National College of Ireland

Detecting Pests on Tomato Plants using Convolutional Neural Networks.

MSc Research Project
Data Analytics

Srivenkateswara Rao Vatti
Student ID: x18181104

School of Computing
National College of Ireland

Supervisor: Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Srivenkateswara Rao Vatti
Student ID:	x18181104
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Rashmi Gupta
Submission Due Date:	17/08/2020
Project Title:	Detecting Pests on Tomato Plants using Convolutional Neural Networks.
Word Count:	XXX
Page Count:	20

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:	Srivenkateswara Rao Vatti
Date:	27th September 2020

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):	

Detecting Pests on Tomato Plants using Convolutional Neural Networks.

Srivenkateswara Rao Vatti
x18181104

Abstract

Image processing is widely used in various industries around the globe such as bioinformatics, space, weather forecasting, disease diagnosis and so on. With the recent advancements in the field of deep learning and artificial intelligence (AI), GPU-powered deep learning frameworks such as TensorFlow, Pytorch, Microsoft Cognitive Toolkit and others, many challenging problems in computer vision such as image classification, object detection and many more can be solved. The current research emphasises on detecting and classifying the pest that is formed on tomato plants. The analysis carried out in this paper is based on convolutional neural networks(CNN). The usage of scouting robots in the agriculture industry is increasing than ever before and many big organizations are investing large amounts of money on the same. The primary goal of developing robotic solutions through AI is to reduce the manpower that is utilized for harvesting purposes. As part of the current research, initial blocks from a set of CNN models such as VGG16, VGG19, Xception, ResNet50 and Inception V3 along with additional convolutional layers are applied on the dataset chosen and results are evaluated with the help of various standard metrics. Maximum classification accuracy of 0.95 is achieved through a CNN model, which is implemented with a set of convolutional layers from the VGG16 and additional layers which are added explicitly. Also, a comparative analysis is carried out with other models developed with transfer learning. Detection and classification of pests or insects for the selected dataset is a challenging task since the size of them is very minute. The results that are obtained in terms of accuracy for detecting and classifying the type of pest/insect can be crucial in integrating the developed models to a scouting robot which can accept an image as an input and identifies the type of pest in that image. Early detection of pest/insects can minimise the usage of pesticides and increase the overall productivity of the crop. The integration of the models implemented with scouting robots is not covered as part of this research.

1 Introduction

Agricultural sector plays a vital role in terms of the economy of any country. Nowadays,this sector is facing a couple of issues such as shortage of labour, the diseases that attack crops which in turn reduces the overall productivity. Hence the detection of pest on crops at

the early stage is essential in terms of avoiding large amounts of pesticides and achieving better productivities. Observation methods through the naked eye are not appropriate and sufficient for crops on a large scale. The recent advancements in computer vision and deep learning methods can play a major role in improving productivity with minimum usage of pesticides. Image processing through convolutional neural networks (CNN) and deep learning are playing a key role in image classification tasks in the modern Artificial Intelligence(AI) world. Deep neural network architectures can be leveraged in real-world applications such as disease diagnostics, satellite image processing, robotic harvesting, time series analysis and weather forecasting. Especially with the recent emerging frameworks in deep learning such as TensorFlow, Keras and others, there is a huge scope for developing better deep learning models for tasks like object detection, image processing and image classification.

The current research is motivated by Gutierrez et al. (2019) in which a comparative analysis is carried out with the help of a couple of pre-trained deep learning models as well as a combination of models implemented with computer vision and machine learning. The research mentioned in ibid focuses mainly on choosing the best method in terms of accuracy for the detection and identification of pests. A dataset is generated with a huge set of images of infected tomato plants with harmful pests to evaluate the machine learning and computer vision models implemented. In the research mentioned in ibid, two types of colour cameras namely AP-3200t-PGE and DataCam 2016R are used to capture the images of pests on tomato plants which are cultivated inside a set of completely enclosed boxes. Once the dataset is created in this way, 4330 images are considered and labelled with a labelling tool called LabelImg [1], an open-source project which is released under t MIT license. The primary objective of this analysis is to improvise the accuracy of pest detection by implementing a deep convolutional neural network(CNN) with recent frameworks such as TensorFlow and Keras. Also, different latest pre-trained models can be applied to the dataset selected to observe the accuracy and other metrics.

This research is based on detection and classification of pests and insects that are formed on tomato plants by leveraging various CNN architectures. A set of pre-trained models such as VGG16, VGG19, ResNet50, InceptionV3 and Xception will be applied on the selected data set and the accuracies achieved by these models will be compared with the one trained with custom convolutional layers.The major limitations in the current research are that the implemented models are not integrated with scouting robots and hence the real-time efficiency of the models can not be evaluated. Also, the current research does not cover implementing an integrated system which can accept an image and outputs the percentage of different type pests that can be observed in that image.

The diagram shown in Figure 1 shows a set of 9 images which are cropped from the actual images as per the coordinates in the annotations. The below set of images are cropped from the raw images of pests and insects and resized into 32 X 32 X 3 to pass to the proposed convolutional neural network as inputs. There is a set of 4 'bt' and 5 'wf' images in the below set of images. The terms "bt" and "wf" indicates Bemisia tabaci and Trialeurodes vaporariorum insects, respectively. The major goal of this analysis is to detect them and classify the images based on their category.

[1]Tzutalin: https://github.com/tzutalin/labelImg

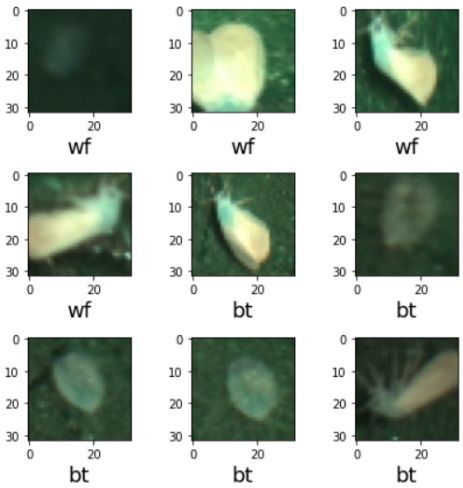

Figure 1: Cropped images from a raw image

Challenges: The major challenges that may be encountered as part of this analysis are identification and classification of pests that are formed on tomato leaves since the size of insects/eggs of various pests is extremely low and the models that are going to be developed as part of this research are trained in such a way that the pixel values must be extracted as per the annotation files and the images should be cropped accordingly.

2 Related Work

The recent literature in image processing through deep learning and convolutional neural networks is mentioned below. Various aspects of image processing through CNN architectures such as data collection, pre-processing, modelling and evaluation are discussed briefly from the respective research papers.

