*A Minor Project Report*

*on*

**Plant Leaf Disease Identification Model Using Machine learning And Genetic Algorithm**

*carried out as part of the course CS1634 Submitted by*

***Shubhra Bhunia Ghosh***

***179301197***

***VI-CSE***

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

In

**Computer Science & Engineering**



**Department of Computer Science & Engineering,**

**School of Computing and IT,**

**Manipal University Jaipur,**

***June, 2020***

**CERTIFICATE**

This is to certify that the project entitled “Plant Leaf Disease Identification Model Using Machine learning And Genetic Algorithm” is a bona fide work carried out as part of the course Minor Project CS-1634, under my guidance by Shubhra Bhunia Ghosh, student of Bachelor Of Technology (B.Tech.) in Computer Science & Engineering (CSE) at the Department of Computer & Communication Science & Engineering , Manipal University Jaipur, during the academic semester -VI of year 2019-20.

Place:

Date: Signature of the Instructor (s)

**DECLARATION**

I hereby declare that the project entitled “Plant Leaf Disease Identification Model Using Machine learning And Genetic Algorithm” submitted as part of the partial course requirements for the course Minor project (CS-1653), for the award of the degree of Bachelor of Technology in Computer Science & Communication Engineering at Manipal University Jaipur in the semester during academic year 2019-20, has been carried out by me. I declare that the project has not formed the basis for the award of any degree, associate ship, fellowship or any other similar titles elsewhere.

Further, I declare that I will not share, re-submit or publish the code, idea, framework and/or any publication that may arise out of this work for academic or profit purposes without obtaining the prior written consent of the Course Faculty Mentor and Course Instructor.

Signature of the Student:

Place:

Date:

**Abstract**

Agriculture is one of the prime sources of the economy and a large number of people are involved in cropping various plants based on the environmental conditions. A number of challenges are faced by the farmers including different types of plant diseases. The detection and prevention of plant diseases is a serious concern and should be treated well on time for increasing productivity. Therefore, an automated plant disease detection system can be more beneficial than individually checking plants for diseases and monitoring the plants. Generally, most diseases may be detected and classified from the symptoms appearing on the leaves. A number of ways exist to detect a disease from leaf images such as SVM, LDA, kNN, and ZeroR; and CNN is one of them.  A good CNN architecture is capable of extracting complex features from the given data. But even the best CNN requires tweaking the parameters of CNN model and parameters such as test - train ratio, number of nodes in convolution layer, size of the image, number of detectors etc. For that here we are using genetic algorithms to get an optimum result from the model or to get an optimum model for plant diseases classification. A public dataset of 54K images of diseased and healthy plant leaves was used which was collected under controlled conditions, on this dataset I trained a CNN to identify 3 diseases (or absence thereof) for the apple leaves. The trained and tuned model achieved an accuracy of 93.86% on the test set, demonstrating the feasibility of this approach. Here, a genetic algorithm (GA) is used to optimally determine the architecture of a convolutional neural network (CNN) that is used to determine plant disease classification. The CNN is a class of deep feed-forward networks, which have seen major success in the field of visual image analysis. The proposed method was tested on the Plant diseases dataset. The results show that the genetic algorithm is capable of optimizing a CNN model successfully. The proposed method performs the entire process of optimization without any human intervention.

**Table of Contents (with page numbers)**

|  |  |  |
| --- | --- | --- |
| Sr No. | Content | Page No. |
| 1. | Introduction | 7 |
| 1.1 | Motivation | 8 |
| 2. | Literature Review | 9 |
| 2.1 | Outcome of Literature Review | 9 |
| 2.2 | Problem Statement | 10 |
| 2.3 | Research Objectives | 10 |
| 3. | Methodology and Framework | 11 |
| 3.1 | Algorithms, Techniques etc. | 12 |
| 3.2 | Detailed Design Methodologies | 15 |
| 4. | Work Done | 16 |
| 4.1 | Results and Discussion | 17 |
| 5. | Conclusion and Future Work | 18 |
| 5.1 | Proposed Work plan of the project | 18 |
|  | References | 19 |

**List of Figures List of Tables (if any)**

|  |  |  |
| --- | --- | --- |
| Sr. No. | Content | Page No. |
| 1 | Literature Review Table | 9 |
| 2 | Figure 1: Representing the flow chart of Genetic Algorithm | 13 |
| 3 | Figure 2: Array of RGB Matrix | 14 |
| 4 | Figure 3: A complete flow of CNN to process an input image and classifies the objects based on values. | 14 |
| 5 | Figure 4: Detailed design and flow of the program and explanation of code | 15 |

**1.  Introduction**

As the world around us becomes more and more automated so is our farming. Not only our traditional farming is becoming more and more machines worked where human involvement is decreasing with the introduction of Tractors, planters, harvesters, etc. but automation in agriculture has started to go to another level of automation all together. Slowly vertical farming is taking the place of traditional and “fancy farming” (organic farming, natural farming, etc). Vertical farming is a highly automated farming system where most of the monitoring and condition control is done by machines. When applied in a large scale manual checking of plants will become more and more difficult, hence there will be a requirement of automated systems of disease identification and prevention.

