Tree Leaves Based Disease Prediction Using Data Mining Approaches

Kiruthiga R[1], Mr. P.Anbumani[2]
[2] ASST.PROF.
[1, 2] Krishnasamy college of Engineering and Technology,Cuddalore

Abstract- *Agricultural productivity is something on which economy highly depends. This is the one of the reasons that disease detection in plants plays an important role in agriculture field, as having disease in plants are quite natural. If proper care is not taken in this area then it causes serious effects on plants and due to which respective product quality, quantity or productivity is affected. Plant diseases are extremely significant, as that can adversely affect both quality and quantity of crops in agriculture production. Plant disease diagnosis is very essential in earlier stage in order to cure and control them. Generally the naked eye method is used to identify the diseases. In this method experts are involved who have the ability to detect the changes in leaf color. This method involves lots of efforts, takes long time and also not practical for the large fields. Many times different experts identify the same disease as the different disease. This method is expensive as it requires continuous monitoring of experts. Plant diseases can increase the cost of agricultural production and may extend to total economic disaster of a producer if not cured appropriately at early stages. The producers need to monitor their crops and detect the first symptoms in order to prevent the spread of a plant disease, with low cost and save the major part of the production. Hiring professional agriculturists may not be affordable especially in remote isolated geographic regions. Inspired by the deep learning break through in image based plant disease recognition, this work proposes deep learning models for image-based automatic diagnosis of plant disease severity. Deep learning algorithm in image can offer an alternative solution in plant monitoring and such an approach may anyway be controlled by a professional to offer his services with lower cost. It includes image segmentation which includes active contour method and image classification approach which includes neural network algorithm to predict various types of diseases. Back Propagation Neural Network (BPNN) is a multilayer neural network consisting of an input layer, minimum one hidden layer and an output layer. As the name of the algorithm implies, the errors (hence, learning) propagate backwards from the output layer to the input layer. Therefore, the process of back propagation is used to estimate weights which minimize the errors in reference with the network's modifiable weights. And also extend the approach to recommend the fertilizers based on severity analysis with measurements. Finally provide alert about fertilizer to farmers in the form of alert.*

Keywords- Data mining, Image processing, Segmentation, Classification

I. INTRODUCTION

Data mining is the process of extracting important information from large data sets. In this survey we are defining the application of data mining techniques in agricultural field. In this field, many methods have been discovered to help farmers in identifying diseased plant. The most widely used method is the image processing along with classification and clustering using neural networks. Throughout the years many algorithms were created, developed and modified to extract the knowledge from data sets. This also supports machine vision to provide image based automatic process control. This study presents a procedure for an automatic classification of different leaf diseases with special focus on early detection. Because of the high cost of chemical control, the detection and differentiation of several diseases at early stages epidemics allow more efficient application of agrochemicals. Data mining techniques, the process of extracting important and useful information for large data sets seems to solve the complex agricultural problems.

II. LITERATURE SURVEY

2.1 *TITLE: PLANT SPECIES IDENTIFICATION USING COMPUTER VISION TECHNIQUES: A SYSTEMATIC LITERATUREREVIEW*

AUTHOR: J. WÄLDCHEN AND P. MÄDER

In a manual identification process, botanist use different plant characteristics as identification keys, which are examined sequentially and adaptively to identify plant species. In essence, a user of an identification key is answering a series of questions about one or more attributes of an unknown plant (e.g., shape, color, number of petals, existence of thorns or hairs) continuously focusing on the most discriminating characteristics and narrowing down the set of candidate species. This series of answered questions leads eventually to the desired species. However, the determination of plant species from field observation requires a substantial botanical expertise, which puts it beyond the reach of most nature enthusiasts. Traditional plant species identification is almost impossible for the general public and challenging even for professionals that deal with botanical problems daily, such as, conservationists, farmers, foresters, and landscape architects. Even for botanists themselves species identification is often a difficult task. The situation is further exacerbated by the increasing shortage of skilled taxonomists. The still existing, but rapidly declining high biodiversity and a limited numbers of taxonomists represent significant challenges to the future of biological study and conservation. Recently, taxonomists started searching for more efficient methods to meet species identification requirements, such as developing digital image processing and pattern recognition techniques. The rich development and ubiquity of relevant information technologies, such as digital cameras and portable devices, has brought these ideas closer to reality. Digital image processing refers to the use of algorithms and procedures for operations such as image enhancement, image compression, image analysis, mapping, and geo-referencing. The influence and impact of digital images on the modern society is tremendous and is considered a critical component in a variety of application areas including pattern recognition, computer vision, industrial automation, and healthcare industries.

