Project plan for Degree Project

Version

PA2534: Research Methodology in Software Engineering

March 27, 2022

Thesis	Tentative title	Pepper Leaf Disease Detection using Convolutional Neural Network
	Classification	Convolutional Neural Network, Deep Learning, Pepper Leaves Disease Detection
Student 1	Name	Mohit Battu
	e-Mail	mobt20@student.bth.se
	Social security nr	991007-T175
Student 2	Name	Sai Chetan Poluri
	e-Mail	sapo20@student.bth.se
	Social security nr	980501-0213
Student 3	Name	Monica Gattupalli
	e-Mail	moga20@student.bth.se
	Social security nr	991130-T308
Student 4	Name	Gowtham Kumar Sandaka
	e-Mail	gosa20@student.bth.se
	Social security nr	001117-T071
Supervisor	Name and title	Nauman Ghazi, Senior Lecturer
	e-Mail	nauman.ghazi@bth.se
	Department	Department of Software Engineering

1 Introduction

The main issue that farmers face is detecting the disease at the final stages of the crop, which causes a wide spread of disease throughout the farm, resulting in crop damage. Currently the impact of diseases is identified by taking an infected leaf to a laboratory and detect the disease. In order to identify how many plants are infected with the disease, the farmer has check each and every plant manually and based on that he must spray the insecticides, which is a very big process and takes loads of time [1].

To reduce the efforts of continuous monitoring of crops agricultural scientist are conducting many experiments to bring out the best solution to the farmers. Few computer scientist have made this work easier by employing advanced concepts such as Artificial Intelligence, Machine Learning, and Deep Learning [2]. These advanced tools assisted in detecting diseases in the leaves at an early stage, allowing farmers to remove the infected crop before it spread throughout the farm. Further the tools also provides clear statistical data to farmers, allowing them to take the necessary actions. We are going to use the concept of Convolutional Neural Networks (CNN) which belongs to the family of Deep Learning [3].

Many researchers have experimented using CNN techniques as it give promising results with great accuracy especially in the applications such as data obtained from satellites, systems of surveillance, medical treatment and digitalized judicial system. It is used to reduce the uncertainty of image recognition to some extent. CNN tried to take us to the future due to its robust feature extraction, information mining and learning mechanism. It is a common technique used in a variety of applications like object recognition, image super-resolution, semantic segmentation, predictions etc [4].

The Convolution Neural Network is made up of several layers, including pooling, fully connected and convolutional layers. The filters, such as kernels or weights, are presented in each layer of the convolution network, which emits a similar number of feature maps by sliding the filters through feature maps from the previous layer [5]. The general architecture of Convolutional Neural Network is as shown in Figure 1.

The input data contains a collection of images depicting various types of pepper leaves. These images are fed into the Convoultional Neural Network, which is used to train and test the model's performance. The Convolution Neural Network training consists in two phases called as Forward and Backward. In forward phase every layer caches the data such as inputs and weight values that is needed for the backward implementation.

As a result, every backward implementation must be preceded by a forward implementation. Each layer will receive and return a gradient during the backward phase. It will receive the loss gradient with respect to its outputs and return the loss gradient with respect to its inputs. Thus resulting in updating the weights of a Convolutional Neural Network thereby increasing the efficiency of predicting the leaf disease detection of a pepper.