Segmentation of the cell nucleus in histopathological images is implemented through CNN Pang et al. (2010) based architecture in which a data set of HE breast cancer biopsy images are used for the analysis. A CNN with 8 feature maps and 3 hidden layers implemented as part of this research and the results are impressive when compared to other pixel classification models such as SVM. A machine learning approach is used in this research to get the domain-specific knowledge which is key for implementing a segmentation strategy. Secondly, CNN is utilized to segment the images-based weight optimization through various filters in the learning process.

Restring an image from another image which is taken from a window covered by rain or dust is explained in Eigen et al. (2013). A data set of clean and corrupted pairs of images are collected to train a customized form CNN which learns to map corrupted image patches to clean ones by using the characteristics of water and dirt appears in the natural images. Different convolutional networks are trained for dirt and rain which in turn allows the overall model to predict the corrupted patches effectively. The network is

trained with 5.8 million samples of synthetic dirt paired with the realistic clean patches. The models implemented as part of this research are effective in terms of dirt removal when tested in outdoor conditions.

A real-time facial expression recognition function which can be used on a smartphone is implemented in Song et al. (2014) through a deep CNN architecture on a GPU. There are 5 layers in the network that is built for facial expression recognition. To reduce the problem of overfitting, techniques such as dropout and data augmentation are used. A smartphone app is developed which can classify a facial expression and is based on a deep neural network. The network structure which is used as a base for the implementation in this research is trained with benchmark CIFAR-10 dataset and showed an impressive performance.

For the operation of computer-aided diagnostic systems, automatic segmentation of human anatomy is a major function. The classification of human anatomy is difficult because of the variability and complexity of it. A medical image classification method is implemented in Roth et al. (2015) for classification of medical images of anatomy acquired using computed tomography (CT) using ConvNets. There are 4298 2D key-images with 5 anatomical classes are utilized in the research and these images are collected from 1675 patients. Data augmentation technique is used in this data set to enhance the data variability and the method obtained impressive results with an AUC (area under the curve) of 0.998.

A new framework is implemented in Peng et al. (2018) in which histopathology image analysis is carried out to classify tissues in histopathologic images. This research is based on Haematoxylin and Eosin (HE) images of stomach tissues which are collected at real-time. The results obtained in this analysis showed comparable accuracies in terms of both classification and segmentation. Two architectures of CNN are developed in this research namely, a path-based CNN and a fully convolutional network (FCN) in two stages. The methods implemented in this analysis are based on pre-trained models such as AlexNet and GoogleNet and most of the parameters from those networks are utilized to train the selected data set.

A new hyperspectral image classification (HIS) is implemented in Feng et al. (2019) with a specific type of CNN namely convolutional long short-term (MCNN-ConvLSTM). Initially, through CNN, features are extracted. An end-to-end classification with the help of multi-layer spatial-spectral features which shallow layers can be used to get complimentary information and deep layers to get abstract information

A new model, evolving deep convolutional neural networks(EvoCNNs) is proposed in Sun et al. (2019) by leveraging genetic algorithms in which weight initialization values and architectures of convolutional neural networks(CNNs) are evolved to address image classification problems. A strategy namely variable-length gene encoding is designed in this research paper to describe the potential depth in the convolutional layers and to explain different building blocks in the model. In the proposed algorithm, a new mechanism is discussed in which connection weights of deep CNNs are effectively initialized and this can prevent the networks getting stuck into a local minimum which is a major issue in backward gradient-based optimization.

Agricultural productivity plays a vital role in the economy of a country. The major problem faced by cultivation is pests that are formed on the leaves of plants and various types of diseases. Hence detection of pests or diseases in the early stages of crops is essential to get maximum productivity. The classification of disease on tomato plants is implemented in Batool et al. (2020) with a data set which contains 450 images. The features of these images are extracted using several models and the classification is achieved through K-nearest neighbours' algorithm.

In the area of disease diagnosis, the classification of medical images plays a vital role as it can guide doctors to make accurate decisions as part of the overall diagnosis. The CNNs such as DenseNet can effectively classify medical images Huang et al. (2020) but the only caveat is that it needs large amounts of properly labelled training data. A light-weighted hybrid neural network is proposed in this research In combination with a Principle Component Analysis Network (PCANet) and a less complex DenseNet.The architecture of the proposed network can be explained in two stages and each stage, multiple kernels will be used for learning. The standard metrics that are derived from a ROC such as Sensitivity, Specificity and Area Under Curve(AUC) are compared among other pre-trained models such as AlexNet, VGG-13, ResNet-50 and so on. The hybrid neural network architecture implemented as part of this analysis outperformed when compared with other models with an accuracy of 83%.

A DenseFood model is proposed in Metwalli et al. (2020) in which a densely connected CNN model with multiple convolutional layers is discussed. The primary goal of this research is to implement a CNN model that can classify the food images according to their category. A combination of centre loss and softmax loss is utilized in the training phase to reduce the variation among the same category of food images and maximise the variation among the different category of the same. The DenseFood model implemented in ibid is based on DenseNet architecture and the major component is the set of dense convolutional layers connected which can reuse the features and lower the required number of parameters. The size of the data set used to train the model implemented contains 110,241 images which are relatively low in size. The total number classes are 172 i.e. on an average basis 641 per class. The DenseFood model achieved an accuracy of 83.92%.

A Faster region convolutional neural network (F-RCNN) is designed in Laishram and Thongam (2020) in the primary objective of the research is to classify dental and oral pathologies. This research provides a method that can be used to identify and classify types of teeth and other oral anomalies which lies internally such as impacted teeth and partial denture which is fixed. As part of this research, image processing is performed by using various techniques. This research is based on the concept of Anchors in which each input signal that contains the teeth of a person will be passed to the model implemented. The algorithm works based on bounding boxes than the manual separation of specific teeth among the set of teeth of a signal. A data set of Orthopantomogram(OPD) images is used to train the model implemented and this data set is acquired using an OPD instrument called Care stream (KODAK) Dental's CS8100. The proposed model achieved an accuracy of above 90% for detection and classification, it achieved more than 99%.

A classification of drones is implemented in Rahman et al. (2020) with a CNN model developed. The primary aim of this research is to create a large data set of micro-doppler spectrogram images of flights and drones which are in motion and to develop

a CNN model to classify these images. Two datasets are used as part of this research where the first one is with Red, Green and Blue (RGB) images and this data set is mainly utilized for training of GoogleNet based architectures whereas greyscale data set was used for the model that is developed as part of this research. Micro-doppler spectrograms are obtained from real-time experimental trails to ensure the better performance of the model implemented. Each data set is divided into two categories such as 2-class (drone and non-drone) and 4-class (noise, clutter, drone and bird). The trained models are tested with a set entirely unseen data which is unlabelled as well. The accuracies for validation and testing are for two-class data set are 99.6% and 94.4%. On the other hand, the accuracies for a four-class data set are 99.3% and 98.3%, respectively.

