SURVEY ON AUTOENCODER BASED DETECTION OF NUTRITIOUS LEAVES

Mr. Raghavendrachar. S [1], Dr. Rekha B Venkatapur[2], Bhoomika.A.M[3], Bhoomika. K[4],

K. Kishan[5], K.R. Vageesh[6]

Assistant Professor, Dept of Computer Science, K S Institute of Technology Bengaluru, Karnataka[1]

Professor & Head, Dept of Computer Science, K S Institute of Technology, Bengaluru, Karnataka[2]

Dept Of Computer Science, K S Institute of Technology, Bengaluru, Karnataka[3-6]

Abstract: Several studies have been made on identifying diseases in mulberry leaves, however, identifying nutrient deficiency in mulberry leaves has not been accomplished. The silkworms that feed on nutrient-deficient mulberry leaves produce low-quality silk. There is a great need for identifying nutrient-rich and healthy mulberry leaves for feeding the silkworm to get good quality silk yield. This paper is focused on segregating nutritious mulberry leaves for feeding the silkworms for cocoon formation. The process involves image acquisition, processing, segmentation, feature extraction, and classification. Auto-Encoder is used for feature extraction from mulberry leaves and for discrete them into nutritious and nutrient-deficient leaves. The real-valued feature vectors are passed to machine learning algorithms like the Naïve Bayes classifier algorithm, Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) for classification. Among them KNN provides higher accuracy for segregating the leaves.

Keywords: Nutrient deficiency, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Naïve Bayes Classifier Algorithm, Auto-Encoder.

I.   INTRODUCTION

Mulberry leaf quality directly affects the growth and food consumption of larvae. It also significantly influenced the weight of the shell and pupal. Silkworms should be repeatedly fed with quality leaves. According to the explanation given, due to repeated contact with food, the time taken for conversion of assailants into body tissue is reduced so that more of the digested food is stored to be silk fibers. Feeding high quality mulberry leaves throughout the growing phase lowers mortality rate of the silkworms (FAO 1990). So, the ultimate factor that affects the growth and development of the silkworm and also the overall silk production is the quality of the mulberry leaves. Feeding deficient mulberry hampers silkworm growth and its economic value. It also results in bad cocoon formation, and hence lowers the quality of the silk produced. So, there is a need to segregate leaves, and feed only good and healthy leaves.

II.   RELATED WORK

Paper 1: Segregation of Mulberry Leaves using Image Processing [8]

This paper is designed to segregate these leaves using an intelligent monitoring platform framework. Image pre-processing, segmentation, feature extraction and classification are the steps performed. Color sensor is used to detect the leaf. K-Nearest Neighbor (KNN), Support Vector Machine(SVM) and Naive Bayes classifier algorithms for classification. Images are captured by Raspberry camera. Servo motor is connected to Raspberry camera and it does segregate the leaves by rotating the plates attached. Segregation is done based on the class specified by the Raspberry Pi.

Paper 2: Mulberry Leaf Disease Detection using Deep Learning[9]

This paper briefs about mulberry leaf diseases and their detection. In this project different algorithms and techniques are used in classification based on the different criteria for image segmentation. This system uses CNN to identify and classify mulberry leaf diseases.

Paper:3 Mulberry leaf disease detection using YOLO [10]

This paper proposes a model that detects the infection in mulberry leaves using Convolution Neural Networks (CNN) and You Look Only Once (YOLO). Before image processing, the image is divided into several grids. Accuracy and speed of detection are relatively high.

III.   OBJECTIVES

1. Objective is to detect healthy Mulberry leaves fed to silkworms, producing good-quality silk. Silk production yields to the economy of the country.

2. To use CNN, KNN for Classification, Auto-Encoder for dimensionality reduction, and techniques used to process images.

IV.   METHODOLOGY

In this section, the essential datasets are segregated and stored in an individual folder and are referred to as the Training folder. This folder consists of both healthy mulberry leaves and nutrient-deficient mulberry leaves. The datasets in this folder are named training datasets, which are used to train the system, where the morphological characteristics and features from it are extracted and compared with the test images to get the required result.