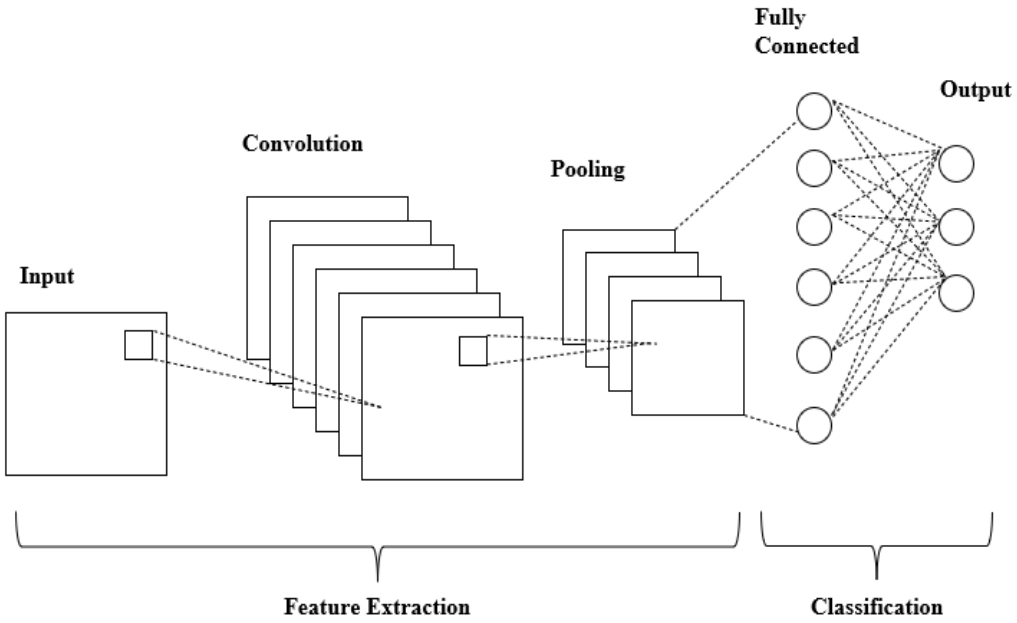

Figure 1: CNN Architecture

2 Related work

Many experiments are conducted on detecting the disease at an early stage in plants. Most of them are achieved by using Artificial Intelligence and Machine Learning Algorithms. Some of such experiments to detect the infected or diseased leaves using CNN are discussed below.

Two different experimentation were done, one on Bell Pepper leaves and other on Corn leaves. Both of these experimentation is done by using CNN techniques like VGG16, ResNet50, and InceptionV3 for identifying disease type, disease area and estimating severity. The experimentation on pepper leaves is done using a data set with 2442 images with around 99% accuracy [6] but the experimentation on corn leaves is done using a data set with 59000 images with around 98.34% accuracy. [7] Both the experimentation resulted in good accuracy but experimentation on corn can be considered as more appropriate because the data set used in corn experimentation is very large than the data set used on bell pepper.

The author has used Deep Learning with Convolutional Neural Networks (CNN) to classify the leaf diseases in Bell Pepper. They have used image data sets to train, Keras and tensor flow libraries were used to analyse the images and further it classifies if the leaves are healthy are not with an accuracy of 96.78%. Thus the algorithm produced by them helps in classifying the leaves as Healthy or Infected [8]. But the experimented results should have been more clear to confirm their accuracy.

The concept of Deep Learning with CNN in detecting and classifying the infected leaves is used. They have experimented on three plants namely tomatoes, pepper and potatoes. They have used a data set containing 20636 images of leaves and they have classified them into 15 classes (12 for infected leaves and 3 for healthy leaves) using Convolutional Neural Network (CNN) with an accuracy of 98.029% [9]. The experimentation result would have been more accurate if the data set used train was at least 20000 in each category.

According to research papers cited above, we can observe that most experiments in the related area rely on Convolutional Neural Network (CNN) concepts, which proved to provide promising results as well as an unexpected degree of accuracy, However the Detailed Cross Validation measures would have resulted in strong claim of their mentioned accuracies.

3 Aim and Objectives

Aim:

The aim of the thesis is to optimize the performance of detecting the diseases in the pepper leaf by using Convolutional Neural Network. The following objectives are considered in order to select the best performance model for pepper leaf disease detection:

Objectives:

- Configuring several sorts of parameters and activation functions such as sigmoid, relu, and softmax to construct the optimal Convolutional Neural Network model by recognizing the elements that have the least impact on the efficacy of Convolutional neural network's.
- Examining the performance measures of the traditional Convolutional Neural Network and various transfer learning architectures of Convolutional Neural Network using performance metrics.