In general, most of the CNNs accepts images of similar size for the classification purposes though there few models which can accept images of different sizes. The disadvantage is that the performance of these models will be lowered when the dissimilar images are fed as inputs. An image classification method is discussed Park et al. (2020) in which images of different sizes are passed as input for the CNN model called MarsNet which is trained as part of this analysis. In this method, a dilated residual Network (DRN) is upgraded such that features maps with higher resolution will be obtained. Further, to facilitate multi-label classification, two modules are designed namely multi-label scoring and threshold estimation. The implemented CNN model is also trained on different datasets and standards metrics such as classification accuracy are compared. The results obtained by training the models on two data sets namely SPI data set and VOC 2007 data set showed excellent performance of the model.

Classification of eye states is crucial is many real-world eye-related applications such as fatigue detection, a device controlling in smart devices which are generally used at home and the analysis of psychological states. There are several existing neural networks in place for classification tasks of eye states and all these methods achieved accuracy over 96% as far as the eye state prediction is concerned. An improvised eye states classification system namely EyeNet is explained by Rahman et al. (2020) in which the proposed model is tested on three data sets such as MRL Eye, CEW and ZJU. As deep neural networks required large amounts of data for training the model, in this research a couple of data augmentation techniques are implemented to enlarge the size of the data set. ReLU is used as the activation function in the proposed CNN model. The model implemented obtained an accuracy of 99% in terms of eye state classification.

In the modern world, Astronomy is an area in which huge amounts of image data need to be analysed and this analysis may not be done by experts alone. Hence, astronomers generally depend on amateur people. On the other hand, lots of images are captured by new generation telescopes it is quintessential to leverage the capabilities of machine learning to build models that can classify these images. An algorithm is proposed in Jiménez et al. (2020) which is based on the gene encoding strategy of variable-length. The main objective of this research is to design two strategies for classification of galaxy images. Firstly, a combination of feature extraction along with a classifier is leveraged followed by a state-of-the-art feature extractor, WND-CHARM is compared with the model proposed in this research. This paper is based on the concept of Auto Encoder(AEs) for feature extraction. A comparative analysis was carried out in this model by two different models based on AEs, the first one is deep AE model and the second one is a convolutional AE model.

In recent times, there are a couple of methods such as deep networks and convolutional networks by which the classification models of hyperspectral images are implemented. On -e issue with these models is that there is a necessity of large amounts of data to train these models and this data should be labelled for training purposes. Gathering adequate amount of training data is a costly matter and time consuming as well. A semi-supervised model which is graph-based is discussed in Mou et al. (2020). The proposed nonlocal graph convolutional network (nonlocal GCN) in this research takes the whole image as input, unlike CNN or RCNN model that receives patches or pixels of hyperspectral images as inputs. To extract features of images, a couple of graph convolutional layers are connected in the proposed model. The nonlocal GCN is achieved competitive results when compared with spectral-spatial classification networks and state-of-the-art spectral classification models.

In the clinical practice, early identification and detection of breast cancer are crucial as far as the treatment is concerned. However, the detection of breast cancers in the early stages is still a challenging task and the deep learning models that are built for this purpose have large scope for future improvement in terms of accuracy is concerned. A classification framework is implemented in Wang et al. (2020) which is based on histology images along with a combination of deep learning and machine learning techniques. A multi-network feature extraction model by leveraging a pre-trained deep convolutional neural network model (DCNN) along with an effective dimensionality reduction technique. The proposed model will be trained with an ensemble support vector machine (E-SVM). The proposed model is evaluated with a breast cancer image data set called public ICIAR 2018 Challenge dataset and an accuracy of 97.70% is achieved.

Recent convolutional neural network models(CNNs) have achieved significant results in terms of classifying skin lesions automatically. Still, there are some issues in terms of developing models for this purpose. The primary concern is that the availability of training samples is getting difficult and for CNN based classification model, large amounts of training data are required for better performance. In Zhou et al. (2020), a skin lesions classification using convolutional spiking neural networks (SNNs) is discussed in which an average accuracy of 83.8 is achieved. In this research, feature selection is proposed to select more diagnostic features to improve the performance of the model in terms of classification. Experimental results as part of this analysis proved that SNNs performed better than CNN's with better accuracies and runtime efficiency.

Early detection of lung cancers is crucial as far as the treatment is concerned and the survival rates of the patients depend on the same. Hence it is the most critical step is to identify lung nodules in computed tomography (CT) images accurately. A two-stage Convolutions Neural Network (TSCNN) is proposed in Cao et al. (2020). The first stage is based on a U-Net segmentation network to identify lung nodules. A new sampling strategy is used to get better accuracy in this stage. A dual pulling structure is built in the second stage to reduce the false-positive rate. A random mask is designed as part of data augmentation since a significant amount of data is required to train the model. The proposed model is verified on LUNA data set and the results obtained by the model are competitive.

Cancer is one of the primary causes of death across the entire world. Hence early detection and treatment is the key in terms of increasing the survival rates of

human beings. The diagnosis of cancer through histopathology is essential since the diagnosis of malignancy is mostly done with a pathological confirmation. The scarcity of pathologists to diagnose large amounts of data in the form of pathological images lead to delay in the treatment of cancer. In Sun et al. (2020) a CNN model is implemented to detect histopathologic cancer in which different type activation functions and a gradient descent optimization algorithm are used to analyse the accuracy of detections. The model is tested on a benchmark data set namely Patch Camelyon which consists of both cancerous and non-cancerous images. It was identified that the maximum accuracy of the model implemented is around 97.94%.

The state recognition of food images is gaining huge interest in the computer vision domain in recent times. Recently a dataset is published by researchers but unfortunately there is no information about categories. In the research paper Ciocca et al. (2020), a CNN is proposed with a new data set where there are 20 categories of food are taken from fruits vegetables. There are also different states like solid, creamy and paste. In this analysis, as there is unavailability of sufficient labelled data to train the model, deep features are extracted from CNNs combined with Support vector machines. The model implemented in this research is used for different recognition tasks: food states , categories and both food states and categories. The results obtained by the model are impressive and the performance can be matched with that of a state-of-the-art method.

The Table 1 summarizes the previous literature in the field of image processing through convolutional neural networks.