2.2 TITLE: HOW DEEP LEARNING EXTRACTS AND LEARNS LEAF FEATURES FOR PLANTCLASSIFICATION

AUTHOR: S. H. LEE, C. S. CHAN

Deep learning is a class of techniques in machine learning technology, consisting of multiple processing layers that allow representation learning of multiple level data abstraction. This paper begins with an introduction to deep learning. Next, we proceed to a critical and comprehensive review of existing methods and a description of the context of plant identification - i.e. how species are delimited by botanists using morphology. Then, we introduce the idea of deep learning for automatic processing and classification in order to learn and discover useful features for leaf data. We describe how computational methods can be adapted and learnt using visual attention. The universal occurrence of variability in natural object kinds, including species, will be described, showing first how it can confound the classification task, but also how it can be exploited to provide better solutions by using deep learning. In the plant identification domain, numerous studies have focused on procedures or algorithms that maximize the use of leaf databases, and this always leads to a norm that lea features are liable to change with different leaf data and feature extraction techniques. Heretofore, we have been engaged with ambiguity surrounding the subset of features that best represent the leaf data. Hence, in the present study, instead of delving in to the creation of feature representation as in previous approaches, we reverse engineer the process by asking DL to interpret an delicit the particular features that best represent the leaf data. By means of these interpretation results, we are able to perceive the cognitive complexities of vision for leaves as such, reflecting the trivial knowledge researchers intuitively deploy in their imaginative vision from the outset. And quantify the characteristics of features in each CNN layer and find that the network exhibits layer-by-layer transition from general to specific types of leaf feature. We find that this effect emulates the botanists' character definition sused for plant species classification.

2.3 TITLE: INTERACTIVE PLANTI DENTIFICATION BASED ON SOCIAL IMAGE DATA

AUTHOR: A.JOLY, H.GOËAU, P.BONNET

Building accurate knowledge of the identity, geographic distribution and uses of plants is essential for a sustainable development of agriculture as well as for biodiversity conservation. Unfortunately, such basic information is often only partially available for professional stakeholders, teachers, scientists and citizens, and often incomplete for ecosystems that possess the highest plant diversity. A noticeable cause and consequence of this sparse knowledge, expressed as the taxonomic gap, is that identifying plant species is usually impossible for the general public, and often a difficult task for professionals, such as farmers or foresters and even for the botanists themselves. Speeding up the collection and integration of raw botanical observation data is a crucial step towards a sustainable development of agriculture and the conservation of biodiversity. Initiated in the context of a citizen sciences project, the main contribution of this paper is an innovative collaborative workflow focused on image-based plant identification as a mean to enlist new contributors and facilitate access to botanical data. Since 2010, hundreds of thousands of geo-tagged and dated plant photographs were collected and revised by hundreds of novice, amateur and expert botanists of a specialized social network. An image-based identification tool – available as both a web and a mobile application – is synchronized with that growing data and allows any user to query or enrich the system with new observations. An important originality is that it works with up to five different organs contrarily to previous approaches that mainly relied on the leaf. This allows querying the system at any period of the year and with complementary images composing a plant observation. Extensive experiments of the visual search engine as well as system-oriented and user oriented valuations of the applications how that it is already very helpful to determine a plant among hundreds or thousands of species.

