracy			0.75	22619
macro avg	0.70	0.65	0.67	22619
weighted avg	0.75	0.75	0.75	22619

Table 7: Classification Accuracy of ResNet50+VGG16 Hybrid model trained with Aug-Mix and Noise on 102 classes

Figure 5: Classification Result

6.9 Discussion

Looking at Figure 3, Figure 4, and Figure 5 show that there are still cases where the model gets confused. For example, a single insect remained common across almost all the methods. Which is the pest from the Alfalfa category. In Figure 5 it's the second column's first image. There is still scope for improvement as this approach was able to achieve an accuracy of 75%

Following is the Accuracy of each individual model prepared for the sub-classification. This shows that, if the superclass is having about 100% test accuracies, then the proposed methodology can max provide accuracy of 79% .However, the ResNet model at best was able to provide 78% accuracy on the superclass. This means roughly about 20% of data is misclassified already after passing through the super classifier. So basically, models for sub-classification are only working on the 80% data.

The following table summarizes the performance of each approach and shows that the Method With AugMix is the clear winner. And indicates how important is data Augmentation when it comes to multi-class classification.

The dataset has 102 classes, and it has a total of 75k images out of which approximately 45K are in the training dataset. Which makes roughly about 442 images per class, which seems to be quite low. But dataset is not balanced so some classes are having more images as compared to others. Data augmentation using the AugMix creates additional images by applying transformations like rotation, flip, cropping, stretching, etc, during the training. This solves the problem of fewer data and improves the class imbalance.

Noise is added to see the robustness of the classification. Noise is used to simulate the issues when collecting or taking pictures in low light or out of focus. The addition of the gaussian noise with random between 0.1 and 0.2, showed the very little impact. And the second method of adding noise is where 20% values of a random array of the same size as the image are added to the image which resembles noise that may be introduced in low-light photos. In this second type of noise, there was more impact. So in final models along with AugMix this type of noise is also added as the added feature extraction method.

Table 8: Performance Summary

METHODOLOGY	DESCRIPTION	ACCURACY
HIERARCHICAL CLASSIFICATION USING ALL RESNET50 MODEL	This Methodology uses 9 Resnet 50 models. Dataset consists of hierarchical taxonomy, where there are 8 super classes and each of one has sub class. So out of 9 model 1 model is for detecting the super class of the given image, then secondly using other 8 models which are for each of the 8 super class is used to determine the final class.	65%
HIERARCHICAL CLASSIFICATION USING ALL RESNET50 AND VGG16 MODEL	Same as the previous one but, for the super user model Resnet 50 was used and for the sub class the VGG16 was used	62%
VGG16 RESNET 50 HYBRID	Similar to first one but Out of 9 model for super class classification VGG16 was used and for rest of other Resnet50 was used	64%
GAUSSIAN	Here the Resnet50-VGG16 Hybrid model is tested on test data by applying the Gaussian Noise And checked test accuracy	64%
NOISE	Here Resnet50-Resnet50 Hybrid model is used and test accuracy is measured by applying normal distributed noise.	53%
CASCADE TRAINING FOR RESNET50-RESNET50 HYBRID	Here the 8 models used for the sub class classification are trained in cascading fashion, which means once the first model is trained, that same models weights are used while starting the training of the second model and so on. Transferring the knowledge learned during training of each sub class.	64%
CASCADE TRAINING FOR RESNET50-VGG16 HYBRID	Process is same as previous one but instead of the Resnet50 model for 8 sub class, VGG 16 is used	64%
RESNET50-VGG16 HYBRID WITH NOISE AND AUGMIX	Here the Resnet50-VGG16 Hybrid model is trained by utilizing the AugMix for data augmentation and adding the Random Noise in the training image to increase the features from the image as well as increase the robustness of the model	75%

7 Conclusion and Future Work

In this research, with the usage of the IP102 dataset which is 102 classes of pest dataset consisting of its images are used to identify the pest in agriculture. Two approaches were studied in this research which shall highlight the difference in the model and its usage as well as its performance when the combination of the models is used. Dividing the identification method as firstly identifying the class of the pest each of the unknown image or pest belongs shall provide us the usability of the two models with which we shall be able to classify the image / Pest among 102 classes which are performed by ResNet50. For classification into the main classes, in the first approach, VGG16 followed by ResNet50 is used. In the second approach ResNet50 followed by ResNet50 is used which shall be showing better accuracy in classification than the first approach. With the VGG16 and ResNet50, an average accuracy of 62.20% whereas the macro average of 56.80% and weighted average of 62.20% is achieved. With the ResNet50 and ResNet50 approaches, 65.39% of average accuracy and the macro average of 59.76% , and a weighted average of 65.44% are observed. Observing the results and analyzing them, a noticeable difference in both approaches indicate the better approach shall be a combination of ResNet50 and ResNet50 which had showcased better performance in the classification of the pest among 102 classes. However further investigation into the study revealed that the accuracy can be improved further just by improving the accuracy of the superclass, which was achieved by employing a more aggressive data augmentation process using the AugMix technique. Which drastically improved the performance of the superclass classifier which was the ResNet50 Model. Also, this study investigated other novel approaches which were identified during the course of the research work. However, those did not perform well for this particular problem but could open doors for other research. And the final best model is where the ResNet50 and VGG16 Hybrid model is trained with Noise in the image and the AugMix method is used.