Generally, leaves of the plants are the first source to detect most of the plant diseases. Yellow and brown spots, primary and late blister, and other ailments caused by bacteria, virus and fungus can be detected automatically through efficient image processing techniques. Here we are using CNN to detect plant disease and Genetic algorithm to optimise the CNN model.

**1.1. Motivation**

Slowly vertical farming is taking the place of traditional and “fancy farming” (organic farming, natural farming, etc.). Vertical farming is a highly automated farming system where most of the monitoring and condition control is done by machines. When applied in a large scale manual checking of plants will become more and more difficult, hence there will be a requirement of automated systems of disease identification and prevention. Here we are using CNN to detect plant disease and Genetic algorithm to optimise the CNN model.

Deep CNN models are famous for their accuracy but they require a large amount of computational capacity, (big CPU or GPU) which are costly and hard to maintain. Instead of using high-end devices (hardware with high computational capacity) for deep CNN, here we propose an optimized shallow CNN model with 2 Convolutional Layer, 2 Pooling Layer, and 1 Fully-Connected Layers. Which can run on a small device or devices with limited computational capacity, with greater speed and good accuracy.

**2.  Literature Review**

**2.1.     Outcome of Literature Review**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Paper | Publisher and year | Objective | Method | Accuracy |
| Using Deep Learning for Image-Based Plant Disease Detection | Frontiers in plant science - September 2016 | To implement  a deep convolutional neural network to identify 14 crop species and 24 diseases | Convolutional Neural Network | 99.35% |
| Plant Leaf Disease Identification using Exponential Spider Monkey Optimization | Elsevier - October 2018 | To introduce exponential spider monkey optimization (ESMO), feature extraction using SPAM, a new approach for selection of feature subset, classifying using kNN, SVM, ZeroR, and LDA classifiers. | ESMO based feature selection method with SVM classifier. | 92.12% |
| Designing Convolutional Neural Network Architecture Using Genetic Algorithms | Int'l Conf. Artificial Intelligence - 2018 | To design CNN arch dynamically with high accuracy, less computational time and less computational resource. | Genetic algorithm and CNN | Na |

The main outcome of the literature review was the attainment of knowledge of possibilities of research in Genetic algorithm and CNN and the knowledge of the explored and unexplored areas of plant disease detection and CNN optimization.

**2.2.     Problem Statement**

To develop a CNN Model to detect plant disease using plant leaf images.

To keep the model small and fast so that it can predict quickly.

To optimise the model using Genetic algorithm.

**2.3.     Research Objectives**

To develop a small CNN model with accuracy more than 92.12%

**3.  Methodology and Framework**

**3.1. System Architecture**

Function load\_images

This function loads the image in a text file and classifies it into its respective diseases or healthy category using numbers. It also divides the dataset into test-dataset and train-dataset.

Function create\_train\_test

This function trains and tests the model for each set of given parameters.

Model: 2 Convolutional Layer, 2 Pooling Layer, and 1 Fully-Connected Layer (hidden layer)

Parameters: train\_test\_prop, detector\_value(Number of detectors), pixel( size of image in pixels), convolution\_node (number of convolution nodes).

Function cal\_pop\_fitness

Calculates the fitness of a given population i.e. calculates the test set accuracy for a given set of parameters.

Function select\_mating\_pool

Selects Parents for the next generation from the current generation.

Function crossover

Creates offspring generation from the current generation using one point crossover.

Function mutation

Creates changes in the new generation by introducing mutation in them using uniform mutation.

Main Function calls all the functions one by one and repeats the process for many generations.

**3.2.  Algorithms, Techniques etc.**

Genetic algorithm

A genetic algorithm may be a search heuristic that's inspired by Charles Darwin’s theory of natural evolution. This algorithm reflects the method of survival where the fittest individuals are selected for reproduction so as to provide offspring of subsequent generation.

Five phases are considered in a genetic algorithm.

* Initial population

The process begins with a group of values which is named a Population. Each individual may be a solution to the problem you would like to solve.

An individual is characterized by a group of parameters (variables) referred to as Genes. Genes are joined into a string to make a Chromosome (solution).

* Fitness function

The fitness function determines how fit a private is (the ability of a private to compete with other individuals). It gives a fitness score to each individual. The probability that a private are going to be selected for copy is predicated on its fitness score.