III. PROPOSED METHODOLOGY

Even when considering trees only, leaves show an impressively wide variety in shapes. It is however necessary to come up with a representation of what a leaf is, that is accurate enough to be fitted to basically any kind of leaf. The general shape of a leaf is a key component of the process of identifying a leaf. Botanists have a whole set of terms describing either the shape of a simple leaf, of the lobes of a palmate leaf, or of the leaflets of a compound leaf. The problem being that the borders between the different terms are not well defined, since leaves can naturally have non-canonical, intermediate shapes. The margin of the leaf is also a very important feature to spot. Its shape can be determining when trying to discriminate two species that have more or less the same global shape. It may consist of teeth of various sizes and frequencies, regularly arranged or not, from large spiny points, to small regular saw-like teeth, or even to a smooth entire border. We present a study on segmentation of leaf images restricted to semi-controlled conditions, in which leaves are photographed against a solid light-colored background. Such images can be used in practice for plant species identification, by analyzing the distinctive shapes of the leaves. We restrict our attention to segmentation in this semi controlled condition, providing us with a more well-defined problem, which at the same time presents several challenges. The most important of these are: the variety of leaf shapes, inevitable presence of shadows and specularities, and the time constraints required by interactive species identification applications. We evaluate several popular segmentation algorithms on this task. In everyday more urbanized and artificial world, the knowledge of plants, that used to constitute our most immediate environment, has somehow been lost, except for a handful of specialists. What is allegedly seen as unquestionable progress also scattered away the names and uses of so many trees, flowers and herbs. But nowadays, with a certain resurgence of the idea that plant resources and diversity ought to be treasured, the will to regain some touch with nature feels more and more tangible. And making it possible, for whoever feels the need, to identify a plant species, to learn its history and properties, is as much a way to transmit a vanished knowledge, as to allow people to get a glance at nature's unfathomable richness. The identification of species is the first and essential key to understand the plant environment. Botanists traditionally rely on the aspect and composition of fruits, flowers and leaves to identify species. But in the context of a widespread non-specialist-oriented application, the predominant use of leaves, which are possible to find almost all year long, simple to photograph, and easier to analyze from two-dimensional images, is the most sensible and widely used approach in image processing. In the process of tree identification from pictures of leaves in a natural background, retrieving an accurate contour is a challenging and crucial issue. In this paper we introduce a method designed to deal with the obstacles raised by such complex images, for simple and lobed tree leaves. A first segmentation step based on a light polygonal leaf model is first performed, and later used to guide the evolution of an active contour. Combining global shape descriptors given by the polygonal model with local curvature-based features, the leaves are then classified over leaf datasets. In this project we introduce a method designed to deal with the obstacles raised by such complex images, for simple and lobed tree leaves. A first segmentation step based on graph cut approach is first performed, and later used to guide the evolution of leaf boundaries. And implement classification algorithm to classify the diseases and recommend the fertilizers to affected leaves.

IV. DIFFRENT PLATFORMS

.NETFRAMEWORK -The .NET Framework (pronounced dot net) is a software framework developed by Microsoft that runs primarily on Microsoft Windows. It includes a large library and provides language interoperability (each language can use code written in other languages) across several programming languages. Programs written for the .NET Framework execute in a software environment (as contrasted to hardware environment), known as the Common Language Runtime (CLR), an application virtual machine that provides services such as security, memory management, and exception handling. The class library and the CLR together constitute the .NET Framework.

Design Features

Interoperability

Because computer systems commonly require interaction between newer and older applications, the .NET Framework provides means to access functionality implemented in newer and older programs that execute outside the .NET environment. Access to COM components is provided in the System. Runtime. Interop Services and System. Enterprise Services namespaces of the framework; access to other functionality is achieved using the P/Invoke feature.

Common Language Runtime engine

The Common Language Runtime (CLR) serves as the execution engine of the .NET Framework. All .NET programs execute under the supervision of the CLR, guaranteeing certain properties and behaviors in the areas of memory management, security, and exception handling.

Language independence

The .NET Framework introduces a Common Type System, or CTS. The CTS specification defines all possible data types and programming constructs supported by the CLR and how they may or may not interact with each other conforming to the Common Language Infrastructure (CLI) specification. Because of this feature, the .NET Framework supports the exchange of types and object instances between libraries and applications written using any conforming .NET language.

Base Class Library

The Base Class Library (BCL), part of the Framework Class Library (FCL), is a library of functionality available to all languages using the .NET Framework. The BCL provides classes that encapsulate a number of common functions, including file reading and writing, graphic rendering, database interaction, XML document manipulation, and so on. It consists of classes, interfaces of reusable types that integrate with CLR (Common Language Runtime).

SQLSERVER:

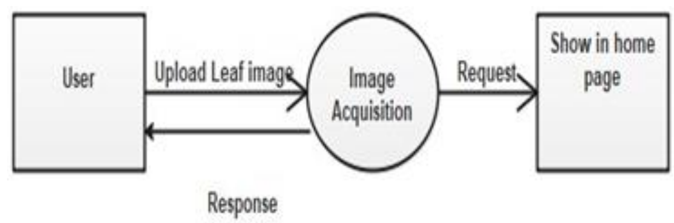
Microsoft SQL Server is a relational database management system developed by Microsoft. As a database server, it is a software product with the primary function of storing and retrieving data as requested by other software applications—which may run either on the same computer or on another computer across a network (including the Internet). Microsoft markets at least a dozen different editions of Microsoft SQL Server, aimed at different audiences and for workloads ranging from small single-machine applications to large Internet-facing applications with many concurrent users. Data storage is a database, which is a collection of tables with typed columns. SQL Server supports different data types, including primary types such as Integer, Float, Decimal, Char (including character strings), Varchar (variable length character strings), binary (for unstructured blobs of data), Text (for textual data) among others.

