Journal of Engineering Sciences
Vol 12, Issue 06, JUNE /2021
ISSN NO:0377-9254

Image-Based Plant Disease Detection: A Comparison of Deep Learning and Classical Machine Learning Algorithms

R.SRINIVAS, GADDALAPATI AMRUTHA VALLI, SANAGA PUSHPANJALI, SHAIK RIZWANA, VANGAVOLU PAVAN KUMAR, VENNA LAKSHMI HARITHA REDDY
DEPT OF CSE
KRISHNACHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES

ABSTRACT

Rapid human population growth requires corresponding increase in food production. Easily spreadable diseases can have a strong negative impact on plant yields and even destroy whole crops. That is why early disease diagnosis and prevention are of very high importance. Traditional methods rely on lab analysis and human expertise which are usually expensive and unavailable in a large part of the undeveloped world. Since smartphones are becoming increasingly present even in the most rural areas, in recent years scientists have turned to automated image analysis as a way of identifying crop diseases. This paper presents the most recent results in this field, and a comparison of deep learning approach with the classical machine learning algorithms.

I.INTRODUCTION

Human population steadily continues to grow, and along with it the need for food production increases. According to the UN projections [1], human population is expected to reach 9.7 billion in 2050, 2 billion more than today. Considering that most of the population growth is to occur in the least developed countries (around 80% increase in the next 30 years), where the food scarcity is the main problem, it is easy to conclude that minimizing food loss in those countries is a primary concern. It is estimated that the yield loss worldwide is between 20 and 40 percent [2], with many farms suffering a total loss.

Traditional methods for detecting diseases require manual inspection of plants by experts. This process needs to be continuous, and can be very expensive in large farms, or even completely unavailable to many small farm holders living in rural areas. This is why many attempts to automate disease detection have been made in the last few decades. One of the notable approaches is the use of hyperspectral imaging. Hyperspectral images are usually taken by satellites or airborne imaging devices and used for monitoring large areas. A downside of this approach is extremely high equipment cost, as well as high dimensionality and small number of samples which make them unsuitable for machine learning (ML) analysis.

Because of the recent breakthroughs in computer vision and the availability of cheap hardware, currently the most popular approach is the analysis of RGB images. The other motive for analysing RGB images is that with the current smartphone ubiquitousness these solutions have potential to reach even the most rural areas. RGB images can be analysed by classical ML algorithms or the deep learning (DL) approach.

Classical methods rely on image pre-processing and the extraction of features which are then fed into one of the ML algorithms. Popular algorithm choices are Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), Fully Connected Neural Networks (FCNN), Decision Trees, Random Forests etc. In the last few years, the researchers shifted almost exclusively to the DL methods for image classification tasks. The reason is that they almost always outperform classical algorithms when given reasonably sized dataset, and can be implemented without the need for hand-engineered features. In this paper, we compare the DL approach with classical ML algorithms for the study case of plant disease classification.

II.LITERATURE SURVEY

1)The global burden of pathogens and pests on major food crops
AUTHORS: Savary, Serge, et al.

Crop pathogens and pests reduce the yield and quality of agricultural production. They cause substantial economic losses and reduce food security

at household, national and global levels. Quantitative, standardized information on crop losses is difficult to compile and compare across crops, agroecosystems and regions. Here, we report on an expert-based assessment of crop health, and provide numerical estimates of yield losses on an individual pathogen and pest basis for five major crops globally and in food security hotspots. Our results document losses associated with 137 pathogens and pests associated with wheat, rice, maize, potato and soybean worldwide. Our yield loss (range) estimates at a global level and per hotspot for wheat (21.5% (10.1–28.1%)), rice (30.0% (24.6–40.9%)), maize (22.5% (19.5–41.1%)), potato (17.2% (8.1–21.0%)) and soybean (21.4% (11.0–32.4%)) suggest that the highest losses are associated with food-deficit regions with fast-growing populations, and frequently with emerging or re-emerging pests and diseases. Our assessment highlights differences in impacts among crop pathogens and pests and among food security hotspots. This analysis contributes critical information to prioritize crop health management to improve the sustainability of agroecosystems in delivering services to societies.

2)Using deep learning for image-based plant disease detection

AUTHORS: Mohanty, Sharada P., David P. Hughes, and Marcel Salathé

Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

3)A practical plant diagnosis system for field leaf images and feature visualization

AUTHORS: Fujita, E., et al.