Table 1: Literature Review Summary Table

Author(s) and Title	Study Aims and Objectives	Research Design	Sample	Findings
Pang et al. (2010)	To implement Segmentation of Cell Nucleus in Color Histopathological Imagery.	CNN with 3 hidden layers	58	A comparative analysis of a couple of segmentation algorithms such as SVM is examined.
Eigen et al. (2013)	To restore an Image taken through a window covered with dirt or rain	CNNs	5.8 million	Model showed effectiveness in terms of removing dirt and rain from images which are taken in outdoor conditions
Song et al. (2014)	To recognise facial expression using deep neural networks.	CNN deployed on a smart phone	1400, 7970, 8510, 1860	accuracies of 99.2%, 97.1%, 95.5% and 84.5% achieved
Roth et al. (2015)	To classifiy anatomy-specific medical images.	A deep CNN model	4298	The average AUC of 0.998 is achieved
Peng et al. (2018)	To classify and segment tissue histopathology images using deep CNN model.	A fully connected CNN	300	A model to classify histopathology images is implemented which is faster than the tested state-of-the-art methods.
Feng et al. (2019)	To Classify hyperspectral images	A convolutional long short-term and a CNN based architecture	200	The results showed that the model implemented in this analysis outperforms existing methods which are based on CNN and CRNN.
Sun et al. (2019)	To address general image classification problems	A CNN model namely evolving deep convolutional neural networks (EvoCNNs)	60000 and more	A new algorithm is proposed to address local minimum which is a major issue in backward gradient-based optimization.
Huang et al. (2020)	To classify medical images as part of diognosis	Principle Component Analysis Network (PCANet) and a less complex DenseNet	4641	Classification accuracy of 83% achieved
Metwalli et al. (2020)	To classify the food images	A CNN with DenseNet architecture	110,241	Classification accuracy of 83.92% achieved
Laishram and Thongam (2020)	To classify dental and oral pathologies	A Faster region convolutional neural network (F-RCNN)	145	Accuracy of above 90% for detection and 99% for classification.
Rahman et al. (2020)	To implement a CNN model for Eye-State Classification.	A CNN based architecture namely EyeNet.	2423	Classification accuracy of 99% achieved
Park et al. (2020)	To classify multi-label images of different sizes.	A CNN based architecture namely MarsNet	75600	Competitive Results obtained
Jiménez et al. (2020)	Galaxy image classification	A CNN model in combination with a state-of-the-art feature extractor , WND-CHARM and a set Auto-Encoders	667944	Competitive Results obtained
Mou et al. (2020)	Hyperspectral image classification	Nonlocal graph convolutional network (nonlocal GCN)	42766, 9671, 53329	Competitive Results obtained
Wang et al. (2020)	Breast cancer image classification	Deep convolutional neural network model (DCNN)	400	Accuracy of 97.70% achieved
Zhou et al. (2020)	To implement a model to classify skin lesion images.	A CNN architecture based model called spiking neural networks (SNNs)	1081	Accuracy of 83.80% achieved

Cao et al. (2020)	For lung nodule detection	A two-stage Convolutions Neural Network (TSCNN)	1186	Competitive Results obtained
Sun et al. (2020)	For histopathologic cancer detection.	A regular CNN architecture.	220,025	Accuracy of 97.94% achieved
Ciocca et al. (2020)	For state recognition of food images	A regular CNN architecture in combination with Support Vector Machines (SVM)	11943	Competitive Results obtained

3 Methodology

As per this research, step by step process is followed to implement proposed CNN models to identify and classify pests that are formed on the leaves of tomato plants. A step by step process is followed in this research and the stages that are part of this exercise are data pre-processing and transformation, research methodology and evaluation. All models that are applied to the selected data set will be evaluated by a set standard metrics.

3.1 Data acquisition and description

The data for this research is obtained from the Technological Centre CTC [2] which is a private non -a profit-making organization established in the year 2000 in Spain and the main motive of this company is to develop innovative solutions in the fields of energy, robotics and advanced materials.

In the data set acquired for this research, there are 4300 images of different type of pests and insects. The annotation files for each image is created are already a part of the data set acquired. These annotation files are named in such a way that the name of an annotation file in .xml format is like its corresponding .jpg file. Hence there are 4300 annotation files which can be further used in data pre-processing and data transformation stages. These images are captured in artificial cultivation chambers with the help of an automatic dataset generator and a colour camera namely GigE UI-5240CP as per Gutierrez et al. (2019).

3.2 Image pre-processing and transformation

The image pre-processing stage is crucial in this research as important metrics such as accuracy, precision, recall and F1 score are largely impacted by the quality of labelling the data set. The image data set is uploaded into Google drive and set of python scripts are used to download both the annotations(labels) and images into to Google Colab environment The images are cropped as per the coordinates identified in annotation files. The bounding boxes are extracted from each image and the respective class label, coordinates are parsed from each annotation file. The cropped images are labelled as per the bounding boxes in the annotation files.

[2]Data Source: https://centrotecnologicoctc.com/en/who-we-are-ctc/

A sample raw image taken from the data set is shown in Figure 2. From the figure, it can be noted that the size of insects or eggs is very minute and the classification task through deep learning models is challenging. An image may contain multiple types of pests or insects and the respective areas in an image can be identified with the help pixel values mentioned in the annotation files.

Figure 2: Raw Image from the Data Set

The image pre-processing stage is crucial in this research as important metrics such as accuracy, precision, recall and F1 score are largely impacted by the quality of labelling the data set. The images are cropped as per the coordinates identified in annotation files. The diagram shown in the Figure 3 describes various steps involved in this stage. The data set selected for this research is uploaded on to google drive and then fetched into the working directory in Google Colab. The next step is to extract the bounding boxes with their coordinates and the corresponding labels to prepare training and testing data. The bounding boxes are extracted from each image and the respective class label, coordinates are parsed from each annotation file.

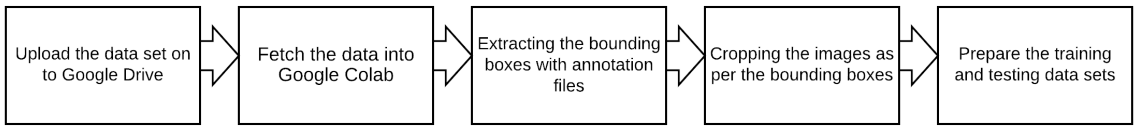

Figure 3: Data pre-processing Stages

In total there 56205 cropped images are generated and tabulated as follows. The Table 2 explains the type of insect/egg included as part of this analysis and description as well as their count after cropping the images as per the annotation files. There are 6 categories including normal images of tomatoes are selected as part of this research and the total cropped images are segregated based on the respective category.Various metrics that can be derived from a confusion matrix can be defined as follows.

wf: Trialeurodes vaporariorum eggs are also called as greenhouse whitefly. Controlling these type pests is difficult as generally, they are reluctant to pesticides as per [3].