4 Research Question

- **RQ1: Which of the Convolutional Neural Network transfer learning models has the highest accuracy for detecting diseases in pepper leaves ?**
- **RQ2: What hyper-parameters are to be considered in Convolutional Neural Network model for yielding the best accuracy ?**

5 Research Methodology

The first research question can be answered based on both systematic literature review and experiment review, and the second research question can be satisfied by conducting the experiment.

RQ 1: To select the best transfer learning models for handling the image data a systematic literature review technique is used. As the part of the systematic literature the search terms will be "heart diseases", "Convolution neutral network" and the criteria of search will be "Deep learning techniques in medical field". Entire process of the SLR performed in the experiment is given in the Figure-2. .

As a part of the Experiment a new CNN architecture is designed and the "pepperleafdata" dataset is obtained from Kaggle in order to carry out the experiment. Data cleaning and different pre-processing techniques are performed on the dataset to remove null values, handle

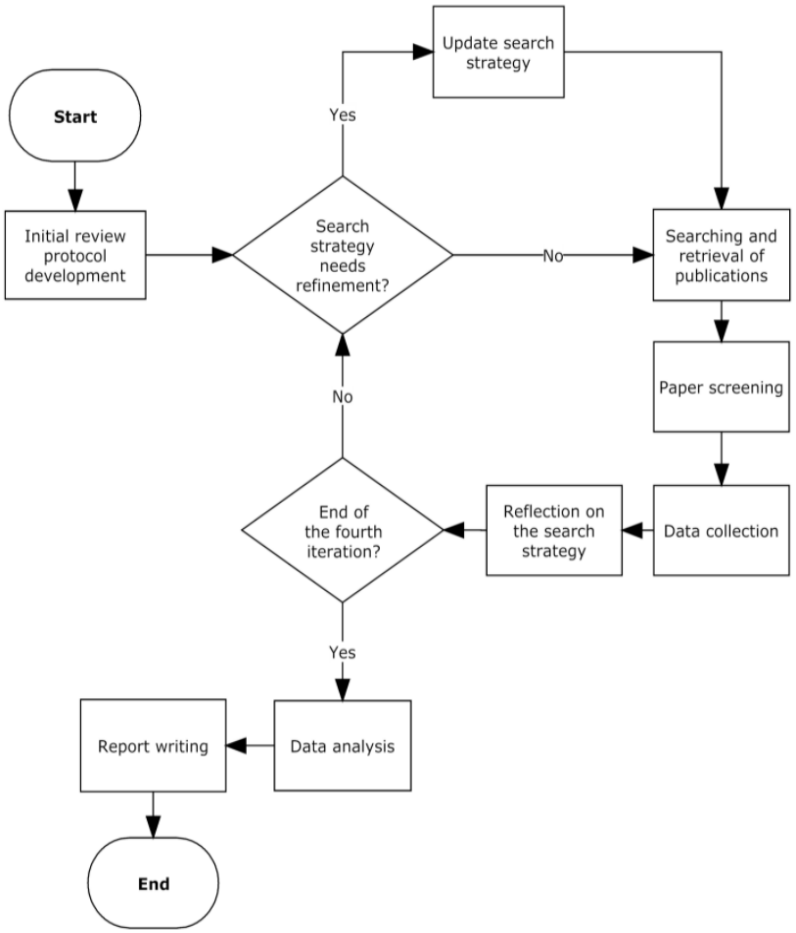

Figure 2: Systematic Literature Review Architecture

missing values, normalize the data, sharpen images and rescaling are performed to standardize the image data. Based on the quality of dataset obtained after applying pre-processing techniques we have to implement the convolutional neural network architecture.The workflow of the Experiment design can be seen in the Figure-3.

A four-layered convolution neural network architecture is designed as discussed above, with Max pooling of window size (2 X 2) and Batch Normalization layers follows each layer. Each layer has 64 kernels, and the filter window size is (3 x 3). Every layer employs the 'Relu' activation function to eliminate negative pixel values. Finally, the flatten layer receives all of the output feature maps. In the flatten layer, a dense layer of 64 units with the 'ReLU' activation function is used to gain information from the previous layers of the neural network. A dropout layer is created using a dense layer of two units and an activation function called a 'sigmoid' is used as the final layer of the architecture to classify the data accurately.