As far as the research question goes, Hierarchical classification does help in classifying the 102 classes but loses the portability, since those models themselves require somewhere around 1GB to 5GB of disk space depending on the combination used. And best method is the AugMix which is quick and easy to integrate. There are more complex approaches but do not work as expected.

The hybrid model creates multiple models for each subcategory of the class. So, each model has to classify between classes fewer than 102 classes. Training of individual models had been carried out and since each model also has only the data which belongs to the particular category, those models learn the features of that class only. Results showed improvement when paired with proper data augmentation techniques.

So final Recommendation is to add noise that resembles the noise in the image when captured in low light. This is like the salt and pepper noise as well as speckle noise. This noise impacts the fine details of the images. Thus, augmenting images with these types of noise can create additional images with variations in fine details.To enhance the image features, a uniformly distributed Noise with 20% of its magnitude is applied to the training image. Since it was discovered throughout the research that introducing noise to the images lowers the model's overall performance. However, several photos that were incorrectly classified are classified correctly. Noise is therefore employed as a feature enhancer.

As the limited number of datasets available for the pest as well as with the small categories of 102 classes of pests had limited our research which can be furthermore explored in the future. As far as its application is concerned, minimizing the processing part of the model as well as its integration into the cloud-based system which can be accessed from mobile applications shall be a future scope of this research.

References

Aly, M. (2005). Survey on multiclass classification methods, *Neural Netw* **19**(1): 9.

Ayan, E., Erbay, H. and Varçın, F. (2020). Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks, *Computers and Electronics in Agriculture* **179**: 105809.

Bhadane, G., Sharma, S. and Nerkar, V. B. (2013). Early pest identification in agricultural crops using image processing techniques, *International Journal of Electrical, Electronics and Computer Engineering* **2**(2): 77–82.

Deepa, N. and Chokkalingam, S. (2022). Optimization of vgg16 utilizing the arithmetic optimization algorithm for early detection of alzheimer's disease, *Biomedical Signal Processing and Control* **74**: 103455.

Gomes, J. C. and Borges, D. L. (2022). Insect pest image recognition: A few-shot machine learning approach including maturity stages classification, *Agronomy* **12**(8): 1733.

Han, C., Rundo, L., Araki, R., Nagano, Y., Furukawa, Y., Mauri, G., Nakayama, H. and Hayashi, H. (2019). Combining noise-to-image and image-to-image gans: Brain mr image augmentation for tumor detection, *Ieee Access* **7**: 156966–156977.

Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J. and Lakshminarayanan, B. (2019). Augmix: A simple data processing method to improve robustness and uncertainty, *arXiv preprint arXiv:1912.02781* .

Karar, M. E., Alsunaydi, F., Albusaymi, S. and Alotaibi, S. (2021). A new mobile application of agricultural pests recognition using deep learning in cloud computing system, *Alexandria Engineering Journal* **60**(5): 4423–4432.

Kasinathan, T., Singaraju, D. and Uyyala, S. R. (2021). Insect classification and detection in field crops using modern machine learning techniques, *Information Processing in Agriculture* **8**(3): 446–457.

Li, W., Zhu, T., Li, X., Dong, J. and Liu, J. (2022). Recommending advanced deep learning models for efficient insect pest detection, *Agriculture* **12**(7): 1065.

Mohsin, M., Bin, R., Ramisa, S. A., Saad, M., Rabbani, S. H., Tamkin, S. et al. (2022). *Classifying insect pests from image data using deep learning*, PhD thesis, Brac University.

Nanni, L., Maguolo, G. and Pancino, F. (2020). Insect pest image detection and recognition based on bio-inspired methods, *Ecological Informatics* **57**: 101089.

Nanni, L., Manfè, A., Maguolo, G., Lumini, A. and Brahnam, S. (2022). High performing ensemble of convolutional neural networks for insect pest image detection, *Ecological Informatics* **67**: 101515.

Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning, *arXiv preprint arXiv:1712.04621* .

Ren, F., Liu, W. and Wu, G. (2019). Feature reuse residual networks for insect pest recognition, *IEEE access* **7**: 122758–122768.

Silva-Palacios, D., Ferri, C. and Ram´ırez-Quintana, M. J. (2017). Improving performance of multiclass classification by inducing class hierarchies, *Procedia Computer Science* **108**: 1692–1701.

Wu, X., Zhan, C., Lai, Y.-K., Cheng, M.-M. and Yang, J. (2019). Ip102: A large-scale benchmark dataset for insect pest recognition, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8787–8796.

Yang, G., Chen, G., Li, C., Fu, J., Guo, Y. and Liang, H. (2021). Convolutional rebalancing network for the classification of large imbalanced rice pest and disease datasets in the field, *Frontiers in Plant Science* p. 1150.