[3]wf reference: https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/trialeurodes-vaporariorum

Cleaning of weed around the leaves of tomato plants can help in controlling.

egg_wf: Eggs laid by Trialeurodes vaporariorum insects

bt: Bemisia tabaci (bt) is a polyphagous insect that normally survives in the humid and hot conditions in tropical and subtropical areas. This type of insects spread gradually and is observed on many crops especially vegetables and cotton [4].

egg_bt: Eggs laid by Bemisia tabaci insects.

tomato: Normal images of tomatoes.

Table 2: Category wise Pest Count

Disease Category	Size
egg_wf	25,834
egg_bt	9,713
egg_ta	4
wf	13,511
bt	6,507
tomato	636
Total	56,205

4 Design Specification

In general, it may take days or even weeks to train a deep learning model and high configurations of hardware are required to carry out the processing of huge amounts of data. An alternative approach is transfer learning in which a model is trained on a benchmark data set and the weights of the model can be reused further by other models. We can also use specific blocks of layers of pre-trained models and can integrate these layers to other models to increase the efficiency in terms of accuracy and other standard metrics. The detection and classification of pests are implemented based on the concept of transfer learning and a CNN implemented from scratch with a couple of convolutional layers. The basic design of the models can be explained in two stages

Pre-trained layer extraction: In this stage individual blocks of convolutional layers from the selected pre-trained models will be chosen.

[4]bt reference: https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/bemisia-tabaci

Common block of convolutional layers: The block of layers from stage 1 is integrated into a common block of additional convolutional layers to create the final neural network.

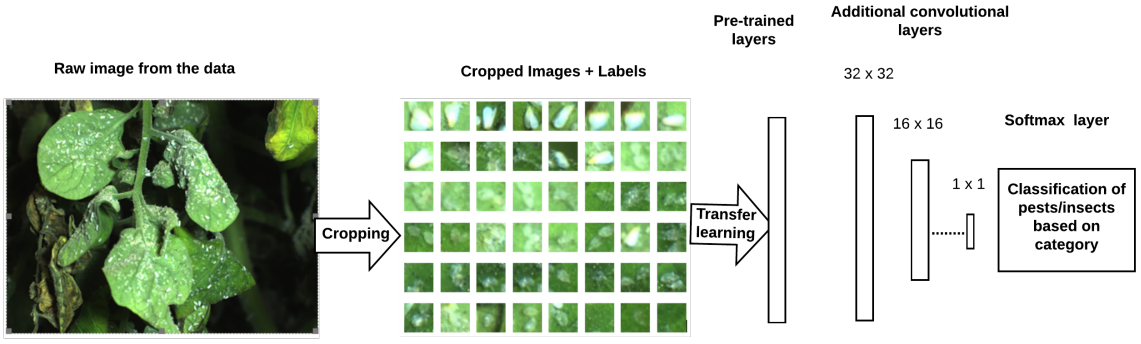

Figure 4: High-level architecture of a hybrid model

The Figure 4 shows the overview of various stages of the processing. A raw image will be cropped into a number of smaller images and labelling to each cropped image is done with few python scripts. The cropped data created is passed to a specific block of a pre-trained model and the output that is obtained will be the input for the second stage of the common block. The final layer is a dense layer of size six with softmax as the activation function. The model trained in this way will be used for prediction of the type of pests and further evaluation.

5 Implementation

5.1 Pre-trained layer extraction

In total six models are implemented for the comparative analysis of detection and classification of pests on the tomato plants. Five different CNN architectures are developed with above-mentioned blocks of pre-trained models and a common block of additional layers and these models are evaluated with a set of standard metrics. The sixth model is a convolutional network with custom layers with softmax as the activation function.

The pre-trained models considered for this analysis are VGG16, VGG19, InceptionV3, Xception and ResNet50. The initial block of these models is reused as part of transfer learning and the performance of each model is compared with others for accuracy in terms of classification. The Table 3 describes the layers of each pre-trained model used along with one common block of additional convolutional layers. The basic idea here is that the initial layers of pre-trained models can identify shapes like lines, squares, circles and other general shapes. The data will be initially trained on these layers and the output obtained will be passed as an input to the final block of additional convolutional layers.

Table 3: Transfer learning layers

Model	Initial Block	Number of parameters	convolutional (2D) layers
VGG16	block2_pool	260,160	4
VGG19	block1_pool	38,720	2
InceptionV3	conv2d_3	28,576	3
Xception	block1_conv2	19,360	2
ResNet50	conv1	9,472	1

5.2 Common block of convolutional layers

The Figure 5 shows the two stages of the basic architecture of models implemented through transfer learning as part of this research. As per the figure, the initial layer with 260160 parameters is the pre-trained layer of VGG16. The rest of the block contains additional convolutional and dense layers. This block is a common component integrated to all the models implemented with transfer learning.

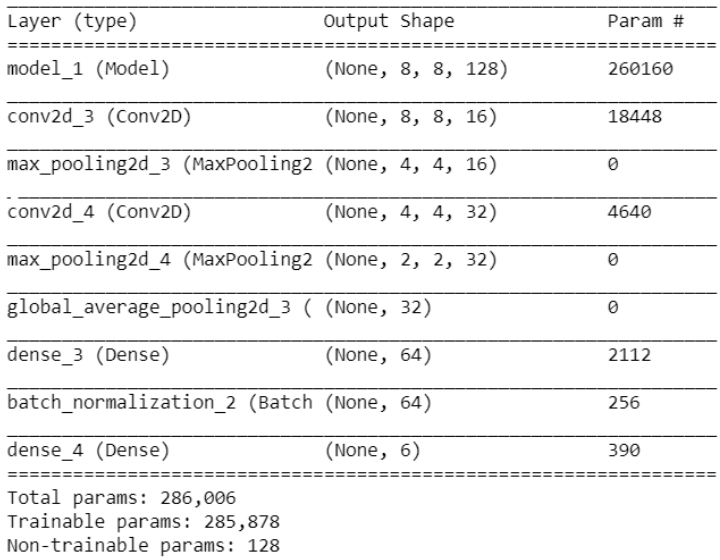

Figure 5: Final architecture of customized VGG16

6 Evaluation

There are three methods chosen for evaluating the performance of the models and results obtained as part of this analysis. The first approach is to analyse the graphs drawn for accuracy and loss of both the testing and training datasets. Secondly, evaluating the results obtained based on a confusion matrix. Finally, the models can be validated

against a few images from the validation data to analyse the predictions from the models versus the ground reality.