The selected algorithms and the designed CNN architectures consisting of four layers are trained and tested against the chosen dataset, where 50 epochs are used to train the three architectures [10]. These epochs aid in the extraction of minute information from images, as well as the construction of efficient feature maps, which aid in disease prediction. We can analyze the better performing architecture using epochs by calculating accuracy, precision, and recall.

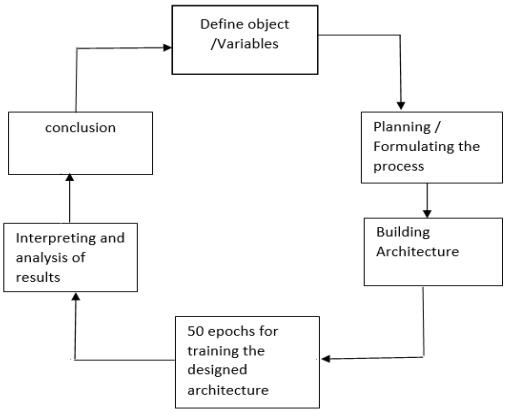

Figure 3: Experimentation Architecture

RQ 2: We are using Hyper parameter tuning technique to attain the better performance from the architectures. The designed 4 layered architecture gives the scope to change the parameters such as activation function from "ReLU" to "sigmoid", number of convolution layers from 4 to 3, number of units 64 unit to 32 units, number of epochs from 50 to 80, batch size from 16 to 8, dropout layer value from 0.5 to 0.2 etc when compared to the transfer learning architecture they don't have a possibility to change those parameters because they are pre-trained and imported from the libraries. Changing these parameters significantly affect the performance of the general CNN architecture. In this, research hyper-parameter tuning is applied on the designed architecture of the CNN to yield the better accuracy.

6 Expected Outcomes

The expected outcomes from this Thesis are given below,

EO1: Performing experiment on the selected two transfer learning techniques and designed CNN architecture on the dataset gives an vision about the CNN performance.

EO2: Hyper-parameter tuning helps in getting an insight about the parameters of the algorithm, with this technique one can identify the effect and dependency of the each parameter on different datasets.

7 Study Design

The elements given below are considered for the Study Design.

Subject: Pepper leaf image data is utilized in the application such as disease affected pepper leaf images and phytopathology research is considered as a subject. The application developed and executed by utilizing CNN prediction models that have been optimized to anticipate outcomes reliably and precisely.

Object: The objective is to get the accuracy of the result derived from the Convolution neural networks model of the selected data-set comprising pepper leaf image data.

Variables: The activation functions, learning rate, optimizing algorithms such as RMS prop used for the CNN Architecture and the data set including pepper leaf image data utilized for training Convolution Neural Networks.

Factors: The data set utilized for training, transfer learning methods, pre-processing techniques, and the activation function for generating a Convolution Neural Networks are the primary factors that might influence the results of this experiment. These variables will have an impact on the experiment's outcome.

Investigation: The CNN layer of the chosen algorithms is investigated in order to determine the total trainable amount of parameters by using the Hyper-parameter tuning. This identifies the effect and dependency of each parameter on different datasets. The outcomes and errors are analyzed when the experiment is completed.

8 Time and activity plan

Start Date	End Date	Activity	Week Activity Details
21-01-2022	15-02-2022	Topic Preparation	It is essential to read the relevant articles and understand the project specifications thoroughly.
16-02-2022	26-03-2022	Preparing the project plan	Research is conducted to gather the relevant information and write the project proposal.
27-03-2022	27-03-2022	Submitting the project plan	After submitting the project plan some of the required improvements will be done based on suggestions.
27-03-2022	31-03-2022	Gathering previous literature works	The project work begins with the collection of data that is relevant to work by doing some research.
01-04-2024	05-04-2022	Structuring of the method	Collecting all relevant information and structuring the detailed methodology and algorithms for modeling the project.
06-04-2022	11-04-2022	Experimentation and modelling solution	The modelling of the solution is done by using necessary research approaches according to the outcome.
12-04-2022	25-04-2022	Analysis of the results	Analyse the result solutions and making necessary changes based on the research questions.
26-04-2022	12-05-2022	Evaluating the outcomes	The collected results will be checked and evaluated, if necessary, they are re-executed.