An accurate, fast and low-cost automated plant diagnosis system has been called for. While several studies utilizing machine learning techniques have been conducted, significant issues remain in most cases where the dataset is not composed of field images and often includes a substantial number of inappropriate labels. In this paper, we propose a practical automated plant diagnosis system. We first build a highly reliable dataset by cultivating plants in a strictly controlled setting. We then develop a robust classifier capable of analyzing a wide variety of field images. We use a total of 9,000 original cucumber field leaf images to identify seven typical viral diseases, Downy mildew and healthy plants including initial symptoms. We also visualize the key regions of diagnostic evidence. Our system attains 93.6% average accuracy, and we confirm that our system captures important features for the diagnosis of Downy mildew.

4)Textural features for image classification

AUTHORS:Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein

Texture is one of the important characteristics used in identifying objects or regions of interest in an image, whether the image be a photomicrograph, an aerial photograph, or a satellite image. This paper describes some easily computable textural features based on gray-tone spatial dependancies, and illustrates their application in category-identification tasks of three different kinds of image data: photomicrographs of five kinds of sandstones, 1:20 000 panchromatic aerial photographs of eight land-use categories, and Earth Resources Technology Satellite (ERTS) multispecial imagery containing seven land-use categories. We use two kinds of decision rules: one for which the decision regions are convex polyhedra (a piecewise linear decision rule), and one for which the decision regions are rectangular parallelpipeds (a min-max decision rule). In each experiment the data set was divided into two parts, a training set and a test set. Test set identification accuracy is 89 percent for the photomicrographs, 82 percent for the aerial photographic imagery, and 83 percent for the satellite imagery. These results indicate that the easily

computable textural features probably have a general applicability for a wide variety of image-classification applications.

5)Support-vector networks

AUTHORS:Cortes, Corinna, and Vladimir Vapnik

The support-vector network is a new learning machine for two-group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high-dimension feature space. In this feature space a linear decision surface is constructed. Special properties of the decision surface ensures high generalization ability of the learning machine. The idea behind the support-vector network was previously implemented for the restricted case where the training data can be separated without errors. We here extend this result to non-separable training data.High generalization ability of support-vector networks utilizing polynomial input transformations is demonstrated. We also compare the performance of the support-vector network to various classical learning algorithms that all took part in a benchmark study of Optical Character Recognition.

III.EXISTING SYSTEM:

Human population steadily continues to grow, and along with it the need for food production increases. According to the UN projections [1], human population is expected to reach 9.7 billion in 2050, 2 billion more than today. Considering that most of the population growth is to occur in the least developed countries (around 80% increase in the next 30 years), where the food scarcity is the main problem, it is easy to conclude that minimizing food loss in those countries is a primary concern. It is estimated that the yield loss worldwide is between 20 and 40 percent [2], with many farms suffering a total loss.Easily spreadable diseases can have a strong negative impact on plant yields and even destroy whole crops. That is why early disease diagnosis and prevention are of very high importance.

DISADAVANTAGES:

- Data Collection Problem
- It searches from a large sampling of the cost surface.

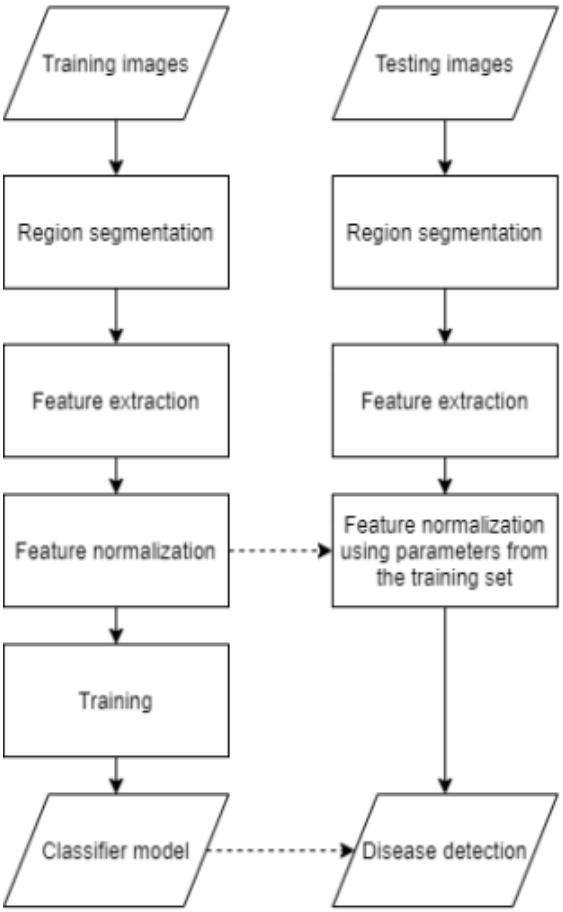
IV.PROPOSED SYSTEM:

Traditional methods for detecting diseases require manual inspection of plants by experts. This process needs to be continuous, and can be very expensive in large farms, or even completely unavailable to many small farm holders living in rural areas.the PlantVillage Dataset is used [3]. It consists of images of plant leaves taken in a controlled environment. In total, there are 54 306 images of 14 different plant species, distributed in 38 distinct classes given as species/disease pair.Classical methods rely on image pre-processing and the extraction of features which are then fed into one of the ML algorithms. Popular algorithm choices are Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), Fully Connected Neural Networks (FCNN), Decision Trees, Random Forests etc

ADAVANTAGES:

- Machine learning algorithm optimizes both variables efficiently, continuous or discrete
- Gives a number of optimum solutions, not a single solution. So different image segmentation results can be obtained at the same time
- Large number of variables can be processed at the same time.
- It can optimize variables with highly complex cost surfaces.

V.SYSTEM ARCHITECTURE:



UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

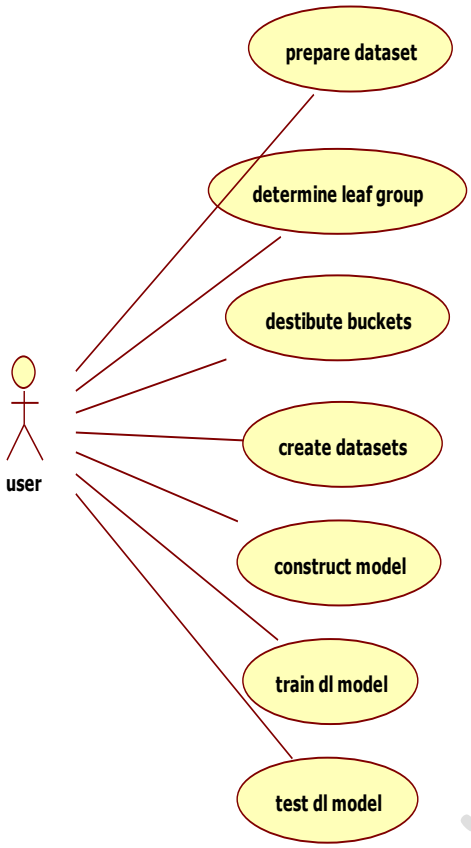
GOALS:

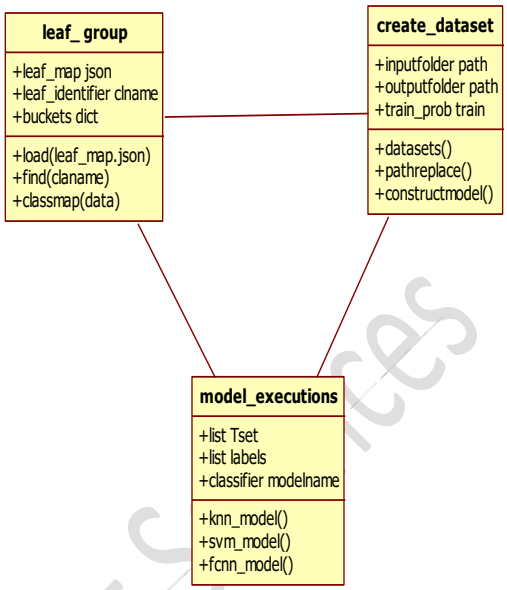
The Primary goals in the design of the UML are as follows:

- Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
- Provide extendibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development process.
- Provide a formal basis for understanding the modeling language.
- Encourage the growth of OO tools market.
- Support higher level development concepts such as collaborations, frameworks, patterns and components.
- Integrate best practices.

USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



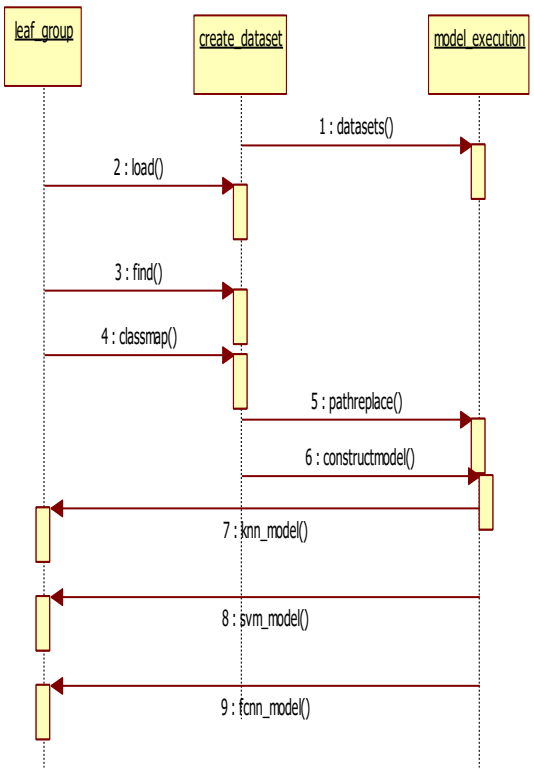


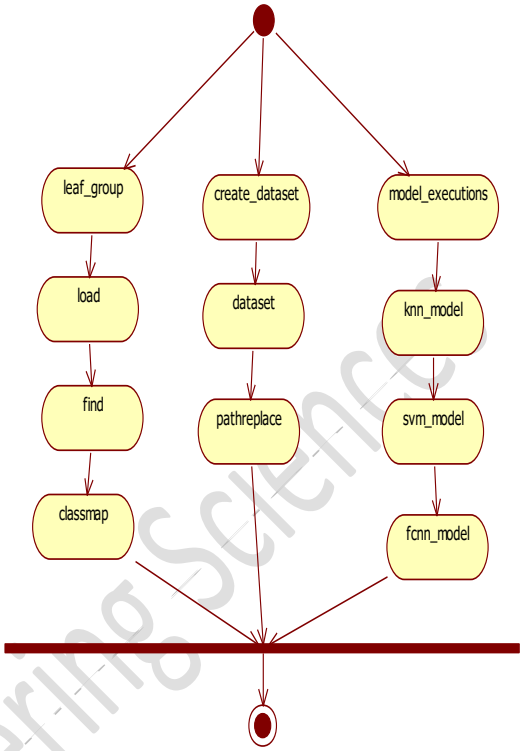
CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.





ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

VI.SYSTEM SPECIFICATION:

HARDWARE REQUIREMENTS:

- **System** : Intel i3
- **Hard Disk** : 1 TB.
- **Monitor** : 14' Colour Monitor.
- **Mouse** : Optical Mouse.
- **Ram** : 4GB.

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 10.
- **Coding Language** : Python.
- **Front-End** : Html. CSS
- **Designing** : Html,css,javascript.

Data Base : SQLite.

VII.CONCLUSION

This paper presents the dominance of the DL method over the classical ML algorithms. Both the simplicity of the approach and the achieved accuracy confirm

that the DL is the way to follow for image classification problems with relatively large datasets.

As the achieved accuracy of the DL method is already very high, trying to improve its results on the same dataset would be of little benefit. Further work with the DL model could be done by expanding the dataset with more diverse images, collected from multiple sources, in order to allow it to generalize better.

The considered ML algorithms achieved relatively high accuracy, but with error rates still an order of magnitude higher than the DL model. Further work in improving accuracy of the classical approach can be done by experimenting with other algorithms and by improving the features, as most likely they are the limiting factor of this approach.

VIII.REFERENCES

[1]United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).

[2] Savary, Serge, et al. "The global burden of pathogens and pests on major food crops." Nature ecology & evolution 3.3 (2019): 430.

[3] Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." Frontiers in plant science 7 (2016): 1419.

[4] Fujita, E., et al. "A practical plant diagnosis system for field leaf images and feature visualization." International Journal of Engineering & Technology 7.4.11 (2018): 49-54.

[5] Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein. "Textural features for image classification." IEEE Transactions on systems, man, and cybernetics 6 (1973): 610-621.

[6] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

[7] Cunningham, Padraig, and Sarah Jane Delany. "k-Nearest neighbour classifiers." Multiple Classifier Systems 34.8 (2007): 1-17.

[8] Haykin, Simon. Neural networks: a comprehensive foundation. Prentice Hall PTR, 1994.

[9] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[10] Duan, Kai-Bo, and S. Sathiya Keerthi. "Which is the best multiclass SVM method? An empirical study." International workshop on multiple classifier systems. Springer, Berlin, Heidelberg, 2005.

[11] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.