The classification of pests on tomato plants, once they are detected, is a classification problem. Hence evaluation of the models can be carried out through a standard set of metrics derived from a confusion matrix. The Figure 6 is extracted from Bittrich et al. (2019) in which a classification model is implemented for classification of early folding residues during protein folding. Primary metrics such as accuracy, precision, recall and F1 score can be derived from the confusion matrix and the formulae for all of these are given in the form of equations 1, 2, 3 and 4

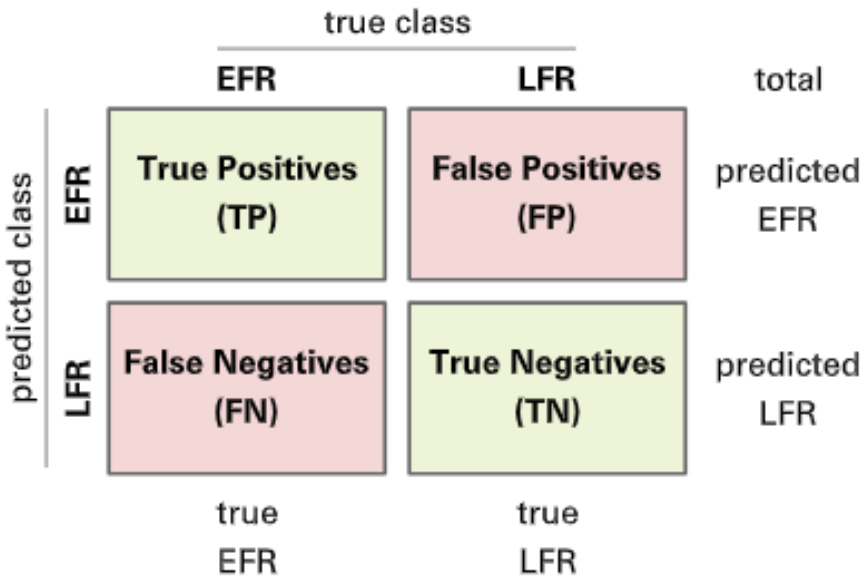

Figure 6: Confusion matrix

Various metrics that can be derived from a confusion matrix can be defined as follows.

Accuracy:

Accuracy is the proportion of the predictions which are correct to the overall predictions made by a machine learning model.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

Precision:

Precision measures the efficiency of the model in terms of predicting a positive class.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall:

The recall is the proportion of actual positive classes that are predicted correctly by the model.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1 score: F1 score shows the overall efficiency of the model by combining precision and recall. In other words, F1 score can be calculated by taking the mean of precision and recall.

$$F1score = \frac{2TP}{2TP+FP+FN} \tag{4}$$

6.1 Accuracy and loss variation

The graphs shown below diagrammatically represent the accuracy and loss variation for both testing and training datasets for all the models implemented as part of this research.

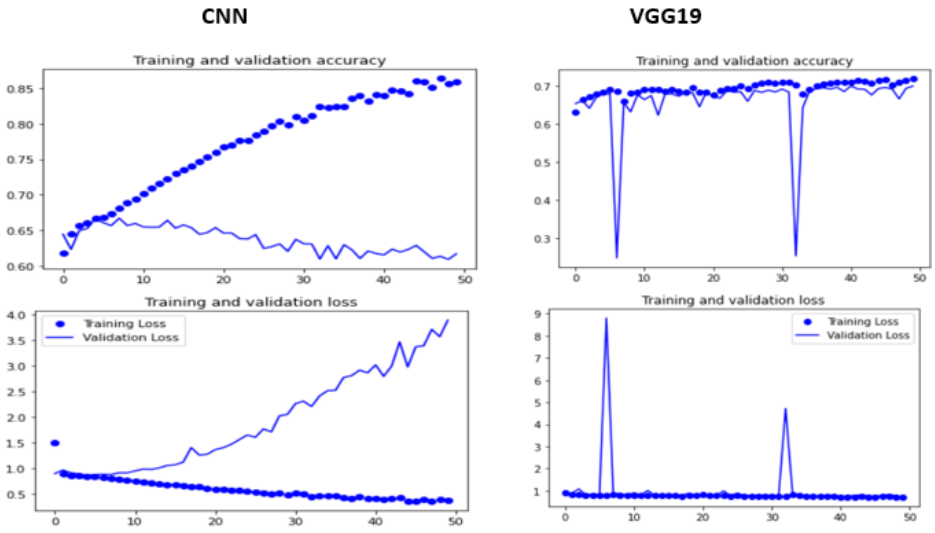

Figure 7: Accuracy and loss variation for CNN and VGG19

The big gap between training and testing curves for models CNN,Xception as per Figure 7, Figure 8 and Figure 9 shows respective model's over fitting. The over fitting is slightly reduced in the model Inception V3 as the curves are separated by a little space when compared to others.The model built with Inception V3 is relatively better since both the training and validation loss decreased steadily whereas the training and validation accuracy increased gradually. The loss for both training and validation datasets is decreasing in case of Inception V3 whereas the accuracy is steadily increased with the number of epochs is increasing for training dataset. Though there are few fluctuations for validation dataset in terms of accuracy.