Microsoft SQL Server also allows user-defined composite types (UDTs) to be defined and used. It also makes server statistics available as virtual tables and views (called Dynamic Management Views or DMVs). In addition to tables, a database can also contain other objects including views, stored procedures, indexes and constraints, along with a transaction log. A SQL Server database can contain a maximum of 231 objects, and can span multiple OS-level files with a maximum file size of 260 bytes (1 exabyte). The data in the database are stored in primary data files with an extension .mdf. Secondary data files, identified with a .ndf extension, are used to allow the data of a single database to be spread across more than one file, and optionally across more than one file system. Log files are identified with the .ldf extension.

V. MODULE DESCRIPTION

Leaf Image Acquisition:

Tree leave is snap shot of the whole plant User can upload the leave images as Leaf images Image can be any size and any resolution This image is in RGB (Red, Green And Blue) form. Color transformation structure for the RGB leaf image is created.



Preprocessing:

Convert RGB image into gray scale image Using median filter algorithm to eliminate the noises in images The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing.

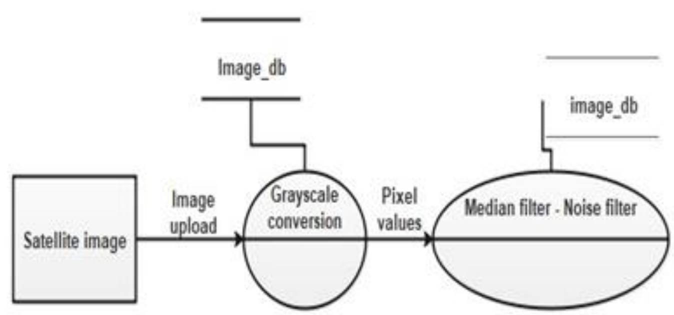
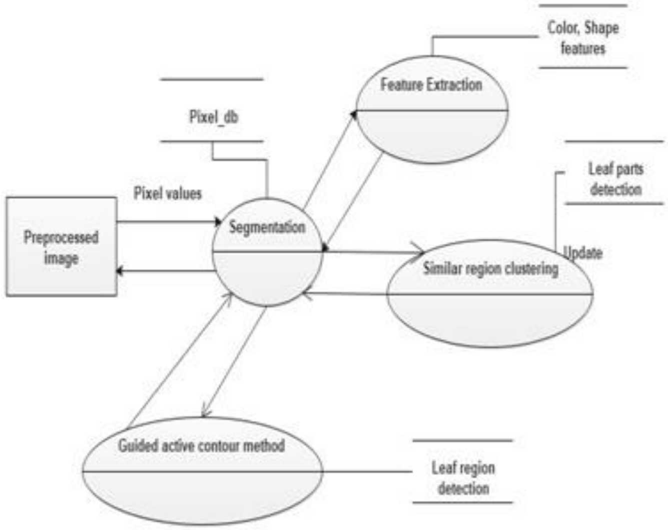


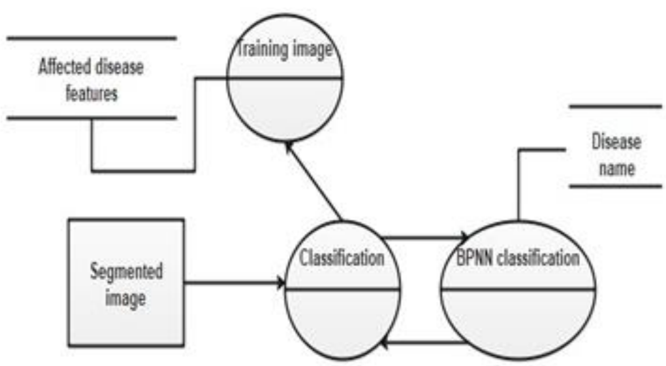
Image Segmentation :

Implement guided active contour method to segment the tree leaves At first leaves features are tracked and pointed as high level features Based on feature values, foreground leaves part is detected using contour method algorithm.