* Selection

The idea of selection phase is to pick the fittest individuals and allow them to pass their genes to subsequent generation.

Two pairs of values (parents) are selected on the basis of their fitness scores. The chance of being selected for reproduction depends on the fitness value.

* Crossover

Crossover is the most vital phase in a genetic algorithm. For each pair individual of the parent generation to be mated, a crossover point is chosen randomly from within the genes.

* Mutation

In certain new offspring formed, a number of their genes are often subjected to a mutation with an occasional random probability.

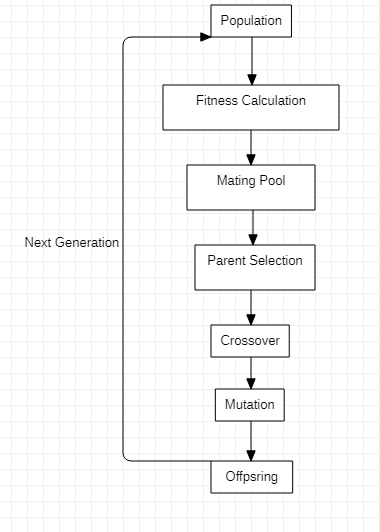


Figure 1: Representing the flow chart of Genetic Algorithm

In neural networks, Convolutional neural network (ConvNets or CNNs) is one among the most categories to do images recognition, images classifications. Objects detections, face recognition, are a number of the areas where CNNs are mainly used.

CNN image classifications takes an input-image, process it and classify it under given categories. Computers see the input-image as an array of pixels depending on the image resolution. Based on the image resolution, it'll see h x w x d( h = Height, w = Width, d = Dimension ). Eg., a picture of 6 x 6 x 3 array of a matrix of RGB (3 refers to RGB values) and a picture of 4 x 4 x 1 array of a matrix of a grayscale image.

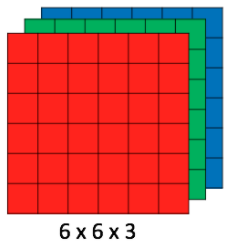


Figure 2: Array of RGB Matrix

Technically, on training and testing deep-learning CNN models each input image will pass through a series of convolution layers with filters, pooling layer, fully connected layers and apply Softmax function to classify an object with values between 0 and 1.

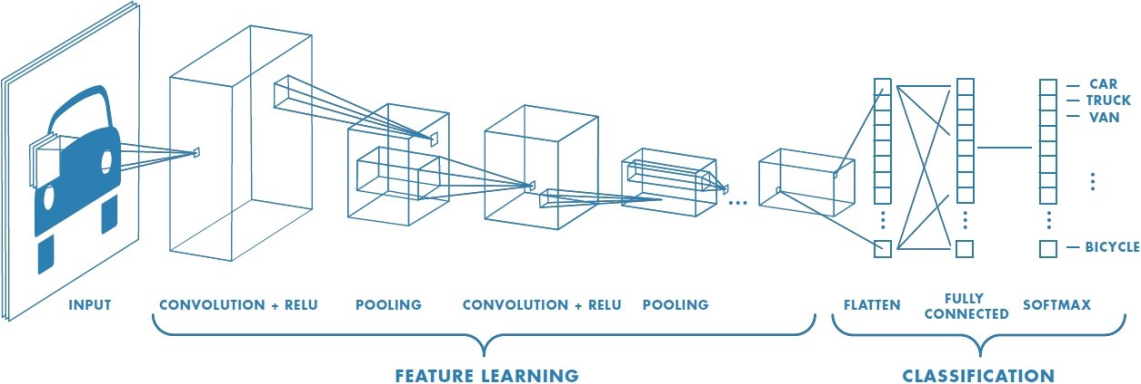


Figure 3: A complete flow of CNN to process an input image and classifies the objects based on values.