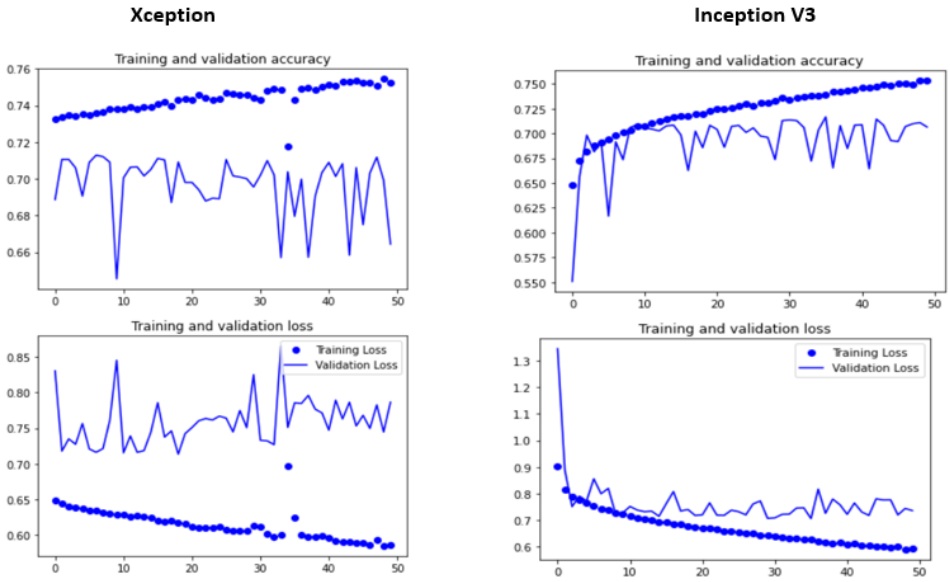

Figure 8: Accuracy and loss variation for Xception and Inception V3

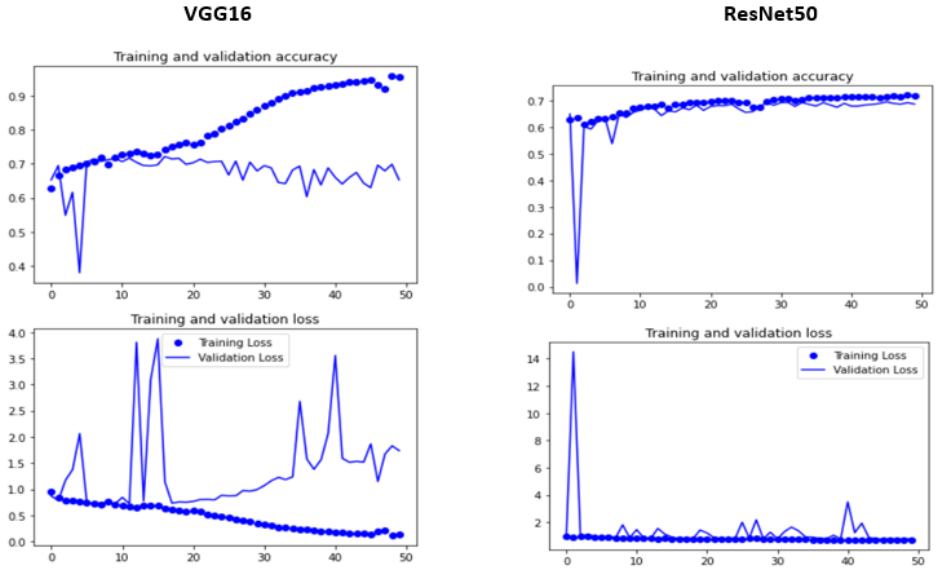

Figure 9: Accuracy and loss variation for VGG16 and ResNet50

6.2 Metrics obtained from confusion matrix

The data filled in Table 4 shows the core results that are obtained as part of this analysis. Detection and classification of the type of pests is a classification problem and can be evaluated with metrics such as accuracy, precision, recall and F1 score. As per the table, the model which is built with a single block from VGG16 and a common block of convolutional layers shows an accuracy of 0.95, the validation accuracy of 0.65 shows

that the model is slightly over fitted.The images in the current dataset are cropped and labelled according to the annotation files. The classes are not properly balanced and the same can be observed in the Table 2. In these scenarios, the F1 score is the better metric compared to accuracy to evaluate the overall goodness of fit of the models implemented. As per the Table 4, Inception V3 model records better F1 score compared to other models except for Xception.

Table 4: Core evaluation metrics

Model	Training Accuracy	Validation Accuracy	Precision	Recall	F1 Score
VGG16	0.95	0.65	0.55	0.54	0.55
VGG19	0.71	0.70	0.45	0.46	0.41
InceptionV3	0.75	0.70	0.70	0.66	0.66
Xception	0.75	0.66	0.67	0.66	0.66
ResNet50	0.71	0.68	0.44	0.47	0.40
CNN	0.85	0.61	0.48	0.44	0.45

A comparison of training and validation accuracies for the models implemented are compared with the help of bar charts shown in the Figure 10. From this figure, it can be inferred that all the models showing a similar effect when it comes to validation accuracy.

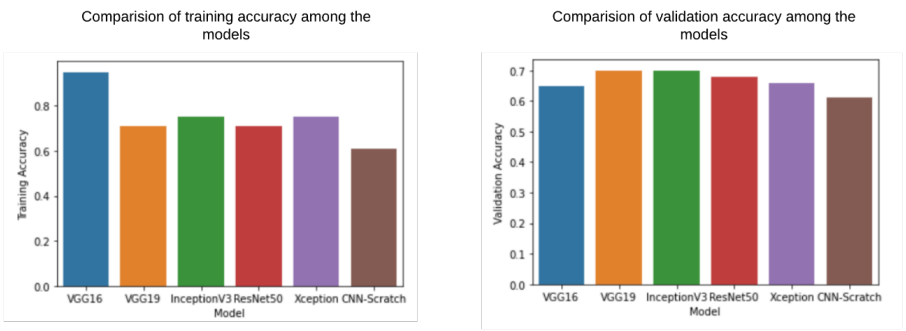

Figure 10: A comparison of training and validation accuracies

6.3 Actual vs predicted analysis

When it comes to prediction, figures Figure 11 and Figure 12 show the prediction results when the validation data is passed to the models implemented as an input. Each image is associated with the respective prediction output. The output text which is in blue indicates the correct prediction by the model whereas the one in red tells that the prediction by the model is inaccurate. Also, for each prediction, the probability is given as well in the parenthesis.