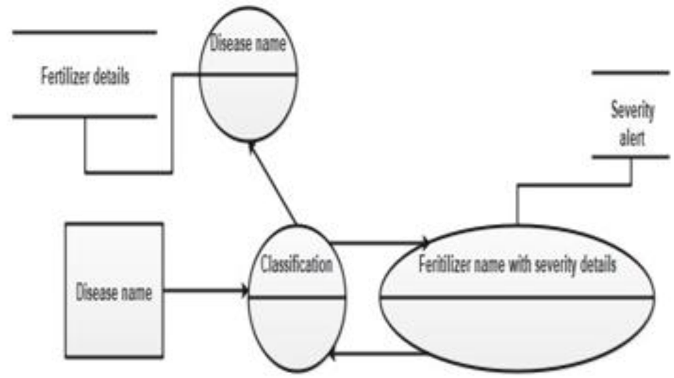
Disease prediction:

Track the features and calculate affected part features Implement Neural classification algorithm to predict the diseases Predict diseases as bacteria, fungi and other diseases



Fertilizer Recommendation:

Predict the diseases based on trained features Based on disease name, recommend the fertilizers that are added in trained side Provide improved classification rate based on severity levels.



VI. CONCLUSION

In this project, we overview the various techniques and algorithms are proposed for segmentation and classification methods for improve the quality of segmentation. But the result shows that segmentation algorithms do not work properly and can't implement in large datasets rather than proposed graph cut model. We have presented a method designed to perform the segmentation of a leaf in a natural scene, based on the optimization of a polygonal leaf model used as a shape prior for an exact active contour segmentation. It also provides a set of global geometric descriptors that, later combined with local curvature-based features extracted on the final contour, make the classification into tree species possible. The segmentation process is based on a color model that is robust to uncontrolled lighting conditions. But a global color model for a whole image may sometimes not be enough, for leaves that are not well defined by color only. The use of an additional texture model or of an adaptive color model could lead to a good improvement. Finally implement neural network classification algorithm to classify the leaf diseases as bacteria, fungi and virus. Then recommend the fertilizers to affected leaves based on measurements

VII. FUTURE ENHANCEMENTS

We can extend the framework to implement various classification algorithms and also classify the diseases not only leaves also in various vegetables and fruits with improved accuracy rate.

REFERENCES

[1] J.Wäldchen and P.Mäder, "Plant species identification using computer vision techniques: A systematic literature review," Archives of Computational Methods in Engineering, pp.1–37,2017.

[2] S.H.Lee, C.S.Chan, S.J.Mayo, and P.R emagnino, "How deep learning extracts and learns leaf features for plant

classification," Pattern Recognition, vol.71, pp.1–13, 2017.

[3] A.Joly, H.Goëau, P.Bonnet, V.Baki´c, J.Barbe, S.Selmi, I.Yahiaoui, J.Carré,E. Mouysset,J.-F.Molinoetal., "Interactive plant identification based on social image data," Ecological Informatics, vol.23, pp.22–34, 2014.

[4] A.R.Sfar, N.Boujemaa, and D.Geman, "Confidence sets for fine grained categorization and plant species identification," International Journal of Computer Vision,vol.111,no.3,pp.255–275,2015.

[5] J.Chaki,R.Parekh, and S.Bhattacharya, "Plant leaf recognition using texture and shape features with neural classifiers," Pattern Recognition Letters, vol.58, pp.61–68, 2015.

[6] N.Kumar, P.N.Belhumeur, A.Biswas, D.W.Jacobs, W.J.Kress, I.C.Lopez, and J.V. Soares,"Leafsnap: A computer vision system for automatic plant species identification," in Computer Vision–ECCV2012. Springer, 2012, pp.502–516.

[7] Y.Nareshand H.Nagendraswamy, "Classification of medicinal plants: An approach using modified lbp with symbolic representation," Neuro computing, vol.173,pp. 1789–1797,2016.

[8] G.L.Grinblat,L.C.Uzal,M.G.Larese, and P.M.Granitto, "Deep learning for plant identification using vein morphological patterns," Computers and Electronics in Agriculture,vol.127,pp.418–424,2016.

[9] J.Charters, Z.Wang, Z.Chi,A. C.Tsoi, and D.D.Feng, "Eagle: A novel descriptor for identifying plant species using leaf laminavascular features," in Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on. IEEE, 2014, pp.1–6.

[10] M.G.Larese, R.Namías, R.M.Craviotto, M.R.Arango, C.Gallo, and P.M.Granitto, "Automatic classification of legumes using leaf vein image features," Pattern Recognition, vol.47, no.1, pp. 158–168,2014.