Identifying the Amount of Pesticides Required for Rice Crops Using Image Processing System.

Toshal More[1], Divyesh Govekar[2], Sourabh Kokane[3], Dr. Sunil Chavhan[4]

[1-4]Department of Computer, Smt. Indira Gandhi College of Engineering, Navi Mumbai, India.

---***---

***Abstract*—**The world is heavily inclining towards Artificial Intelligence. It is making a significant impact in all walks of life and disciplines. Agricultural Science is the field which will be the most benefitted from this inclination towards Artificial Intelligence after Finance and Aerospace. Agricultural Science and research is widely welcoming new technology because this field had been isolated from the technological changes for an extended period of time. Taking all these facts into account we aim to develop a system which "Identifies the amount of Pesticides Required for Rice Crops using the Image Processing System.

1. INTRODUCTION

Computers With an evolution in technology, there has been a rapid evolution in diseases which affect the human life directly or indirectly. The diseases which affect the crops are of the category which affect humans indirectly. They lead to economic losses on a large scale. We aim at building an efficient system which will quantify the problems with the crop by taking its image as the input. The system's input will be the image. The output will be the Disease from which the rice crop is suffering from and the amount of fertilizer which should be sprayed to cure the disease. This system will be built by integrating a number of intelligent technologies like Image Processing and Machine Learning. The system should predict the amount of fertilizer to be sprayed on the crop depending on the disease which has affected the crop. We will be using a dataset of 1000 images for this purpose. These images are of 3 diseases which rice crops generally suffer from. The data is labelled, i.e. the images have been separated into relevant folders according to the diseases. We will use relevant image processing techniques on these images and obtain relevant features of the images. These features will be stored in an excel file. This excel file will be used to create classification models. We will be creating 2 classification models based on the best algorithm amongst following classification algorithms:

- ❖ Support Vector Machines (SVM)
- ❖ Back Propogation Neural Network (BPNN)

2. Image Processing -:

Canny Edge Detection

Canny edge detection is a technique to extract useful structural information from different vision objects and dramatically reduce the amount of data to be processed. It has been widely applied in various computer vision systems.

The general criteria for edge detection include:

1. Detection of edge with low error rate, which means that the detection should accurately catch as many edges shown in the image as possible
2. The edge point detected from the operator should accurately localize on the center of the edge.
3. A given edge in the image should only be marked once, and where possible, image noise should not create false edges

Among the edge detection methods developed so far, Canny edge detection algorithm is one of the most strictly defined methods that provides good and reliable detection. Owing to its optimality to meet with the three criteria for edge detection and the simplicity of process for implementation, it became one of the most popular algorithms for edge detection.

The images have to be segmented for extracting features Segmentation is the process of identifying boundaries for the image under consideration

Canny Edge detection Algorithm is used for detecting the boundaries in segmentation

3. Canny Edge Detector Algorithm

- It calculates the gradient of every point on the image
- Gradient is a physical quantity which has both magnitude and direction
- As the magnitude of gradient is calculated at every point, if there is a steep change in value, then it implies the presence of an edge
- If there is no steep change in the value of the gradient, then there is no edge detected at that point
- Calculation of the magnitude of gradient gives the direction of edge (if any is present)

4. Normalization

In image processing, normalization is a process that changes the range of pixel intensity values. Applications include photographs with poor contrast due to glare, for example. Normalization is sometimes called contrast stretching or histogram stretching. In more general fields of data processing, such as digital signal processing, it is referred to as dynamic range expansion.[1] The purpose of dynamic range expansion in the various applications is usually to bring the image, or other type of signal, into a range that is more familiar or normal to the senses, hence the term normalization. Often, the motivation is to achieve consistency in dynamic range for set of data, signals, or images to avoid mental distraction or fatigue. For example, a newspaper will strive to make all of the images in an issue share a similar range of grayscale

5. Classification Algorithms

Support Vector Machine algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification challenges. However, it is mostly used in classification problems. In this algorithm, plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Support Vectors are simply the coordinates of individual observation. In this paper mainly we will consider the input is based upon Support Vector Machine as training data, testing data is decision value. In this method we consider the following steps like Load Dataset, after loading the dataset will Classify Features (Attributes) based on class labels then estimate Candidate Support Value, like the condition is While (instances!=null), Do condition if Support Value=Similarity between each instance in the attribute then finding the Total Error Value. Suppose if any instance < 0 then the estimated decision value = Support Value\Total Error, repeated for all points until it will be empty. Therefore mainly we have calculated the entropy and gini index.

