Machine Learning for Cactus (Beles) Diseases Detection

A thesis submitted to

In partial fulfillment of the requirements for the award of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

By

HAILAY BEYENE BERHE

(Roll No. 160500401001)

DEPARTMENT OF COMPUTER SCIENCE

PARUL UNIVERSITY

LIMDA, WAGHODIYA, VADODARA-301760, GUJARAT, INDIA

Under the guidance of

Dr. NARAYAN A. JOSHI

Department of MCA, Dharmsinh Desai University, Nadiad

December, 2019.

Abstract

Machine learning is very important technology that can support people in different disciplines (Agriculture, health centers, household, transportation, etc) and different levels of life. Machine learning increases accuracy of performance (prediction). It uses various types of data (image, video, audio and text) for different purposes and applications. Our work has focused on cactus diseases detection to early prevent the reduction of productivity (quantitatively and qualitatively) of the cereal. To do this, we have used unhealthy and healthy cactus images. The images were enhanced, noises were removed and images were segmented to create better model using imadjust, guided filter and K-means clustering techniques respectively. These image preprocessing techniques were selected from many techniques after implementing each technique and measuring their performances. As part of creating the model, feature extraction techniques (Color histogram, Bag of features and GLCM) were applied to extract color, bag of features and texture features respectively. After testing the model applying these features, bag of features were found to be best for creating better model and they were selected as features of our model. We created our machine learning model using bag of features applying linear SVM. Other machine learning algorithms were used to train and test the model for detecting the diseases, but linear SVM was found with best performance (97.2%). In this task, 75% of each class was used for training and 25% was used for testing the model. Finally, the similarity for classification was checked using linear kernel, RBF kernel and Polynomial kernel and an average accuracy of 94% was achieved though linear kernel was the best classifying method with an accuracy of 98.951%.

Keywords: Machine learning, supervised learning, diseases detection, training, classification, bag of features, K-means, MSE, PNSR, and linear SVM, cactus.