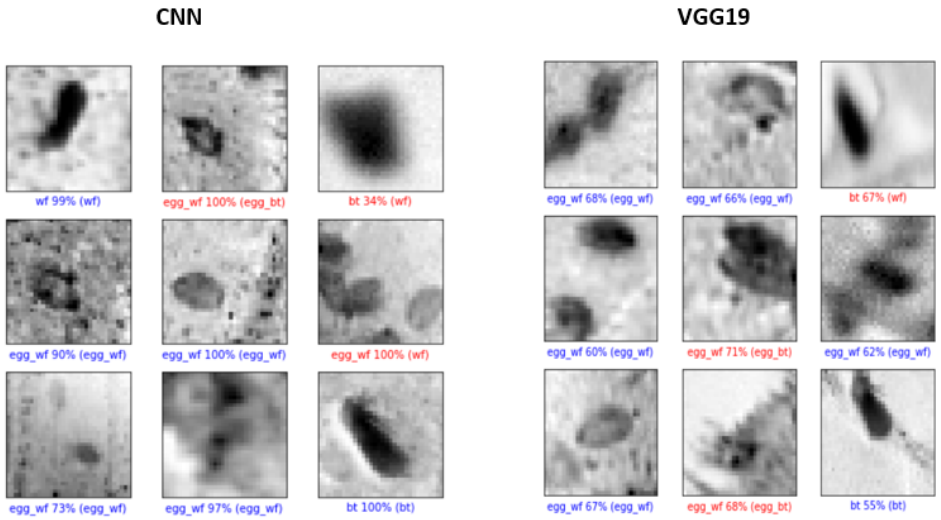

Figure 11: Actual vs Predicted for CNN and VGG19

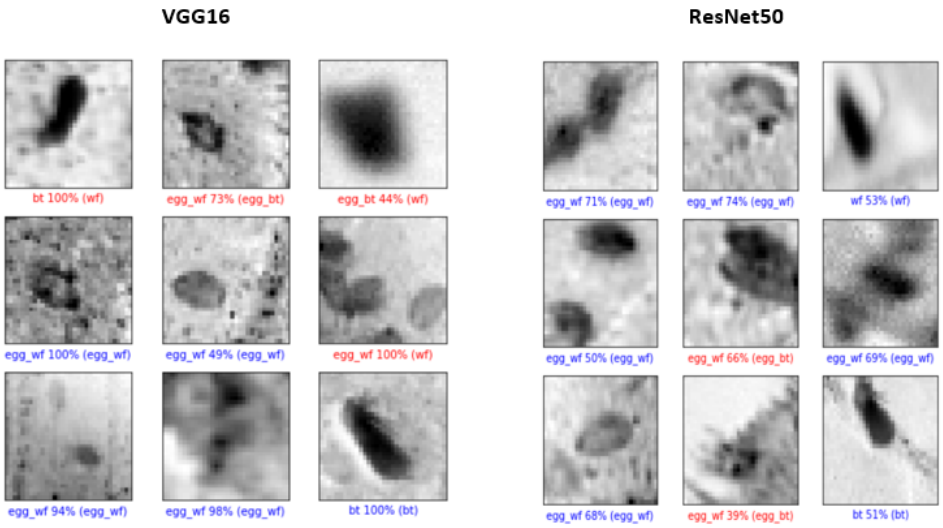

Figure 12: Actual vs Predicted for VGG16 and ResNet50

7 Conclusion and Future Work

Detection and classification of pest on tomato plants through CNN is crucial in terms of identifying the pests in the early stages of the crop and increasing the overall productivity. A set of 2 stages CNN models by leveraging the initial block of various pre-trained models and a common layer of additional convolutional layers are implemented and the results are evaluated with a standard set of metrics. An impressive 0.95 training accuracy is obtained with transfer learning using VGG16 though the model is slightly over fitted. There is a good scope for future work in this area such as integrating the implemented models to scouting robots and the performance of the model in the real-time scenarios

can be evaluated. Also, a framework can be developed like an image can be passed to the framework and various types of pests and their proportion can be tested so that necessary preventive measures can be taken to reduce the usage of pests and increase the productivity.

References

Batool, A., Hyder, S. B., Rahim, A., Waheed, N., Asghar, M. A. et al. (2020). Classification and identification of tomato leaf disease using deep neural network, *2020 International Conference on Engineering and Emerging Technologies (ICEET)*, IEEE, pp. 1–6.

Bittrich, S., Kaden, M., Leberecht, C., Kaiser, F., Villmann, T. and Labudde, D. (2019). Application of an interpretable classification model on early folding residues during protein folding, *BioData mining* **12**(1): 1.

Cao, H., Liu, H., Song, E., Ma, G., Jin, R., Xu, X., Liu, T. and Hung, C.-C. (2020). A two-stage convolutional neural networks for lung nodule detection, *IEEE Journal of Biomedical and Health Informatics* .

Ciocca, G., Micali, G. and Napoletano, P. (2020). State recognition of food images using deep features, *IEEE Access* **8**: 32003–32017.

Eigen, D., Krishnan, D. and Fergus, R. (2013). Restoring an image taken through a window covered with dirt or rain, *Proceedings of the IEEE international conference on computer vision*, pp. 633–640.

Feng, J., Wu, X., Chen, J., Zhang, X., Tang, X. and Li, D. (2019). Joint multilayer spatial-spectral classification of hyperspectral images based on cnn and convlstm, *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp. 588–591.

Gutierrez, A., Ansuategi, A., Susperregi, L., Tubío, C., Rankić, I. and Lenža, L. (2019). A benchmarking of learning strategies for pest detection and identification on tomato plants for autonomous scouting robots using internal databases, *Journal of Sensors* **2019**.

Huang, Z., Zhu, X., Ding, M. and Zhang, X. (2020). Medical image classification using a light-weighted hybrid neural network based on pcanet and densenet, *IEEE Access* **8**: 24697–24712.

Jiménez, M., Torres, M. T., John, R. and Triguero, I. (2020). Galaxy image classification based on citizen science data: A comparative study, *IEEE Access* **8**: 47232–47246.

Laishram, A. and Thongam, K. (2020). Detection and classification of dental pathologies using faster-rcnn in orthopantomogram radiography image, *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, pp. 423–428.

Metwalli, A.-S., Shen, W. and Wu, C. Q. (2020). Food image recognition based on densely connected convolutional neural networks, *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, IEEE, pp. 027–032.

Mou, L., Lu, X., Li, X. and Zhu, X. X. (2020). Nonlocal graph convolutional networks for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing* .

Pang, B., Zhang, Y., Chen, Q., Gao, Z., Peng, Q. and You, X. (2010). Cell nucleus segmentation in color histopathological imagery using convolutional networks, *2010 Chinese Conference on Pattern Recognition (CCPR)*, IEEE, pp. 1–5.

Park, J.-Y., Hwang, Y., Lee, D. and Kim, J.-H. (2020). Marsnet: Multi-label classification network for images of various sizes, *IEEE Access* **8**: 21832–21846.

Peng, B., Chen, L., Shang, M. and Xu, J. (2018). Fully convolutional neural networks for tissue histopathology image classification and segmentation, *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 1403–1407.

Rahman, M. M., Islam, M. S., Jannat, M. K. A., Rahman, M. H., Arifuzzaman, M., Sassi, R. and Aktaruzzaman, M. (2020). Eyenet: An improved eye states classification system using convolutional neural network, *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, IEEE, pp. 84–90.

Roth, H. R., Lee, C. T., Shin, H.-C., Seff, A., Kim, L., Yao, J., Lu, L. and Summers, R. M. (2015). Anatomy-specific classification of medical images using deep convolutional nets, *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, IEEE, pp. 101–104.

Song, I., Kim, H.-J. and Jeon, P. B. (2014). Deep learning for real-time robust facial expression recognition on a smartphone, *2014 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, pp. 564–567.

Sun, Y., Hamzah, F. A. B. and Mochizuki, B. (2020). Optimized light-weight convolutional neural networks for histopathologic cancer detection, *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*, IEEE, pp. 11–14.

Sun, Y., Xue, B., Zhang, M. and Yen, G. G. (2019). Evolving deep convolutional neural networks for image classification, *IEEE Transactions on Evolutionary Computation* **24**(2): 394–407.

Wang, Y., Lei, B., Elazab, A., Tan, E.-L., Wang, W., Huang, F., Gong, X. and Wang, T. (2020). Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning, *IEEE Access* **8**: 27779–27792.

Zhou, Q., Shi, Y., Xu, Z., Qu, R. and Xu, G. (2020). Classifying melanoma skin lesions using convolutional spiking neural networks with unsupervised stdp learning rule, *IEEE Access* .