In machine learning, **support-vector machines** (**SVMs**, also **support-vector networks**) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The **support-vector clustering**[2] algorithm, created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications.

6. Back Propogation Neural Network

In machinelearning, **backpropagation** (**backprop**, **BP**) is a widely used algorithm in training feed forward neural networks for supervised learning. Generalizations of back propagation exist for other artificial neural networks (ANNs), and for functions generally – a class of algorithms referred to generically as "back propagation".] In fitting a neural network, back propagation computes the gradient of the loss function with respect to the weights of the network for a single input–output example, and does so efficiently, unlike a naive direct computation of the gradient with respect to each weight individually. This efficiency makes it feasible to use gradient methods for training multilayer networks, updating weights to minimize loss; gradient descent, or variants such as stochastic gradient descent, are commonly used. The back propagation algorithm works by computing the gradient of the loss function with respect to each weight by the chain rule, computing the gradient one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms in the chain rule; this is an example of dynamic programming.[3]

The term *back propogation* strictly refers only to the algorithm for computing the gradient, not how the gradient is used; but the term is often used loosely to refer to the entire learning algorithm, including how the gradient is used, such as by stochastic gradient descent. Backpropagation generalizes the gradient computation in the delta rule, which is the single-layer version of back propogation, and is in turn generalized by automatic differentiation, where back propogation is a special case of reverse accumulation (or "reverse mode").

7. OBJECTIVES

We aim at creating a system which objectively finds and quantifies what is wrong with a particular rice crop. The scope of the system encompasses the Agriculture Domain. It aims at making farming easy by introducing smart technologies in it and using the past data. This technology will be of immense use to the farmers. This is a smart way of farming and can save the damage caused by crop diseases to the farmers

Literature Survey

A. Image Acquisition

The real time images are fed directly from the camera. For further analysis, proper visibility and easy analysis of images, white background is created because most of leaves color varies from red to green for exact segmentation.

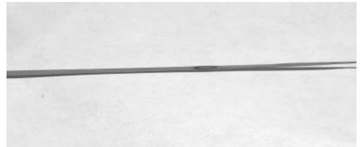
- Loading the image into python using the method "imread_collection()".
- This method takes argument as the path of the folder from which images are to be loaded.
- This method stores the all the images in the folder into a variable

B. Image Preprocessing

Image preprocessing is required to resize captured image from high resolution to low resolution. The image resizing can be done through the process of Canny edge detection algorithm .As the existing system had used Interpolation method, the method contains lots of calculation and it increases the complexity of the system. Captured input image is being converted into a grayscale image using color conversion by the equation

$$Image = 0.3R + 0.59G + 0.11B$$

The captured image placed in white background results in large differences between grey values of object and background.

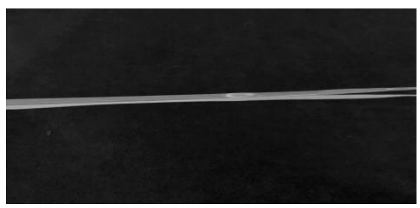


C. Disease Segmentation

Disease Segmentation is an important step to make something that is more meaningful and easier to analyze. The goal of segmentation is to simplify or change the representation of an image into multiple segments for further analysis. The method for segmentation in previous system was Otsu's Threshold , the method contains 3% to 25% error rate for rigid SVM System .In our system the segmentation is done by Canny edge detection algorithm and the error rate is reduced significantly.

Output of Canny Edge Detector Algorithm:

Output of segmentation:



D. Feature Extraction

Feature Extraction is one of the most interesting steps of image processing to reduce the efficient part of an image. After the image has been normalized, we find the following features from each image and store them to an external excel file.

1) Contrast
2) Homogeneity
3) Correlation
4) Dissimilarity
5) ASM
6) Energy
7) Mean
8) Variance

Once the features have been extracted, they may be used to build machine learning models for accurate object recognition or object detection.

For feature extraction of leaves recognition, gray-level co-occurrence matrix method is introduced. Gray-level co-occurrence matrix (GLSM matrices) is designed to measure the spatial relationships between pixels.

E. Classification of Image

Classification of image consists of database that contains pre-defined patterns that are compared with detected objects to classify them in a proper category. Classification will be executed on the basis of spectral defined feature such as density, texture etc. classification is done through three classification algorithms and finding best algorithm from them in terms of complexity and accuracy domain

F. User Interface

The **user interface** (**UI**), in the industrial design field of human-computer interaction, is the space where interactions between humans and machines occur.