Step 1: Image acquisition
Image acquisition is the process of capturing the image from the input source. This captured image will be saved in an individual folder referred to as a Testing folder. The datasets in this folder are named testing datasets.

Step 2: Image pre-processing
This is a technique used for improving the quality of the image and eliminating the noises in the image. The captured images are processed before they are implemented in other steps.

The inputs of images will consist of different sizes and qualities. Both dimensions should match for comparing the test image with the training datasets. The system throws an error if the resizing of the image is not appropriate. The test images are transformed into grayscale images and are reduced to the dimension which matches the training dataset's dimension.

Step 3: Image segmentation
Image segmentation is used for converting digital images into multiple parts consisting of similar identities. This is used for leaf identification and boundary line identification in the image. In our proposed system, the segmentation step consists of cropping, masking, and contouring the image.

Step 4: Feature extraction
Features are nothing but information that is extracted from an image. They are the real-valued vector numbers that describe an image at high-level perspective.

Step 5: Classification
The feature vectors calculated in the feature extraction, becomes the classifier input. Classifiers like K-Nearest Neighbour, Support Vector Machine and Naïve Bayes are used for the computation purpose. Each training feature vectors are compared with the test feature vectors.

The difference is calculated between healthy and nutrient deficient image featured vectors. The test image is classified to certain class (healthy or deficient) which contains the least difference.

a. KNN is one of the machine learning algorithms used for image classification. This algorithm depends on distance between the feature vectors. The unknown data points are classified by finding k-closest examples common classes.

b. SVM is basically a supervised learning algorithm and their main purpose is to classifying the data into different classes. Margin is the representation of distance between either nearest possible point. The goal is among support vectors selecting a hyperplane that has the maximum possible margin in the provided dataset.

c. Naive Bayes algorithm is a probabilistic classifier. They create probability distribution to all the available classes and decide which to use. They work by the principle of conditional probability distribution which involves probability of assigning a given class with distinct set of features.

1.Auto-encoder architecture

Autoencoder is an Artificial Neural Network where initially the network is first trained based on the provided input. The network has the capacity to reconstruct the provided input from the features it has picked up from it and gives an

approximation output as that of input. The training step composes of the error computation and error backpropagating. The architecture of an autoencoder is identical to bottleneck.

For encoding purpose, the encoder part is used. First half of the network has a decreasing number of hidden units in each layer, with the help of this part encoding function is achieved. The decoding function is achieved by the second half. This part contains of increasing number of hidden units in each layer and from encoded data it will reconstruct the original input.

This working tells that the autoencoder structure resembles a bottleneck.

Step 1: The process of encoding the given input data. Using biases and initialized weights it encodes the given data.

Step 2: The process of decoding the given input data. From the encoded data reconstruction of original input takes place which later used for testing the reliability of the encoding process.

Step 3: The process of backpropagating the occurred error. To determine the reliability of the encoding the loss function is computed after the reconstruction takes place. Backpropagation of generated error is done.

