Classification of Plant Seedlings using Machine Learning Techniques

Raviraj Pawar, Vaibhav Patil, Prajakta Patil, Rukmini Chougule, Prajakta Chavan, Rohan Maske, Sayali Sawant, Prajakta Gaikwad, Rutuja Velhal, Sneha Kshirsagar

'Students, Department of Electronics and Telecommunication and Information Technology, Government College of Engineering, Karad'

Abstract

Our project, SeedRec is a process in which different varieties of seedlings are classified into different classes, using machine learning algorithms. Seedlings are classified on the basis of their morphological features. These algorithms are used to learn from data which can be used for predictions, for pattern recognitions and classification of input data. We performed this classification using tools such as Numpy, Keras and Tensorflow.

Keyword: seedling classification, Machine Learning, agricultural activity

I. Introduction

In today's world, most of the agricultural activities are done with a lack of modern technology. Currently, seedling classification is done based on knowledge of human being. This manual classification is inefficient. Specifying the nature of seedlings require a specialist judgment and it is time consuming. To overcome this, we have tried to use modern technology. Our project, SeedRec is a process in which different varieties of seedlings are classified into different classes, using machine learning algorithms. It will help to produce better results and will be efficient. Seedlings are classified on the basis of their morphological features. These algorithms are used to learn from data which can be used for predictions, for pattern recognitions and classification of input data. We performed this classification using tools such as Numpy, Keras and Tensorflow. The developed model is experimented using seedling dataset and then seedling classes are predicted using the developed model. The experiment is conducted on more than 10 species of seedlings. Finally by using the developed model, determinant factors for classifying seedlings are identified and ranked. Being able to identify the different species of seedlings growing in agricultural areas and to automatically detect the presence of invasive species is crucial. Identifying plants is usually a difficult task, sometimes for professionals like farmers, wood exploiters etc. Using content-based retrieval technologies, it is possible for us to classify seedlings.

The content-based seedling classification task, SeedRec is focused on tree, herb and fern species identification based on different types of datasets. The number of species was 6, and there were 5 view-points at their datasets: number of leaves, height of seedling, color, texture and length of leaf. Approximately 150 records are collected. The data was present in form of textual data which was created by observing plant seedlings. Those observations were inserted into the dataset with respect to the above five view-points.

The aim of task is to classify datasets into the known categories (species), but the classification system has to be robust to unseen categories. It was a more difficult problem, because the test set contained datasets of species that were not in the training set which were unseen categories. Besides, the seedling's contextual metadata were also available.

II. Literature Review

Studies have been done on classification of seeds using machine algorithms. These research have used different machine classifiers and have performed accuracy for carrying out their work.

According to paper [1], they assess the discriminating power of different seed characteristics for the unique identification of seeds of weed species. They have used a simple Bayesian approach (naı¨ve Bayes classifier) to evaluate morphological, color and textural characteristics measured from video images, establishing their importance as classification features for weed seeds identification. The final classifier based on the optimal set of features shows a remarkable good performance. In addition, they have presented classification results obtained using the same feature set as input of a committee of ANNs.

A study by [2] deals with finding the accuracy of Seeds data. For doing this, six machine learning algorithms are used, that is, Logistic Regression (LR), Linear Discriminant Analysis (LDA),K Neighbors Classifier (KNN), Decision Tree Classifier (CART), Gaussian NB (NB), Support Vector Machine (SVM).

In this paper [3] , Classification through Weka software is used here for the analysis of structural activity relationship. The above data was then processed for classification which was prepared in ARFF format. Four classification methods are used in our system such as Function, Bayes, Meta and Lazy on seeds datasets to test the classification performance in this system.

In this paper [4], Python 2.7 programming language was used for image processing steps such as seed segmentation and feature extraction with the help of Scipy and Numpy scientific computing libraries. Following programming libraries were utilized to implement image processing algorithms: Mahotas, Scikit-image image processing library, and OpenCV. Matlab (2012a, The MathWorks, Inc., Natick , Massachusetts) computing language was also used for implementing feature selection and neural networks.

III. Materials and Methods

1. Seedling Samples

10 types of seedlings were collected for this research: mango, tomato, peanut, almond, mustard and coconut.

We have built a database containing approximately 150 records of 6 species considered.

Feature extraction:

We measured a number of features from the raw seed images to be later used for classification purposes. As stated above, these features correspond to morphological, color and textural characteristics of the seeds.

2. Programming Environment

In this research,

1. We have collected a dataset of 150 records approximately.

2. **Python 3.6** language was used for content based classification on the parameter values specified in the dataset.

3. We have used **pycharm IDE** for algorithm development.

4. We included some libraries like **numpy** (for arranging data in array like structure), **pandas** (for arranging data in data frame for simplicity)

5. We have used **Random Forest** which is an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

6. Computing languages was also used for implementing feature selection and neural networks.

IV. Design

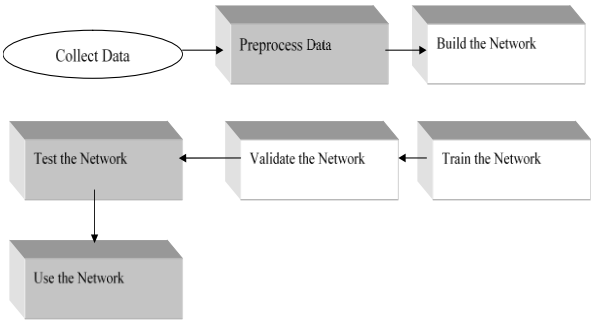

Fig.1. Design of classification of seedling using machine learning techniques.

V. Conclusion

In this study, our goal was to classify seedling accuracy using machine algorithms. For this we used different algorithms to implement and selected one best suited out of them. Five independent variables were selected and examined in developing the model. The model is developed for prediction of determinant factors of seedlings based on Machine Learning.

In this research, we used Logistic Regression (LR), Linear Discriminant Analysis (LDA), K Neighbors Classifier (KNN), Decision Tree Classifier (CART), Gaussian NB (NB), Random forest and Support Vector Machine (SVM). We send information in normal and scaled manner, and then prediction was performed. Based on all the outcomes we concluded that Random forest performs better than all algorithm. The accuracy of this algorithm emerged out to be more than that algorithm. The machine learning approach is recommended as flexible and precise way to solve the stated issue.

VI. Future Scope

1. We can also classify the seedlings using image processing.

2. Its identification system can be run on mobile applications.

3. It can also be implemented using CNN Classifier.

4. Besides there are also un-supervised machine learning techniques, such as clustering to classify the seeds.

5. Moreover we can classify the other categories of seeds using these techniques and classifiers after extracting these features.

VII. Acknowledgement

We are thankful to all of them who have motivated us, supported us, encouraged us and guided us throughout the project sessions and at all times of project.

VIII. References

[1] Granitto P M, Navone H D, Verdes P F, Ceccatto H A. “Weed seeds identification by machine vision. Computers and Electronics in Agriculture”, 2002; 33: 91–103.

[2] Divya Agrawal, Prof. Priyanka Dahiya. “Comparisons of Classification Algorithms on Seeds Dataset using Machine Learning Algorithm”

[3] Raja Haroon Ajaz , Lal Hussain. “Seed Classification using Machine Learning Techniques”

[4] Ferhat Kurtulmuş , İlknur Alibaş , Ismail Kavdir. “Classification of pepper seeds using machine vision based on neural network”