The user will load the real time image(.jpg) into the python server through user interface. All necessary credentials will be taken from user and type of disease is updated to the user to respective Gmail account .

fig Android Application(UI)

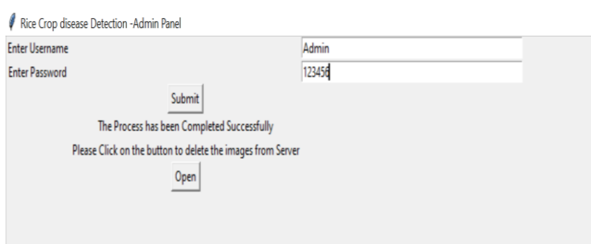


fig Admin Panel

fig Disease and Amount of pesticides is given to the user through mail

```
Hispa 0.5ml fenitrothion 100EC in 1 liter of water for spraying crop 225-230liter/acre
Hispa 10-12 DAS  1ml of methyl parathion 50EC in 1 liter of water for spraying crop 225-230liter/acre
Hispa 1.3ml monocrotophos 36SL in 1 liter of water for spraying crop 225-230liter/acre
Hispa 2ml chloropyriphos 20 EC in 1 liter of water for spraying crop 225-230liter/acre
>>>
```

fig Detected disease and required amount of pesticides

Sr. No	Research Paper Title	Conclusions	Limitations
1	An Image Processing and Neural Network based approach for detection and classification of plant leaf diseases	An automatic image processing neural network based approach has been studied and proposed for plant leaf disease detection.	Only classification algorithm is used the Neural Network and there is no comparison between the accuracies of several algorithm
2	Plant and leaf disease Detection and Classification Using Image Processing techniques .	The study reviews and summarizes image processing techniques for several plant species that have been used for recognizing plant diseases.	There is no comparison between the accuracies of different algorithm.
3	Tomato Plant Disease Classification in Digital Images Using Classification Tree	The study reviews and summarizes image processing techniques for tomato plant diseases and classification tree algorithm is used for classification	No neural network based algorithm is used for classification and also the accuracies between the algorithm is also not considered.

Related Research

The following modules will be used to accomplish the results stated above:

- image processing
- Machine Learning
- Server side scripting/backend
- user interface

8. METHODOLOGY

Image Processing

Step 1: Take the dataset of 1000 images

Step 2: Load the image in python

Step 3: Perform image processing on it to get only the required part of the image

This can include getting rid of the background, finding the outline, etc

Step 4: After the image has been processed, extract the required features from the image

Step 5: Store the extracted features into an excel file

Step 6: This file will form the training and testing data set for the classification algorithms which are to be used for the purpose of classification

Classification

Step 1: Load the data into Python for classification

Step 2: Pre-process the data if required

Step 3: Split the data into training and testing data set

Step 4: Form classifier models of the algorithms stated previously

Step 5: Train the classifier with the training data set obtained in step

Step 6: Test the classifier which has been created in the previous step

9. CONCLUSION

An application of detecting the plant diseases and providing the necessary suggestion for the diseases has been implemented. Hence the proposed objective was implemented on rice crops. The diseases specific to this plant were considered for testing of the algorithm. The experimental results indicate the proposed approach can recognize the diseases with little computational efforts. By this method the plant diseases can be identified at the initial stage itself and the pest control tools can be used to solve pest problems while minimizing risks to people and the environment. In order to improve diseases identification rate at various stages the training samples can be increased.

10. FUTURE WORK

As a part of future enhancement, in order to improve diseases identification rate at various stages, the training samples can be increased with the optimal features given as input condition for diseases identification and fertilization management of the crops. The complete process described in this project can be automated so that the result can be delivered in a very short time.

REFERENCES

1. https://www.kaggle.com/minhhuy2810/rice-diseases-image-dataset
2. https://www.seminarsonly.com/Engineering-Projects/Computer/amount-of-fertilization-required.php
3. *Gonzalez, Rafael (2018). Digital image processing. New York, NY: Pearson.* ISBN 978-0-13-335672-4*.* OCLC 966609831
4. https://arxiv.org/ftp/arxiv/papers/0709/0709.3967.pdf
5. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent System-Aurelien Geron
6. Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn, and Tensor Flow 2, 3rd Edition-Sebastian Raschka
7. https://developer.android.com/studio
8. "Use Java 8 language features | Android Developers"*. developer.android.com*. Retrieved October 25, 2017*.*