The above training process takes place multiple times until reconstruction reaches to an acceptable level in determining discrete nutrient and nutrient deficient mulberry leaves.
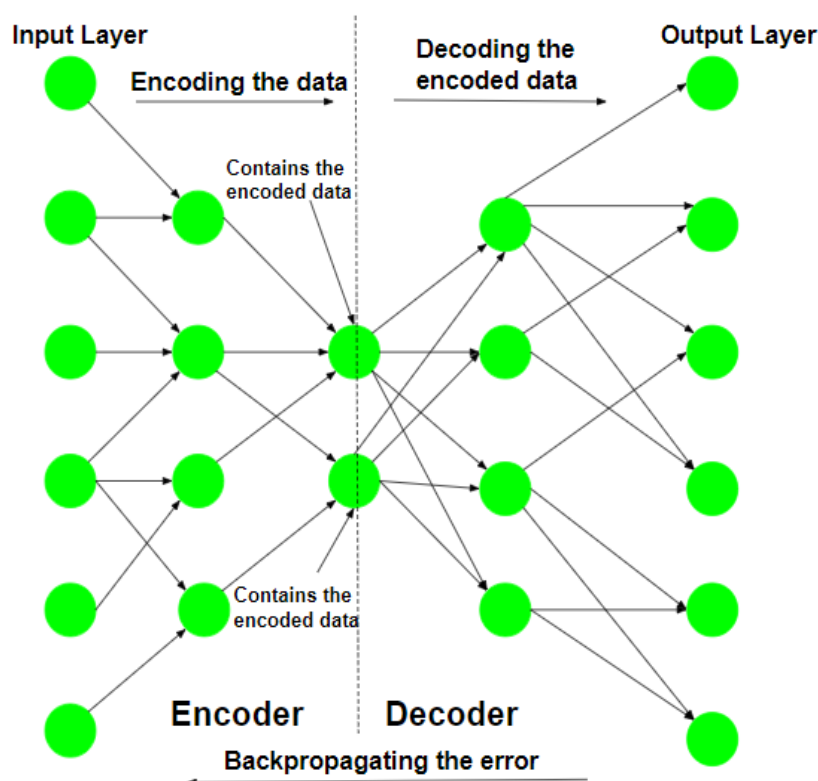
2.Application Requirements

Image Processing:

Image processing is the process of transforming an image into a digital form and performing certain operations to get some useful information from it. The image processing system usually treats all images as 2D signals when applying certain predetermined signal processing methods.

Jupyter Notebook:

The basic objective of the open-source scientific computing programme Jupyter Notebook is to mix equations, visuals, and live code. It supports more than 40 programming languages. The terms Julia, Python, and R are combined to form

the moniker Jupyter. While Anaconda comes preinstalled, Jupyter is primarily designed for data science and analytics applications. Data sets, such as visuals and charts, are produced by modules like Matplotlib, Plotly, or Bokeh in Anaconda.

Open CV:

OpenCV is a sizable open-source library for image processing, machine learning, and computer vision. It now plays a significant part in real-time operation, which is crucial in modern systems. Using it, one may analyse pictures and movies to find faces, objects, and even human handwriting. Python is able to handle the OpenCV array structure for analysis when it is integrated with different libraries, such as NumPy. We use vector space and apply mathematical operations to these features to identify visual patterns and their various features.

ACKNOWLEDGEMENT

REFERENCES

[1] https://vikaspedia.in/agriculture/farm-based-enterprises/sericulture/mulberry-cultivation
[2] https://pyimagesearch.com/2016/08/08/k-nn-classifier-for-image-classification/
[3] https://www.slideshare.net/mohdyounuswani/role-of-micro-nutrients-and-their-deficiency-symptoms-in-mulberry-93681380
[4] https://www.healthline.com/nutrition/foods/mulberries#vitamins-and-minerals
[5] https://www.ijresm.com/Vol.2_2019/Vo2_Iss5_May19/IJRESM_V2_I5_274.pdf
[6] https://www.geeksforgeeks.org/ml-auto-encoders/
[7] https://patents.google.com/patent/CN104761350A/en
[8] ref:https://www.ijresm.com/Vol.2_2019/Vol2_Iss5_May19/IJRESM_V2_I5_274.pdfhttps://www.ijresm.com/Vol.2_2019/Vol2_Iss5_May19/IJRESM_V2_I5_274.pdf
[9] ref: https://www.ijeat.org/wp-content/uploads/papers/v9i1/A1521109119.pdf
[10] ref: https://www.ijariit.com/manuscripts/v7i3/V7I3-2039.pdf