12-05-2022	18-05-2022	Documentation of the required outcomes	Outcomes will be recorded according to the research problem and attaching additional criteria to the execution.
19-05-2022	14-06-2022	Preparation of thesis draft, Presentation of the thesis and finally submitting the report	Preparation of the final thesis draft and necessary changes will be implemented based on supervisor suggestions.

Table 1: Time and Activity Plan

9 Risk management

There is no system in this world that is 100% perfect. Every system shows some errors. There are many factors and constrains that might reduce the efficiency of the system. Some of the constrains identified in our system are tabulated below. Because the data set chosen is relatively mild, the task execution time is also reasonable. There are several approaches to perform, such as by employing ML algorithms. As a result, it is not so ubiquitous that just one sort of approach is used. Similarly, the likelihood of risk is low. Table 2 shows the constrains involved in this system and some possible ways that can be implemented to overcome them.

Risk	**Possibility**	**Severity**	**Diminution**
Memory Constraints	High	Medium	We can use the Cloud Platforms for storing and processing the data.
Distortion Outcomes	Medium	High	Normalizing the data to increase the accuracy.
Complex Computing Issues	Low	Medium	Training the models in the cloud platform to reduce the time factor and model computational complexity with the increased amount of CPU and GPU resources.
Access to data	Medium	Medium	Different sources of data can be combined.
Quality control	Medium	High	Make sure to select the most appropriate algorithms that to give the better results.

Table 2: Risk Management

References

[1] Eftekhar Hossain, Md Farhad Hossain, and Mohammad Anisur Rahaman. A color and texture based approach for the detection and classification of plant leaf disease using knn

classifier. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pages 1–6. IEEE, 2019.

[2] Jay Trivedi, Yash Shamnani, and Ruchi Gajjar. Plant leaf disease detection using machine learning. In *International conference on emerging technology trends in electronics communication and networking*, pages 267–276. Springer, 2020.

[3] SERT Eser. A deep learning based approach for the detection of diseases in pepper and potato leaves. *Anadolu Tarım Bilimleri Dergisi*, 36(2):167–178, 2021.

[4] Manali Shaha and Meenakshi Pawar. Transfer learning for image classification. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 656–660. IEEE, 2018.

[5] K Ramanjaneyulu, K Veera Swamy, and CH Srinivasa Rao. Novel cbir system using cnn architecture. In *2018 3rd International conference on inventive computation technologies (ICICT)*, pages 379–383. IEEE, 2018.

[6] Prateek Thakur, Anuradha Chug, and Amit Prakash Singh. Plant disease detection of bell pepper plant using transfer learning over different models. In *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 384–389. IEEE, 2021.

[7] Aanis Ahmad, Dharmendra Saraswat, Aly El Gamal, and Gurmukh S Johal. Comparison of deep learning models for corn disease identification, tracking, and severity estimation using images acquired from uav-mounted and handheld sensors. In *2021 ASABE Annual International Virtual Meeting*, page 1. American Society of Agricultural and Biological Engineers, 2021.

[8] Monu Bhagat, Dilip Kumar, Rehan Mahmood, Bibhudhendra Pati, and Monu Kumar. Bell pepper leaf disease classification using cnn. In *2nd International Conference on Data, Engineering and Applications (IDEA)*, pages 1–5. IEEE, 2020.

[9] Marwan Adnan Jasim and Jamal Mustafa AL-Tuwaijari. Plant leaf diseases detection and classification using image processing and deep learning techniques. In *2020 International Conference on Computer Science and Software Engineering (CSASE)*, pages 259–265. IEEE, 2020.

[10] Baoqi Li and Yuyao He. An improved resnet based on the adjustable shortcut connections. *IEEE Access*, 6:18967–18974, 2018.