**3.3. Detailed Design Methodologies**

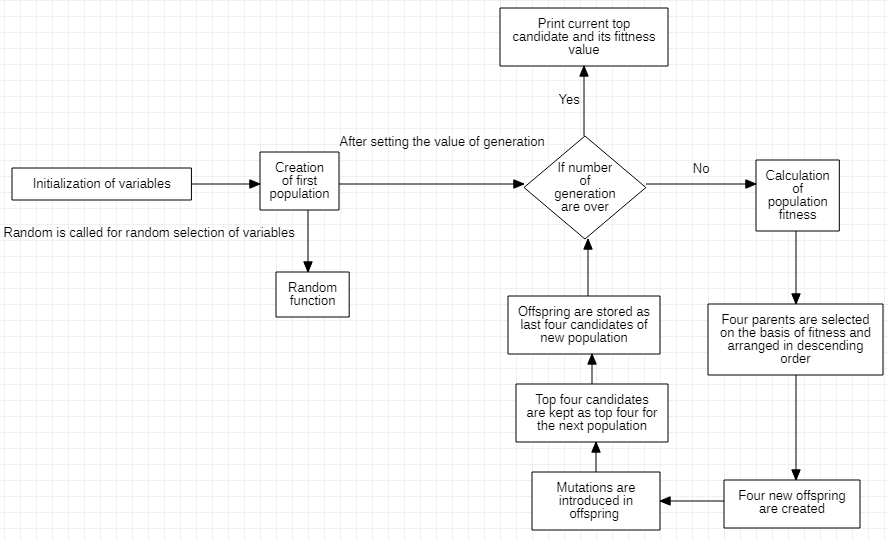
****

Figure 4: Detailed design and flow of the program and explanation of code

**4.    Work Done**

CNN Model:

A Simple CNN model with 2 Convolutional Layer, 2 Pooling Layer, and 1 Fully-Connected Layer (hidden layer).

Variables are:

* Test train proportion
* Number of detectors or filters  in Convolutional Layer
* Size of image in pixels
* Number of Convolution nodes in a fully connected layer

Convolutional Layer:

Filter\_size = 5

Strides = default

Padding = default

Activation function =  ReLU (rectified linear unit)

Fully-Connected Layer:

Activation function =  ReLU (rectified linear unit) and softmax (in output layer)

Genetic algorithm implementation:

Genes (Number of variable parameters in the model) = 4

Chromosomes (Number of solution per populations) = 8

Initial population is randomly generated.

Number of generation (Number iteration in the for loop)  = 8

In the for loop:-

* First fitness is calculated. Fitness value is a float value which is the accuracy attained by the model for the population (variable values in that particular population) values.
* Then the mating pool is selected and arranged in descending order of accuracy.
* Four offspring are generated using one point crossover.
* Uniform mutation is done on each offspring chromosome and the gene to be mutated is chosen randomly.
* A new population is made comprising of offspring and four top parent candidates (values in parent population)

After 8 generation the best chromosome is printed

**4.1. Results and Discussion**

An accuracy of 93.86 by the following values

* Test train proportion = 9:1
* Number of detectors or filters  in Convolutional Layer = 32
* Size of image in pixels = 128
* Number of Convolution nodes in a fully connected layer = 512

The Average prediction time in a system with 8GB CPU and number GPU is 2 minutes 30 seconds per 10K images.

**5.  Conclusion and Future Work**

In current scenario as many automated farms use automated systems for watering plants,

Planning the schedule of applying pesticides, schedules are made automatically for workers in vertical farms and in most factories identification of detected products is done by machines as it can be done faster. In a similar way diseases in plant can also be detected using plant leaf images classified by a simple CNN model.

**5.1. Proposed Work plan of the project**

For small far plant disease apps or websites can be developed where the farmer or worker can upload photos to detect the disease. In bigger more automated farm this can be done by a drone or a mechanised camera which can click photos and directly sent it to a system where diseases detection can be done.

**References**

1. Using Deep Learning for Image-Based Plant Disease Detection - Sharada P. Mohanty, David P. Hughes and Marcel Salathé

<https://www.frontiersin.org/articles/10.3389/fpls.2016.01419/full>

1. Plant Leaf Disease Identification using Exponential Spider Monkey Optimization - Sandeep Kumara, Basudev Sharmab, Vivek Kumar Sharmab, Harish Sharmac, Jagdish Chand Bansald

<https://www.sciencedirect.com/science/article/abs/pii/S2210537918300982>

1. Designing Convolutional Neural Network Architecture Using Genetic Algorithms - Ashray Bhandare and Devinder Kaur

<https://csce.ucmss.com/cr/books/2018/LFS/CSREA2018/ICA3302.pdf>

1. The Genetic Convolutional Neural Network Model Based on Random Sample - You Zhining and Pu Yunming

<http://article.nadiapub.com/IJUNESST/vol8_no11/31.pdf>

1. Introduction to Genetic Algorithms

<https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3>

1. Genetic Algorithm Implementation in Python

<https://towardsdatascience.com/genetic-algorithm-implementation-in-python-5ab67bb124a6>

1. CNN Architectures — LeNet, AlexNet, VGG, GoogLeNet and ResNet

<https://medium.com/@RaghavPrabhu/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet-7c81c017b848>

1. Understanding of Convolutional Neural Network (CNN) — Deep Learning

<https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>

1. Plant Disease Detection model using a Convolutional Neural Network

<https://towardsdatascience.com/plant-ai-plant-disease-detection-using-convolutional-neural-network-9b58a96f2289>

10. The Quest of Higher Accuracy for CNN Models

<https://towardsdatascience.com/the-quest-of-higher-accuracy-for-cnn-models-42df5d731faf>